To track or not to track: examining perceptions of online tracking for information behavior research

Mykola Makhortykh
Institute of Communication and Media Studies, University of Bern, Bern, Switzerland

Aleksandra Urman
Institute of Communication and Media Studies, University of Bern, Bern, Switzerland and
Social Computing Group, University of Zurich, Zürich, Switzerland

Teresa Gil-Lopez
Institute for Communication Psychology and Media Pedagogy, University Koblenz – Landau, Landau, Germany, and

Roberto Ulloa
GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany

Abstract

Purpose – This study investigates perceptions of the use of online tracking, a passive data collection method relying on the automated recording of participant actions on desktop and mobile devices, for studying information behavior. It scrutinizes folk theories of tracking, the concerns tracking raises among the potential participants and design mechanisms that can be used to alleviate these concerns.

Design/methodology/approach – This study uses focus groups composed of university students ($n = 13$) to conduct an in-depth investigation of tracking perceptions in the context of information behavior research. Each focus group addresses three thematic blocks: (1) views on online tracking as a research technique, (2) concerns that influence participants' willingness to be tracked and (3) design mechanisms via which tracking-related concerns can be alleviated. To facilitate the discussion, each focus group combines open questions with card-sorting tasks. The results are analyzed using a combination of deductive content analysis and constant comparison analysis, with the main coding categories corresponding to the thematic blocks listed above.

Findings – The study finds that perceptions of tracking are influenced by recent data-related scandals (e.g. Cambridge Analytica), which have amplified negative attitudes toward tracking, which is viewed as a surveillance tool used by corporations and governments. This study also confirms the contextual nature of tracking-related concerns, which vary depending on the activities and content that are tracked. In terms of mechanisms used to address these concerns, this study highlights the importance of transparency-based

mechanisms, particularly explanations dealing with the aims and methods of data collection, followed by privacy- and control-based mechanisms.

Originality/value – The study conducts a detailed examination of tracking perceptions and discusses how this research method can be used to increase engagement and empower participants involved in information behavior research.

Introduction

The growing use of digital media creates new challenges and opportunities to study information behavior in media environments where users have more choice regarding where and how to engage with information. One such challenge-opportunity is the use of online tracking for the automated capture of information behavior (e.g. visits to specific websites or mobile app usage) [1]. Using approaches that range from HTML extraction from browsers (Adam *et al.*, 2019) to the interception of web traffic via a man-in-the-middle attack [2] (Bodó *et al.*, 2017) to the recording of mobile device screens (Krieter, 2019), tracking provides transactional data about user actions and sheds light on multiple aspects of individual and collective behavior. In this study, we examine the intricacies of using online tracking for behavior research by organizing a set of focus groups and examining the perceptions and concerns of potential participants about tracking as a research method as well as possible mechanisms via which to alleviate these concerns.

Tracking tackles two issues associated with self-reported data on information behavior: the inability of participants to consistently recall their exposure to information (Prior, 2013) and the effect of social desirability on the information reported by participants (Krumpal, 2013). By automatically capturing user behavior, tracking puts less pressure on the participants than self-reporting approaches (e.g. media diaries) and allows the tracing of their interactions with digital platforms and applications in a consistent way. Consequently, it increases the validity of behavioral observations and facilitates research on diverse subjects ranging from information consumption routines (Möller *et al.*, 2019) to incidental information exposure (Thorson, 2020).

Conversely, tracking requires complex technical infrastructure (Bodó *et al.*, 2017; Kreuter *et al.*, 2020) and is often characterized by low participation rates, which can lead to sampling biases (Stier *et al.*, 2020a; Jürgens *et al.*, 2020). Low participation can be attributed to several causes, but the most common causes are privacy and security concerns as well as a lack of incentives (Keusch *et al.*, 2019a). While there are studies (Keusch *et al.*, 2019a, b; Kreuter *et al.*, 2020; Revilla *et al.*, 2019; Ochoa and Revilla, 2018) investigating the concerns preventing participants from engaging with passive data collection approaches, including tracking, the question of how these concerns can be alleviated remains under-studied.

A few exceptions include suggestions to implement strict data-sharing policies (Keusch *et al.*, 2019a) and differentiate between more- and less-intrusive data requests (Kreuter *et al.*, 2020). However, most of these studies rely on quantitative surveys, which tend to focus on mobile tracking and allow fewer chances for a detailed analysis of the relationship between participant concerns and research design. To our knowledge, no study has thus far conducted an in-depth investigation of the methods of alleviating tracking-related concerns via software design or drawn comparisons between mobile and desktop tracking.

To better understand concerns about tracking as a research technique and how they can be addressed, we organized several focus groups to investigate how potential participants in a tracking-based project perceive the use of tracking for studying their behavior. Specifically, we pursue three goals: to examine how participants imagine tracking as a research technique and the kinds of folk theories they may have about it, to scrutinize what concerns participants

have about the use of tracking to study their behavior and to identify design mechanisms that can increase participants' motivation to be involved in tracking research.

The remainder of the paper is organized as follows: first, we describe the benefits and caveats of tracking as a research method. This is followed by a discussion of related work that primarily deals with folk theories of technology, online privacy perceptions and user-centered design. Then, we elaborate on our methodology and explain how we recruited the participants and structured the focus groups. Next, we present our findings pertaining to the participants' perceptions of tracking, their concerns about it and potential ways of alleviating these concerns. The paper concludes by discussing the implications of our findings for behavior research, the limitations of the current study and directions for future studies.

Online tracking as a research method

Online tracking is a form of passive data collection that traces information behavior in digital environments. Most tracking approaches assume that participants install data collection software (e.g. a browser plugin) that captures their behavior, either by intercepting web traffic outgoing from the browser or recording user actions visible on the screen (Christner *et al.*, 2021). Depending on software configuration, it can be installed on desktop (Bodó *et al.*, 2017; Haim and Nienierza, 2019; Menchen-Trevino and Karr, 2012; Stier *et al.*, 2020b) or mobile devices (Festic *et al.*, 2021; Krieter, 2019; Reeves *et al.*, 2021; Van Damme *et al.*, 2020) and capture data from the URLs (Stier *et al.*, 2020b) or HTML code of websites visited (Adam *et al.*, 2019) to the content of the mobile applications opened (Krieter, 2019) or the Facebook feeds scrolled by the users (Dvir-Gvirsman, 2020). Often, such software allows specifying additional conditions, for instance, by collecting data from a certain range of websites (Bodó *et al.*, 2017) or filtering out applications dealing with sensitive content (Krieter, 2019).

Several tracking approaches can be distinguished based on the type of device they are applied to and the way in which data are collected (Christner *et al.*, 2021). For desktop-based tracking, transparent proxy and screen-scraping are the two main approaches. The transparent proxy approach uses a virtual proxy that intercepts web traffic and forwards it to the storage server. Examples of such tools include Roxy (Menchen-Trevino and Karr, 2012) and Robin (Bodó *et al.*, 2017). In contrast, the screen-scraping approach extracts HTML content that appears on the device's screen and then sends it to storage. Screen-scraping tools include Eule (Haim and Nienierza, 2019) and Webtrack (Adam *et al.*, 2019).

There are several mobile-based tracking approaches, but only two are currently available for the research community as functioning tools (Christner *et al.*, 2021). The first approach uses smartphone loggers to collect metadata about user activities (e.g. app usage duration), but this approach is usually unable to retrieve the content that users interact with. An example of such a tool is MobileDNA (Van Damme *et al.*, 2020). The commercial alternatives, such as that provided by Wakoopa (Festic *et al.*, 2021), offer more advanced functionality that also allows capturing information about URLs visited but not the actual content viewed. The second approach uses recording apps to take screenshots of the device's screen at a high frequency (e.g. every 5 s; Reeves *et al.*, 2021) or record what is happening there as a video file (Krieter, 2019).

One major advantage of tracking is that it allows the analysis of information behavior in a comprehensive manner (instead of focusing on a few platforms providing access to users' digital traces) and the identification of which information participants are exposed to. It also allows mapping how users navigate to a particular piece of information (i.e. whether they reach it via social media platforms or search engines; Möller *et al.*, 2019) and their news browsing patterns within specific news websites (Vermeer *et al.*, 2020). In addition, online tracking enables studying algorithmic information curation, particularly personalized content delivery, which may create and exploit individual and societal vulnerabilities to increase commercial profits via individualized brand targeting (Bol *et al.*, 2020).

By combining tracking and survey data, researchers can explore the various factors affecting information behavior. These factors vary from the relationship between political attitudes and media exposure in general (Stier *et al.*, 2020b) and the effect of political interest on online news consumption (Möller *et al.*, 2019) to the role of demographic factors in the use of web search (Urman and Makhortykh, 2021) and news apps (Festic *et al.*, 2021).

Despite the multiple benefits provided by tracking in terms of studying information behavior, its use is associated with certain caveats. Because explicit consent is necessary to install tracking software, participants are aware that their behavior is recorded, which may lead to the behavioral changes known as the Hawthorne effect (McCarney *et al.*, 2007). Such an awareness amplifies privacy and security concerns (Keusch *et al.*, 2019a), which are already high considering the sensitivity of online information behavior. Furthermore, the process of installing tracking software requires additional effort on the part of individuals and raises concerns about the detrimental effects of such software on device performance (e.g. increased battery consumption; Krieter, 2019).

Tracking data collection is also demanding in terms of the requirements which have to be met for its successful implementation. Unlike digital trace data (e.g. tweets), which can be retrieved retroactively and often without the user's explicit consent (Salganik, 2018), tracking data are collected in real time and cannot be acquired unless participants explicitly agree to such data collection (Stier *et al.*, 2020a). The real-time mode of tracking data collection also requires resilient backend solutions to enable the uninterrupted transmission of data from participant devices to the storage server and the maintenance of the resulting database.

The above-mentioned technical and ethical issues complicate the process of recruiting participants for tracking-based projects and stress the need to understand participants' concerns and ways to alleviate them. Monetary incentives remain the most common stimulus for participants (Keusch *et al.*, 2019a), although some studies suggest that their effect on participant willingness to partake in tracking is not necessarily significant (Keusch *et al.*, 2019b). Furthermore, monetary incentives require substantial funding, in addition to the already high costs of infrastructure, and raise ethical issues related to the need to estimate the cost of participant privacy.

These complexities prompt interest in alternative incentives, from appeals to the societal relevance of information behavior research (Van Damme *et al.*, 2020) to opportunities for participants to learn about themselves via their data (Sullivan *et al.*, 2019). While such incentives are increasingly used to motivate participants, their effectiveness remains somewhat unclear. Understanding what could motivate participants entails, first, identifying how they perceive online tracking and, second, what specific concerns they have in relation to it and how these concerns can be alleviated.

Perceptions of online tracking research and design features
Due to the lack of studies conducting an in-depth analysis of the interactions between participants' perception of tracking and tracking tools' design features in the context of information behavior research, we considered the three related areas that can provide insights for our study: folk theories of technology, perceptions of online privacy and user-centered design.

Folk theories of technology
Folk theories are intuitive explanations used by people to make sense of complex phenomena. These phenomena include both broad concepts (e.g. privacy and its explanations; Kwasny *et al.*, 2008) and concrete mechanisms (e.g. the functionality of the Facebook newsfeed; Eslami *et al.*, 2016) in fields ranging from nanotechnologies (Rip, 2006) to journalism (Nielsen, 2016). Unlike scientific theories, folk theories are not formalized or empirically tested but, rather,

implicit and often imprecise (Gelman and Legare, 2011). However, folk theories have a substantial influence on individual and collective behavior by determining how a particular phenomenon is understood by the public (Rip, 2006).

The rise of online platforms has prompted interest in intuitive explanations of how digital communication technologies influence information behavior [3]. Considering the fact that these technologies are often perceived as "opaque" boxes (Pasquale, 2015), folk theories influence how these technologies' functionality is interpreted and what concerns they raise among their users. Examples of such concerns vary from privacy (Kwasny *et al.*, 2008) to the lack of user agency (Eslami *et al.*, 2016) and control over algorithmic systems (Harambam *et al.*, 2019) to the threat of manipulation (Toff and Nielsen, 2018).

The importance of folk theories in shaping user expectations and concerns makes them highly relevant in designing tools and approaches to study information behavior. Because folk theories define the popular understanding of how technologies operate, they can influence user choices about what technologies to use and how to use them (Wash, 2010). Consequently, folk theories can influence the choice of research framing (e.g. to address participant concerns; DeVito *et al.*, 2017) or inspire design elements affecting how users perceive the technology (Eslami *et al.*, 2016).

Considering how important folk theories are for participants' understandings of technology and their decisions to accept or reject certain research designs, we propose the following research question:

RQ1. What are the folk theories of tracking in the context of information behavior research?

Online privacy, its perceptions and concerns about it
Defined as the ability of individuals and groups to determine when and how information about them is communicated to others (Westin, 2003), privacy is a key notion in the online ecosystem. It underwent substantial changes following the digital turn as both private corporations and government agencies benefitted from the increased technical capacities to gather and process individuals' data (West, 2019). The growing volume of behavior data available to these external parties amplifies privacy risks and creates the need for more comprehensive approaches to data protection.

The complexity of the concept of online privacy is reflected in the growing number of calls for its contextual (Nissenbaum, 2011) or subjective (Coll, 2014) interpretation. The shift from a more abstract to a more applied understanding of privacy highlights the importance of scrutinizing individual concerns about privacy in different contexts (e.g. news consumption or entertainment). With privacy concerns being a key factor determining the willingness to disclose personal information online (Dinev *et al.*, 2008), information behavior research can benefit from integrating more contextual understandings of privacy into research designs to enable more active and informed participation.

Despite the growing recognition of the importance of privacy for the willingness to be tracked, thus far, privacy has mostly been addressed only from a legal point of view (Kreuter *et al.*, 2020). Consequently, most of the resulting measures do not go beyond traditional data protection procedures (e.g. data anonymization, strict data access, or encrypted data transfer), with a few minor exceptions (e.g. the use of deny lists identifying websites that are not tracked; Bodó *et al.*, 2017). While such measures are undoubtedly important, the degree to which they address participants' concerns about how their privacy is being handled when they consent to participate in tracking-based research is not fully clear.

In an effort to understand the concerns of individuals who may be involved in tracking research, we ask the following research question:

RQ2. What privacy concerns do participants have in relation to the use of tracking to study information behavior?

While we anticipate privacy to be the primary tracking-related concern, we also contemplate the possibility of additional concerns. One example is the (lack of) control that can be related to two dimensions of tracking: awareness of data collection and data usage (Sheehan and Hoy, 2000). The former dimension is often discussed by studies dealing with permission and disclosure (e.g. Cespedes and Smith, 1993; Nowak and Phelps, 1995), which indicate that participant concerns are amplified when they believe that their data are being collected without their awareness. The latter dimension deals with how collected information is presented and whether it can be used to generate additional insights, including the ones concerning participant psychological or sociological profiles (Pridmore, 2008). A number of studies (e.g. Coll, 2014) connect the second dimension to transparency concerns by arguing that non-transparency, in relation to data derivatives (e.g. profiles obtained via data analysis), can limit participant agency.

To investigate these additional concerns and their potential influence on willingness to be involved in tracking, we propose an additional research question:

RQ3. What other concerns do participants have in relation to the use of tracking to study information behavior?

User-centered design
One potential way to address tracking-related concerns is through the adoption of user-centered design (UCD) in the development of tracking tools, particularly because many academic institutions prefer designing their own tracking software despite this being a rather expensive and demanding endeavor [4]. Introduced by Norman and Draper (1986), UCD is a design paradigm that focuses on user interests and the usability of the product. UCD allows developers to better accommodate the needs of end-users, which makes it more likely that designed products will enable sustainable behavioral changes (McCurdie *et al.*, 2012) and can increase trust in the use of products for sharing sensitive information (Veinot *et al.*, 2013). Consequently, UCD is often applied in designing software products, including those used for research purposes (Macaulay, 2009).

The practical implementation of UCD requires product designers to focus on user concerns and needs from the early stages of development. There are multiple ways to identify user attitudes toward a product, but focus groups are most often used (Preece *et al.*, 2015). Focus groups are particularly handy in the case of "niche" products (e.g. tracking software), for which designers have a particular goal in mind (i.e. to collect data on information behavior) but are not yet sure as to which product features are relevant or important. Often, the discussion of such features is facilitated by card sorting [5], a technique that helps to uncover user mental models by identifying the ways in which participants sort and organize concepts (Rosenfeld and Morville, 2006).

Since its introduction, UCD has established its effectiveness for improving product design in multiple domains, including those dealing with information behavior (e.g. Massanari, 2010). The integration of a user perspective via UCD in such domains (e.g. information security) has enabled a shift from paternalistic design approaches to those that explicitly accommodate user perceptions and concerns (Volkamer and Renaud, 2013). Many of these concerns, as noted in the previous subsection, are associated with privacy and data security, which leads to the growing integration of software features allowing individuals to better control their personal information (Nowak and Phelps, 1995).

To our knowledge, there are no studies conducting an in-depth analysis of the use of UCD in the context of tracking research, despite an increasing recognition of the potential of

amending participant concerns via research software design (Bodó *et al.*, 2017; Keusch *et al.*, 2019a). One common approach involves the integration of privacy into tracking tool design by adding mechanisms for filtering sensitive content via deny and allow lists (Bodó *et al.*, 2017) or image recognition techniques used to identify and remove sensitive content before recording it (Krieter, 2019). Another potential way of integrating privacy into tracking software is by providing participants with an option to temporarily disable tracking to give them more control (Adam *et al.*, 2019). These approaches, however, stem from researchers' own mental models and, therefore, must be contrasted with the perceptions of the individuals participating in tracking. Hence, our last research question is as follows:

RQ4. How can the use of UCD amend participant concerns in relation to online tracking research?

Methodology

To answer our research questions, we organized focus groups (i.e. semi-structured discussions aiming at exploring a particular issue) with students from the University of Bern in Switzerland. To recruit participants, we distributed flyers across the campus inviting students to participate in a moderated discussion on the use of personal data for academic research. The participants were incentivized by offering them a chance to participate in the prize raffle as well as a small monetary award. We recruited 13 participants, predominantly BA students, with a few MAs and one PhD student. The majority of participants were between 20 and 30 years old, with a similar proportion of males and females (6 and 7 participants, respectively). The majority of students were from Switzerland, with several exchange students from China, Iran and the US.

The composition of the participant sample was diverse in terms of the students' majors, ranging from Law and Business Administration (four participants) to Social Sciences and Psychology (four participants) to STEM fields (Molecular Biology – two participants, Nutrition – one participant) as well as Management (one participant) and Archeology (one participant). We assumed that covering a broad range of disciplines is advantageous because majors are important factors influencing the understanding that students have of technology and its potential relation to privacy. While students are often treated as a somewhat homogenous population, we argue that this unidimensional treatment can be oversimplified and it is important to take different student backgrounds into consideration.

The participants were divided into three focus groups moderated by two authors. The number of participants (i.e. three to five per group) and the number of focus groups (i.e. three) were chosen according to existing research recommendations. A small group size is recommended for research dealing with complex subjects (e.g. tracking), particularly when researchers must discuss multiple issues to achieve an in-depth understanding of a subject (Krueger, 2014). Similarly, earlier research has demonstrated that three focus groups are usually sufficient to identify between 80 and 90% of the relevant issues (Guest *et al.*, 2017).

Our decision to use focus groups instead of the surveys used by earlier studies on participant concerns about passive data collection (e.g. Boerman *et al.*, 2018) is attributed to several causes. First, focus groups enable more in-depth interactions with participants, which are essential in obtaining more nuanced responses, as compared with surveys, which rely on a small set of response options (Viseu *et al.*, 2004). Unlike surveys, in which the response options are pre-determined, thus constraining and potentially biasing individual replies (Reja *et al.*, 2003), focus groups are semi-structured, which allows taking into consideration emerging themes while sticking to key goals.

Considering that online tracking is a novel research method, we assumed that having a more open data collection structure will be beneficial in our study because it is scarcely

possible to anticipate the complete range of concerns and perceptions that are held by participants. While it could be possible to incorporate open-ended questions into a survey, they allow little to no clarification and are more often omitted by the respondents as compared to closed-ended items. Finally, focus groups provided us with a deeper understanding of the role of participants' subjectivity in the context of tracking by allowing the participants to interact with one another and the moderators to explore both individual and shared perspectives (Morgan, 1996).

Each focus group was 70–80 min long, took place in a university environment and consisted of three sets of questions. The first set of questions focused on the participants' views on tracking as a research technique and its uses for behavior research by academic scholars. The second set dealt with concerns about tracking and the contextual factors influencing these concerns (e.g. differences between desktop and mobile tracking as well as online activities such as dating or gaming). The third set scrutinized design mechanisms (e.g. tracking tool features), which could alleviate participants' concerns.

The participants did not have previous experiences of being involved in online tracking projects in academic environments. Before conducting the focus groups, we did not provide participants with explicit explanations about the mechanism of tracking, to avoid pre-defining their perception of the technique. However, in the course of the focus group, before discussing the second set of questions, we described a particular scenario that involved using online tracking to help participants contextualize their concerns and mechanisms via which to alleviate these. The description of the scenario is provided in the beginning of the second subsection of the Findings section.

To facilitate the discussion of concerns and UCD solutions, we combined open questions with card-sorting tasks. The latter relied on the use of stacks of cards with predefined items (e.g. UCD design features or tracking-related concerns), which participants were asked to select or order depending on specific criteria (i.e. how greatly specific UCD features affected their concerns). The card-sorting approach is frequently used to investigate user mental models in the context of software product design (Paul, 2008). For the majority of tasks, we used an open card-sorting approach, with participants being encouraged to add their own items to the stack of cards or discard the predefined items if they found them unnecessary. The only case in which we used closed-card sorting was the ranking of information resource types according to user concerns about being tracked when using these resources. All card-sorting tasks were performed individually to provide each participant with more chances to express their own opinions.

To process the collected data, we relied on tape-based analysis (Onwuegbuzie *et al.*, 2009), together with notes about the results of card-sorting tasks. As an analytical method, we used deductive content analysis, facilitated by constant comparison analysis. Using a deductive approach, we identified three categories that we were interested in and that followed the three sets of questions used to conduct the focus groups. These categories included perceptions of tracking research (subcategories: participant associations with tracking, actors using tracking and the role of academia in tracking); concerns about tracking research (subcategories: general willingness to participate, general concerns about the use of tracking, concerns about mobile versus desktop tracking, concerns about specific types of content being tracked, effect of tracking period length/monetary incentive); and mechanisms to alleviate participant concerns (subcategories: transparency-based, control-based, privacy-based, exploration-based and other [6] mechanisms). Then, we used constant comparison analysis (Onwuegbuzie *et al.*, 2009) to facilitate the analysis of data within each category and subcategory and compare emerging trends between different focus groups.

Findings

Perceptions of tracking research

We started our analysis by investigating how the concept of tracking is perceived. We found that our participants primarily considered tracking in the context of corporate business and digital advertisements. When asked what they think about when they hear about tracking, participants responded that it is associated with cookies, big data and online platforms (e.g. search engines or social media). Often, participants related the use of tracking to recent scandals, in particular, Cambridge Analytica, and the possibility of meddling in elections. In several cases, participants also noted that tracking is related to individual surveillance and makes them feel insecure. As one participant noted, the possibility of being subjected to online tracking makes them anxious because "someone is recording or seeing what I am doing" (female social science student 2).

An important component of the folk theories of tracking is the perceived identity of actors that use tracking to advance a particular goal. The majority of participants suggested that tracking is primarily used by tech giants, such as Google or Facebook, because it helps them "sell something" (male law student 1) or "predict almost everything" (male psychology student). While the potential of behavior tracking to increase corporate profits was named as a major motivation for using it, several participants mentioned that governments can also use tracking, particularly in the case of countries known for their surveillance efforts, such as the US and China. When asked about government motivations in using tracking, participants suggested that tracking primarily functions as a means of control to tackle terrorist or extremist threats – ("All the governments, they analyze our data. It is not only in the films, not only in the US, not only in China. They are analyzing what people are doing . . . they are analyzing all our conversations, for example, to control criminal activity and terrorism" – female management student) – and also to "put citizens in the box" (male business administration student) by providing them with certain kinds of information.

Unlike corporations and governments, researchers were rarely mentioned among the actors using tracking. When explicitly asked about the use of tracking by academic scholars, participants suggested that it might be of particular interest for disciplines dealing with behavior research, from the social sciences and economics to psychology and media studies – "it's an easy way to get a lot of data", in the words of one of the students (female archeology student). A few participants also noted that tracking might be of interest for the university administration to gain insights into students' background and increase the effectiveness of services by identifying potential bottlenecks. Related to acquiring tracking data for an academic project, participants suggested that it is more effective to take data from big companies, but "it is more ethical to approach a person and ask them to participate" (female social science student 1). The effectiveness argument was related to the assumption that big companies already have large volumes of data, so it is easier to retrieve such data from them.

In general, folk theories of tracking seem to align with common critical narratives about digital technology as a means of increasing profits for corporations and facilitating surveillance. Recent scandals related to personal data abuses, such as Cambridge Analytica or Snowden's disclosures, have a substantial influence on how tracking is perceived by potential project participants. Such circumstances highlight the impact of recent mediatized events on folk theories and the possibility of them amplifying negative feelings toward the use of tracking for studying information behavior. The few connections drawn between tracking and academia offer both opportunities and risks for academic projects relying on the use of this technique.

Concerns about online tracking research
Following the discussion of tracking perceptions, we moved toward scrutinizing the concerns expressed by the participants in relation to their potential involvement in tracking-based research. To do so, we asked our participants to consider a scenario in which the university-affiliated scholars propose that they join a research project tracking their information behavior via desktop and mobile devices. This scenario was informed by the authors' first-hand experience of participating in the development of tracking tools at an academic institution as well as insights into the tool development shared by other research groups (e.g. Bodó *et al.*, 2017).

According to the scenario, in the case of desktop devices, the project requires participants to install a browser extension that captures all browser traffic, except websites on the deny list, and for mobile, it requires the user to install an app to track all mobile traffic except that coming from apps and websites on the deny list. We did not aim to provide a detailed technical explanation of desktop and mobile tracking as part of the scenario explanation, because we assumed it could be tedious for the participants and result in negative group dynamics (e.g. by confusing the participants or making them feel disempowered because of the complicated explanations). However, we offered the participants an opportunity to ask questions about the scenario if they wanted to know more about it.

We started by asking whether the participants were willing to participate in such a project and what motivated their decision. The reactions were rather mixed, with the majority of participants stating that they were not willing to be tracked. The main reason was the participants perceiving their online behavior as an important ("It is your life, what you think, what you do. Even if the university says it is anonymized and everything, I do not trust academics to make it safe" – male psychology student) and highly personal ("It is very uncomfortable, because it is personal. I do everything on my mobile and my laptop." – female social science student 3) part of their lives that they were not willing to expose to the researchers.

Despite the generally negative view on tracking, a few participants also noted that they might be willing to participate in such a project if they would be able to use it for self-actualization purposes ("I would be interested if I will get the results and I would know about my behavior" – female nutrition student) or had an additional incentive ("I would be interested if there is some motivation" – male law student 1). At least three participants expressed a willingness to join the project without additional stimuli and only under the condition of its transparency ("I would do it for free, but I would need to know exactly what you would be looking at" – male business administration student). The importance of transparency also resonated with some other participants, who noted that "it could make a difference" (female social science student 2) if they knew exactly what data were being collected and for what purpose.

In terms of the specific concerns, the examination of which was facilitated by using a card-sorting task, our observations align with the earlier quantitative studies that stress the importance of privacy- and security-related concerns (Keusch *et al.*, 2019a). We found that our participants were primarily worried about privacy both in terms of it being threatened by tracking in general ("Even when completely anonymous, I still do not feel comfortable. I would feel judged even if they [researchers] would not know it is me" – female social science student 3) and the possibility of exposing certain aspects of their online behavior that they want to keep hidden ("Only people with good Internet behavior will participate – others would hesitate" – female social science student 2).

The second common concern was security, particularly how and where the data will be stored and who will be able to access them. One participant, for instance, drew a comparison between data servers and hedge funds by noting that data storage "is a bit like [hedge] funds. Having a fund in the Bahamas is not the same as in Switzerland" (male law student 1).

A similar sentiment was expressed by several other participants, who noted that, for them, it is important that "laws should be in place for data storage" (male business administration student), as well as that they were more concerned about storing data outside the EU ("If I see that the government can basically do whatever it wants with data, then it is a definite no." – male business administration student). Another important security concern was the duration of the data storage period and whether the data would actually be deleted after its expiration ("When data should be deleted, it should be deleted. No data backup, no anything" – male law student 1).

In addition to privacy and security concerns, several participants also noted other sources of anxiety. The lack of an incentive to participate ("Nobody would do it just because" – male law student 1) was mentioned several times, together with concerns about the lack of control over the collected data – ". . . [the main concern] is losing control somehow. It is about what is happening with the data or the drawbacks which are coming with it . . . you need just a good hacker and then the data are gone" (male psychology student).

Some participants were anxious about their data being monetized or manipulated in the research environments ("In the future, my data can be mishandled and I can be blackmailed" – male molecular biology student 1) and questioned whether their participation could somehow compromise their post-tracking life ("You never know what is happening or what is found in the data depending who gets it and then related to that are negative consequences for your future" – male psychology student). Additionally, one participant expressed fear about the tracking process being "addictive" (male law student 1) and the possibility of researchers using engagement techniques to increase the time spent with tracked devices.

After the examination of general concerns, we shifted toward discussing the effect of contextual factors (e.g. the device or information resource type) on the participants' concerns. When asked about being tracked on mobile versus desktop devices, some participants noted that they do not see much difference and do not mind being tracked on both at the same time. Others, however, argued that they "do different things on different devices" and that the mobile behavior is often more personal and raises more concerns about being tracked as compared with the desktop. When asked about what type of mobile apps raise the most concerns, participants rather uniformly noted mobile messengers (e.g. WhatsApp) and banking applications. Other types of apps (e.g. password messengers and insurance) were also mentioned a few times, but most participants did not express concerns about them. Hence, not all participants expressed different concerns about being tracked on mobile as compared to desktop devices and particular concerns about mobile tracking usually arose from the perception of mobile devices as being used for purposes related to more sensitive information.

We also observed substantial variation in the level of concern depending on the content tracked. To do so, we used a card-sorting task, with different types of online resources being listed on the cards. We asked participants to rank these according to the seriousness of concerns they would have about their visits to these resources being tracked. The possibility of tracking interactions with news media, streaming services and online games evoked few worries. By contrast, tracking personal finance, interpersonal communication via messengers or emails and file sharing raised the most concerns. Importantly, social media remained a gray area: while some participants noted that they preferred to exclude social media from tracking, others argued that it would not make a difference, because social media companies extensively track user behavior anyway.

Finally, we asked participants how their concerns were influenced by two aspects of research design, namely the length of the tracking period and the presence of a monetary incentive. Regarding tracking length, most participants noted that it would not affect their concerns substantially ("The resistance is there, whether it is two days or two years" – male psychology student), although a shorter tracking period might be less damaging for their

privacy ("For one week, I just need to get through it. I can behave good" – male law student 2). One concern voiced with regard to longer tracking periods was the reduced control over the data: "the longer you gather my data, the less safe it might be" (female social science student 3). The general consensus was that the length of tracking does not matter much as long as it is less than a few months, with some participants noting that they were willing to be tracked for "up for a year" (female nutrition student) and others noting that they would initially agree to a shorter tracking period but might later agree to prolong it: "I would start with two months. Then, you could send me an email and ask me how I am doing and whether I want to keep going, and I might agree [to do more]" (male law student 2).

In the case of the monetary incentive, most participants noted that it was hard for them to estimate the value of their privacy; hence, monetary compensation would not affect their (un) willingness to be tracked ("For me even the highest monetary incentive would not matter, because basically there is always a risk that it [data] can be misused" – male molecular biology student 1; "If I could consciously decide, I would not give my data to anyone for any reason" – male psychology student) or could only function as a "symbolic contribution" (female social science student 1), indicating that researchers recognize the participants' effort. Some participants also noted that the effectiveness of a monetary incentive depends on the exact volume and type of data being tracked. One participant, for instance, expressed willingness to provide data on their Google search history for 50 Swiss Francs, whereas another noted that they were eager to be tracked for 10 Francs "if nobody would be able to recognize me" (male law student 2). Another participant noted that they would be more concerned about a large monetary incentive because then "you could think that the consequences for you might be stronger" (female social science student 2) and suggested that the incentive should be "not very low and not very high" (female social science student 2).

Mechanisms for alleviating concerns about online tracking research
Finally, we examined UCD mechanisms, which could alleviate participants' concerns and increase their willingness to become involved in tracking research. Following the same scenario used during the discussion of concerns, we asked our participants to think about design features that could be used to address their anxieties about tracking. Specifically, we proposed that participants write down their ideas and then prioritize them depending on these ideas' potential to alleviate their concerns. To facilitate the process, we also provided the participants with sets of cards describing a number of pre-defined mechanisms that were divided into several categories (i.e. control-based, privacy-based, transparency-based and exploration-based). The results of this card-sorting task were used to facilitate the discussion of participant preferences regarding specific UCD mechanisms.

The discussion showed that the participants view mechanisms dealing with transparency to be those with the most potential to increase their willingness to partake in tracking. Almost all participants listed transparency as an important prerequisite for letting anyone track their information behavior. When asked about what exactly should be made transparent, participants usually responded that it was important for them to know "who can see the data and who cannot" (female social science student 2) as well as "how my data are kept safe" (male law student 2) and "the purpose of the project, the method, how you are doing it" (female management student). These three aspects – i.e. project aims, data analysis methods and data access policies – were mentioned particularly often.

In several cases, participants related transparency with accountability by noting that tracking projects should include information such as "what guarantees I have [and] who is liable if it is mishandled" (male molecular biology student 1) or "an address I can contact [in relation to the project]" (male business administration student) [7]. A related suggestion involved the provision of updates about project progress, such as notifications about project

milestones and outputs, but interest in such updates was relatively limited. Finally, the technical details of tracking evoked relatively little interest, with only one participant (male molecular biology student 1) expressing interest in having access to the source code for the tracking software.

The second most popular sets of mechanisms were control- and privacy-based ones. In the case of privacy-based mechanisms, participants were interested in having the ability to switch to a privacy mode whenever "you want to do something you do not want to show" (female social science student 2). Among such mechanisms, the one that attracted the most interest was a privacy button that allows the disabling of tracking for a certain period of time. Because mobile tracking is subject to stronger privacy concerns, we suggest that adding such a button might be particularly beneficial in recruiting participants for mobile tracking studies. A few participants, however, expressed concerns about their ability to detect whether the privacy button was actually working and argued that it might require them "to dig around the source code to see [if it is actually working]" (male molecular biology student 1).

In the case of control mechanisms, the participants had multiple suggestions, varying from relatively simple design features (e.g. an indicator showing whether the user's activity is currently tracked) to more complex solutions (e.g. individualized deny lists or the ability to review and/or modify tracking data before uploading them to the storage server [8]). The simpler solutions, such as the tracking indicator to "know if you are observed or not" (female social science student 2) or a platform to view the collected data and enable "ownership, lack of manipulation and accountability" (male business administration student) over them were the ones the majority of participants preferred.

While some more sophisticated options were suggested (e.g. individualized deny lists), participants expressed relatively little interest in these. Two concerns were often voiced: the first related to these mechanisms being more obscure and, hence, more difficult to understand, and the second related to the potential for these mechanisms to be abused so as to provide incomplete or censored data (". . . [if participants are able to remove data] it would remove the whole point of the thing, it would give falsified data" – male business administration student).

While mechanisms from the last category – i.e. the exploration-based ones – were prioritized only by a few participants, these mechanisms were referred to as valuable additions that can increase participant motivation to engage in tracking. Specifically, participants were interested in being able to explore their data via an individual or comparative tracking dashboard that would give them statistics on their browsing behavior. In the case of an individual dashboard, participants expressed interest in having something "similar to the ScreenTime app, where you see what apps you spend the most time on" (female social science student 1) or "like a resume of what you did, so you can modify your behavior" (female social science student 2).

The comparative tracking dashboard, on which participants can compare their information behavior with the behaviors of other participants, attracted more mixed reactions. While a number of participants expressed interest in it ("it is cool to compare yourself with your peers" – female nutrition student), some also noted that they were not sure they were "comfortable with others seeing my data" (female social science student 1). Altogether, both individual and comparative dashboards were viewed as a chance for participants "to get something out of it [tracking]" (female social science student 3) and for researchers to "create some added value for participants" (male law student 1), which aligns with the suggestions about the importance of self-actualization mechanisms in motivating the participants.

Discussion
In this study, we used focus groups composed of university students to examine perceptions of the use of online tracking for information behavior research. Using a combination of

deductive content analysis and constant comparison analysis, we investigated which folk theories of tracking were present among the participants and how these influenced the perception of tracking as a research method. We also examined the concerns associated with the use of tracking and how these concerns depend on the specific type of content that is tracked, monetary incentive and tracking length. Finally, we scrutinized the various mechanisms that can be used to alleviate these concerns. By doing so, we aimed to achieve a better understanding of the factors influencing the willingness of participants to partake in the tracking research in academic contexts and also highlight ways to better inform the design of potential tracking tools using the UCG approach.

Our observations offer several insights for research relying on tracking to study information behavior. The intuitive explanations of tracking tend to revolve around the concepts of machine learning and big data, together with online surveillance and algorithmic profiling. These concepts are viewed through the prism of mediatized scandals (e.g. Cambridge Analytica), which leads to tracking often being viewed as an instrument employed by corporations to increase their profits (often in unethical ways) and by governments to surveil their citizens. This situation highlights the importance of taking folk theories of research methods into consideration when communicating project goals to potential participants because these theories can affect the willingness to participate in specific types of research.

It is also important to recognize that the substantial presence of these particular folk theories can be attributed to the specifics of the demographic composition of our sample (i.e. young and well-educated individuals from an environment associated with strong leftist and liberal leanings; Hastie, 2007). However, this specific demographic also intensively uses technology to consume information online (Kalogeropoulos, 2019). Hence, despite the lack of generalizability, our findings are informative of the narratives and idiosyncrasies of a concrete group that can be viewed as one of the main subjects of research dealing with online information behavior.

The predominantly negative folk theories of tracking translate into a limited willingness on the part of potential participants to be tracked because of privacy and security concerns, as is also supported by earlier research (Keusch *et al.*, 2019a). These concerns, however, vary depending on the content being tracked, so by limiting tracking to certain activities (in particular, minimizing the capture of data concerning interpersonal communication and personal finances) and adhering to data protection standards, it may be possible to alleviate many participant anxieties. Only a few participants expressed different concerns about being tracked on mobile versus desktop devices. Among those who did, mobile tracking was associated with stronger privacy concerns because mobile devices were perceived as being more personal and used more frequently for transmitting potentially sensitive data (e.g. private messages or online banking). Because the main difference between mobile and desktop tracking perceptions is the strength of privacy concerns, we suggest that mechanisms aimed at protecting participants' privacy, such as adding a private mode button, are particularly important to implement in the context of mobile tracking.

We also found that the length of the tracking period and the presence of a monetary incentive influence the willingness to be tracked to a certain degree, but none of these factors have a decisive effect. The latter observation also corresponds to the findings of Keusch *et al.* (2019b), who noted that a monetary incentive does not have a statistically significant effect on individual decisions to become involved with tracking. The limited impact of monetary incentives points to the relevance of alternative forms of incentivizing that could address participants' intrinsic motivations (e.g. to help them better understand their information behavior; Sullivan *et al.*, 2019).

In terms of research and software designs that can alleviate participant concerns, we find that transparent communication about the methods and aims of data collection is of

paramount importance. The lack of interest in more transparency about technical implementations of tracking (e.g. tracking software source code) eases the task of researchers because the technical aspects are usually the hardest to communicate to non-experts. In addition to transparency-based mechanisms, we found that participants were interested in other simple mechanisms (e.g. a privacy tab or a tracking indicator) that could address participants' privacy- or control-related concerns or, in the case of exploration-based mechanisms (e.g. tracking dashboards), provide additional incentives. These findings highlight the potential for UCD to inform the design of tracking research in a way that reflects user perceptions of privacy and maintains their agency.

Altogether, these findings expand our understanding of the concerns driving individuals away from tracking-based research, as well as ways to alleviate them. Such an understanding is essential not only to facilitate participant recruitment but also to apply tracking in an ethical way and ensure that the benefits to participants outweigh their risks. The growing interest in tracking, both in academia and industry, prompts the need to provide individuals who are tracked with more control over their own data and enable mechanisms that can hold data collectors accountable for potential misuse of collected data (Fuchs, 2011). By taking into consideration the participant perspective, it will be possible to improve methodological standards and ensure that new approaches in behavior research will empower the participants instead of undermining their agency.

As we conclude, it is vital to address the limitations of the study. Our observations are based on a small sample from a specific social group (i.e. university students); thus, they are not necessarily applicable to other groups. At the same time, members of this specific group are particularly active in terms of using digital media and, thus, are often a primary target of research on online information behavior. Nevertheless, future research can benefit from using a larger and more diverse set of participants and, potentially, include the comparative aspect by considering perceptions of tracking in various national contexts and analyzing the potential differences in concerns arising among various groups of participants. Such a follow-up study could provide valuable insights for understanding how folk theories of technology vary between groups (e.g. age, gender and education ones), which remains a rather under-studied subject.

It will also be beneficial to trace interactions with mechanisms that can alleviate tracking concerns in the course of an online experiment. However, such an experiment would require designing and deploying mechanism prototypes, together with simulating an online tracking environment, which will require substantial technical resources and also funding to recruit participants. Similarly, it will be valuable to examine the differences in perceptions of online tracking on different devices, particularly desktop and mobile ones, in more detail.

Overall, our findings illustrate that there is little knowledge about the academic uses of behavioral tracking, even within a population that is closely related to academia. Assuming that the general population is less familiar with behavioral tracking, our study emphasizes the importance of acknowledging non-expert views on tracking and potentially redefining its use in the light of associations that may be transferred from other realms of knowledge about technology, privacy and surveillance. It also stresses the importance of considering perceptions of tracking technologies and their implications in contexts other than academia, for instance, online commerce or law enforcement.

Notes

1. There are multiple projects developing tracking tools for desktop and mobile devices. For some examples, see Roxy (Menchen-Trevino and Karr, 2012), Robin (Bodó *et al.*, 2017; Möller *et al.*, 2019), Webtrack (Adam *et al.*, 2019), FeedVis (Eslami *et al.*, 2015) and AdCollector (Merrill, 2018).

2. The man-in-the-middle-attack is a common form of cyberattack in which the attacker intercepts the web traffic from a certain source and redirects it to a different destination, thus gaining the ability to

view and potentially modify the intercepted traffic and the information transmitted via it (Conti *et al.*, 2016). In the case of online tracking research, the technique allows capturing participant traffic and storing it to identify what content participants interacted with.

3. See, for instance, research on folk theories of online behavioral advertising (Yao *et al.*, 2017) and news distribution (Toff and Nielsen, 2018).

4. In fact, the majority of existing tracking tools are designed by research groups or individual researchers (Menchen-Trevino and Karr, 2012; Bodó *et al.*, 2017; Haim and Nienierza, 2019; Adam *et al.*, 2019; Van Damme *et al.*, 2020; Krieter, 2019). While there are some research groups (Festic *et al.*, 2021; Stier *et al.*, 2020b) that rely on third-party software used by external market companies to collect data, there is a strong leaning toward self-designed tracking solutions within academic tracking research.

5. There are two types of card sorting approaches depending on whether participants can add their own cards to the sorting task (Paul, 2008). Open sorting allows for the soliciting of more information by letting the participants add their own ideas, but it requires more active involvement, whereas closed sorting requires less involvement but can bias the results by focusing on the researchers' ideas. Therefore, we used a combination of both methods.

6. We included this category to account for the possibility of participants introducing mechanisms that would not fit in any of the subcategories we identified in advance.

7. One participant noted that merely providing an address on the website for submitting formal objections was not sufficient, because "it gives you a corporate feeling" (male business administration student). A proposed alternative was to personalize communications by providing the name and contact details of a member of the research team.

8. The latter option is an example of the overlap between different categories of mechanisms. Several participants noted that the ability to explore their data is essential not only for them being able to better understand their information behavior but also to control their data.

References

Adam, S., Maier, M., Aigenseer, V., Urman, A., Christner, C., Makhortykh, M. and Gil-Lopez, T. (2019), "Webtrack – desktop extension for tracking users' browsing behavior using screen-scraping", *Poster Presented at the 5th International Conference on Computational Social Science*, July 17-20, Amsterdam.

Bodó, B., Helberger, N., Irion, K., Zuiderveen Borgesius, F., Möller, J., van de Velde, B., Bol, N., van Es, B. and de Vreese, C. (2017), "Tackling the algorithmic control crisis – the technical, legal, and ethical challenges of research into algorithmic agents", *Yale Journal of Law and Technology*, Vol. 19, pp. 133-180.

Boerman, S., Kruikemeier, S. and Zuiderveen Borgesius, F. (2018), "Exploring motivations for online privacy protection behavior: insights from panel data", *Communication Research*, Vol. 48 No. 7, pp. 953-977.

Bol, N., Strycharz, J., Helberger, N., van de Velde, B. and de Vreese, C. (2020), "Vulnerability in a tracked society: combining tracking and survey data to understand who gets targeted with what content", *New Media and Society*, Vol. 22 No. 11, pp. 1996-2017.

Cespedes, F. and Smith, H. (1993), "Database marketing: new rules for policy and practice", *MIT Sloan Management Review*, Vol. 34 No. 4, pp. 7-22.

Christner, C., Urman, A., Adam, S. and Maier, M. (2021), "Automated tracking approaches for studying online media use: a critical review and recommendations", *Communication Methods and Measures*. doi: 10.1080/19312458.2021.1907841.

Coll, S. (2014), "Power, knowledge, and the subjects of privacy: understanding privacy as the ally of surveillance", *Information, Communication and Society*, Vol. 17 No. 10, pp. 1250-1263.

Conti, M., Dragoni, N. and Lesyk, V. (2016), "A survey of man in the middle attacks", *IEEE Communications Surveys and Tutorials*, Vol. 18 No. 3, pp. 2027-2051.

DeVito, M., Gergle, D. and Birnholtz, J. (2017), "'Algorithms ruin everything': #RIPTwitter, folk theories, and resistance to algorithmic change in social media", *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, pp. 3163-3174.

Dinev, T., Hart, P. and Mullen, M. (2008), "Internet privacy concerns and beliefs about government surveillance – an empirical investigation", *The Journal of Strategic Information Systems*, Vol. 17 No. 3, pp. 214-233.

Dvir-Gvirsman, S. (2020), "Understanding news engagement on social media: a media repertoire approach", *New Media and Society*. doi: 10.1177/1461444820961349.

Eslami, M., Aleyasen, A., Karahalios, K., Hamilton, K. and Sandvig, C. (2015), "FeedVis: a path for exploring news feed curation algorithms", *Proceedings of the 18th ACM Conference on Computer-Supported Cooperative Work*, Association for Computing Machinery, New York, NY, pp. 65-68.

Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K. and Kirlik, A. (2016), "First I 'like' it, then I hide it: folk theories of social feeds", *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, pp. 2371-2382.

Festic, N., Büchi, M. and Latzer, M. (2021), "How long and what for? Tracking a nationally representative sample to quantify internet use", *Journal of Quantitative Description: Digital Media*, Vol. 1 No. 1, pp. 1-23.

Fuchs, C. (2011), "Web 2.0, prosumption, and surveillance", *Surveillance and Society*, Vol. 8 No. 3, pp. 288-309.

Gelman, S. and Legare, C. (2011), "Concepts and folk theories", *Annual Review of Anthropology*, Vol. 40, pp. 379-398.

Guest, G., Namey, E. and McKenna, K. (2017), "How many focus groups are enough? Building an evidence base for nonprobability sample sizes", *Field Methods*, Vol. 29 No. 1, pp. 3-22.

Haim, M. and Nienierza, A. (2019), "Computational observation: challenges and opportunities of automated observation within algorithmically curated media environments using a browser plug-in", *Computational Communication Research*, Vol. 1 No. 1, pp. 79-102.

Harambam, J., Bountouridis, D., Makhortykh, M. and Van Hoboken, J. (2019), "Designing for the better by taking users into account: a qualitative evaluation of user control mechanisms in (news) recommender systems", *Proceedings of the 13th ACM Conference on Recommender Systems*, Association for Computing Machinery, New York, NY, pp. 69-77.

Hastie, B. (2007), "Higher education and sociopolitical orientation: the role of social influence in the liberalisation of students", *European Journal of Psychology of Education*, Vol. 22 No. 3, pp. 259-274.

Jürgens, P., Stark, B. and Magin, M. (2020), "Two half-truths make a whole? On bias in self-reports and tracking data", *Social Science Computer Review*, Vol. 38 No. 5, pp. 600-615.

Kalogeropoulos, A. (2019), "How younger generations consume news differently", *Reuters Institute Digital News Report*, available at: https://www.digitalnewsreport.org/survey/2019/how-younger-generations-consume-news-differently/ (accessed 26 January 2021).

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. and Kreuter, F. (2019a), "Willingness to participate in passive mobile data collection", *Public Opinion Quarterly*, Vol. 83, pp. 210-235.

Keusch, F., Leonard, M., Sajons, C. and Steiner, S. (2019b), "Using smartphone technology for research on refugees: evidence from Germany", *Sociological Methods and Research*, Vol. 50 No. 4, pp. 1863-1894. doi: 10.1177/0049124119852377.

Kreuter, F., Haas, G., Keusch, F., Bähr, S. and Trappmann, M. (2020), "Collecting survey and smartphone sensor data with an app: opportunities and challenges around privacy and informed consent", *Social Science Computer Review*, Vol. 38 No. 5, pp. 533-549.

Krieter, P. (2019), "Can I record your screen? Mobile screen recordings as a long-term data source for user studies", *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, Association for Computing Machinery, New York, NY, pp. 1-10.

Krueger, R. (2014), *Focus Groups: A Practical Guide for Applied Research*, Sage Publications, Thousand Oaks, CA.

Krumpal, I. (2013), "Determinants of social desirability bias in sensitive surveys: a literature review", *Quality and Quantity*, Vol. 47 No. 4, pp. 2025-2047.

Kwasny, M., Caine, K., Rogers, W. and Fisk, A. (2008), "Privacy and technology: folk definitions and perspectives", *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, pp. 3291-3296.

Macaulay, C. (2009), "Usability and user-centered design in scientific software development", *IEEE Software*, Vol. 26 No. 1, pp. 96-102.

Massanari, A. (2010), "Designing for imaginary friends: information architecture, personas and the politics of user-centered design", *New Media and Society*, Vol. 12 No. 3, pp. 401-416.

McCarney, R., Warner, J., Iliffe, S., Van Haselen, R., Griffin, M. and Fisher, P. (2007), "The Hawthorne effect: a randomised, controlled trial", *BMC Medical Research Methodology*, Vol. 7 No. 30, pp. 1-8.

McCurdie, T., Taneva, S., Casselman, M., Yeung, M., McDaniel, C., Ho, W. and Cafazzo, J. (2012), "mHealth consumer apps: the case for user-centered design", *Biomedical Instrumentation and Technology*, Vol. 46 No. 2, pp. 49-56.

Menchen-Trevino, E. and Karr, C. (2012), "Researching real-world Web use with Roxy: collecting observational Web data with informed consent", *Journal of Information Technology and Politics*, Vol. 9 No. 3, pp. 254-268.

Merrill, J.B. (2018), "What we learned from collecting 100,000 targeted Facebook ads", *ProPublica*, 26 December, available at: https://www.propublica.org/article/facebook-political-ad-collector-targeted-ads-what-we-learned (accessed 26 January 2021).

Möller, J., van de Velde, R., Merten, L. and Puschmann, C. (2019), "Explaining online news engagement based on browsing behavior: creatures of habit?", *Social Science Computer Review*, Vol. 38 No. 5, pp. 616-632.

Morgan, D. (1996), *Focus Groups as Qualitative Research*, Sage Publications, Thousand Oaks, CA.

Nielsen, R. (2016), "Folk theories of journalism: the many faces of a local newspaper", *Journalism Studies*, Vol. 17 No. 7, pp. 840-848.

Nissenbaum, H. (2011), "A contextual approach to privacy online", *Daedalus*, Vol. 140 No. 4, pp. 32-48.

Norman, D. and Draper, S. (1986), *User Centered System Design: New Perspectives on Human-Computer Interaction*, LEA, Mahwah, NJ.

Nowak, G. and Phelps, J. (1995), "Direct marketing and the use of individual-level consumer information: determining how and when 'privacy' matters", *Journal of Direct Marketing*, Vol. 9 No. 3, pp. 46-60.

Ochoa, C. and Revilla, M. (2018), "To what extent are members of an online panel willing to share different data types? A conjoint experiment", *Methodological Innovations*, Vol. 11 No. 2, pp. 1-13.

Onwuegbuzie, A., Dickinson, W., Leech, N. and Zoran, A. (2009), "A qualitative framework for collecting and analyzing data in focus group research", *International Journal of Qualitative Methods*, Vol. 8 No. 3, pp. 1-21.

Pasquale, F. (2015), *The Black Box Society*, Harvard University Press, Cambridge, MA.

Paul, C. (2008), "A modified Delphi approach to a new card sorting methodology", *Journal of Usability Studies*, Vol. 4 No. 1, pp. 7-30.

Preece, J., Sharp, H. and Rogers, Y. (2015), *Interaction Design: Beyond Human-Computer Interaction*, John Wiley & Sons, Hoboken, NJ.

Pridmore, J. (2008), *Loyal Subjects? Consumer Surveillance in the Personal Information Economy*, Queen's University, Kingston.

Prior, M. (2013), "The challenge of measuring media exposure: reply to Dilliplane, Goldman, and Mutz", *Political Communication*, Vol. 30, pp. 620-634.

Reeves, B., Ram, N., Robinson, T., Cummings, J., Giles, L., Pan, J., Chiatti, A., Cho, M., Roehrick, K., Yang, X., Gagneja, A., Brinberg, M., Muise, D., Lu, Y., Luo, M., Fitzgerald, A. and Yeykelis, L. (2021), "Screenomics: a framework to capture and analyze personal life experiences and the ways that technology shapes them", *Human-Computer Interaction*, Vol. 36 No. 2, pp. 150-201.

Reja, U., Manfreda, K., Hlebec, V. and Vehovar, V. (2003), "Open-ended vs. close-ended questions in web questionnaires", *Developments in Applied Statistics*, Vol. 19 No. 1, pp. 159-177.

Revilla, M., Couper, M. and Ochoa, C. (2019), "Willingness of online panellists to perform additional tasks", *Methods, Data, Analyses*, Vol. 13 No. 2, pp. 223-252.

Rip, A. (2006), "Folk theories of nanotechnologists", *Science as Culture*, Vol. 15 No. 4, pp. 349-365.

Rosenfeld, L. and Morville, P. (2006), *Information Architecture for the World Wide Web*, O'Reilly Media, Newton, MA.

Salganik, M. (2018), *Bit by Bit: Social Research in the Digital Age*, Princeton University Press, Princeton, NJ.

Sheehan, K. and Hoy, M. (2000), "Dimensions of privacy concern among online consumers", *Journal of Public Policy and Marketing*, Vol. 19 No. 1, pp. 62-73.

Stier, S., Breuer, J., Siegers, P. and Thorson, K. (2020a), "Integrating survey data and digital trace data: key issues in developing an emerging field", *Social Science Computer Review*, Vol. 38 No. 5, pp. 503-516.

Stier, S., Kirkizh, N., Froio, C. and Schroeder, R. (2020b), "Populist attitudes and selective exposure to online news: a cross-country analysis combining web tracking and surveys", *The International Journal of Press/Politics*, Vol. 25 No. 4, pp. 426-446.

Sullivan, E., Bountouridis, D., Harambam, J., Najafian, S., Loecherbach, F., Makhortykh, M., Kelen, D., Wilkinson, D., Graus, D. and Tintarev, N. (2019), "Reading news with a purpose: explaining user profiles for self-actualization", *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, Association for Computing Machinery, New York, NY, pp. 241-245.

Thorson, K. (2020), "Attracting the news: algorithms, platforms, and reframing incidental exposure", *Journalism*, Vol. 21 No. 8, pp. 1067-1082.

Toff, B. and Nielsen, R. (2018), "'I just Google it': folk theories of distributed discovery", *Journal of Communication*, Vol. 68 No. 3, pp. 636-657.

Urman, A. and Makhortykh, M. (2021), "You are how (and where) you search? Comparative analysis of web search behaviour using web tracking data", arXiv [Preprint], available at: https://arxiv.org/abs/2105.04961 (accessed 26 August 2021).

Van Damme, K., Martens, M., Van Leuven, S., Vanden Abeele, M. and De Marez, L. (2020), "Mapping the mobile DNA of news. Understanding incidental and serendipitous mobile news consumption", *Digital Journalism*, Vol. 8 No. 1, pp. 49-68.

Veinot, T., Campbell, T., Kruger, D. and Grodzinski, A. (2013), "A question of trust: user-centered design requirements for an informatics intervention to promote the sexual health of African-American youth", *Journal of the American Medical Informatics Association*, Vol. 20 No. 4, pp. 758-765.

Vermeer, S., Trilling, D., Kruikemeier, S. and de Vreese, C. (2020), "Online news user journeys: the role of social media, news websites, and topics", *Digital Journalism*, Vol. 8 No. 9, pp. 1114-1141.

Viseu, A., Clement, A. and Aspinall, J. (2004), "Situating privacy online: complex perceptions and everyday practices", *Information, Communication and Society*, Vol. 7 No. 1, pp. 92-114.

Volkamer, M. and Renaud, K. (2013), "Mental models – general introduction and review of their application to human-centred security", in Fischlin, M. and Katzenbeisser, S. (Eds), *Number Theory and Cryptography*, Springer, Berlin and Heidelberg, pp. 255-280.

Wash, R. (2010), "Folk models of home computer security", *Proceedings of the Sixth Symposium on Usable Privacy and Security*, Association for Computing Machinery, New York, NY, pp. 1-16.

West, S. (2019), "Data capitalism: redefining the logics of surveillance and privacy", *Business and Society*, Vol. 58 No. 1, pp. 20-41.

Westin, A. (2003), "Social and political dimensions of privacy", *Journal of Social Issues*, Vol. 59 No. 2, pp. 431-453.

Yao, Y., Lo Re, D. and Wang, Y. (2017), "Folk models of online behavioral advertising", *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, Association for Computing Machinery, New York, NY, pp. 1957-1969.

About the authors
Dr. Mykola Makhortykh is a postdoctoral researcher at the University of Bern, where he studies information behavior in online environments. Before moving to Bern, Mykola defended his PhD dissertation at the University of Amsterdam on the relationship between digital platforms and war remembrance in Eastern Europe and worked as a postdoctoral researcher in Data Science at the Amsterdam School of Communication Research, where he investigated the effects of algorithmic biases on digital news consumption. Mykola Makhortykh is the corresponding author and can be contacted at: makhortykhn@yahoo.com

Dr. Aleksandra Urman is a postdoctoral researcher at the Institute of Communication and Media Studies of the University of Bern and Social Computing Group, University of Zurich. Her PhD dissertation defended in May 2020 examines polarization on social media from a comparative perspective. Aleksandra's research interests include online political communication, algorithmic biases and computational research methods.

Dr. Teresa Gil-Lopez (PhD Communication, University of California, Davis, 2019) is a postdoctoral researcher at the Institute for Communication Psychology and Media Pedagogy at the University of Koblenz-Landau, Germany. She writes about protest media coverage and its impact on public perceptions about dissenting groups and the legitimacy of protest as a political tool. She investigates the ways in which digital technologies may have altered the relationships between social movements, the media and the citizen discourse.

Dr. Roberto Ulloa is a postdoctoral researcher at the Computational Social Science department of GESIS – Leibniz Institute for the Social Sciences. His research interests include the role of institutions in polarization and homogenization of the public opinion.