Data privacy concerns as a source of resistance to complete mobile data collection tasks via a smartphone app

Authors:
Dr. Caroline Roberts (corresponding author)
Senior Lecturer and Researcher
Institute for Social Sciences
University of Lausanne
Bâtiment Géopolis, Quartier Mouline
1015 Lausanne
Switzerland
(+41 21 692 38 53)
caroline.roberts@unil.ch

Dr. Jessica M. E. Herzing
Scientific Collaborator
Interfaculty Centre for Educational Research (ICER)
University of Bern, Switzerland
jessica.herzing@icer.unibe.ch

Ms. Jimena Sobrino Piazza
Researcher
Ipsos Public Affairs, Switzerland
Jimena.Sobrino@ipsos.com

Mr. Philip Abbet
Senior Research and Development Engineer
Idiap Research Institute, Switzerland
philip.abbet@idiap.ch

Professor Daniel Gatica-Perez
Adjunct Professor and Head of Social Computing
EPFL, LIDIAP and Idiap Research Institute, Switzerland
gatica@idiap.ch

Publication

This is the original version of the manuscript. A revised version has been published in the Journal of Social Statistics and Methodology: https://academic.oup.com/jssam/article-abstract/10/3/518/6609828

Caroline Roberts, Jessica M E Herzing, Jimena Sobrino Piazza, Philip Abbet, Daniel Gatica-Perez, Data Privacy Concerns as a Source of Resistance to Complete Mobile Data Collection Tasks Via a Smartphone App, *Journal of Survey Statistics and Methodology*, Volume 10, Issue 3, June 2022, Pages 518–548, https://doi.org/10.1093/jssam/smac018

Acknowledgement of funding sources

This work was supported by the EPFL and University of Lausanne 'Collaborative Research on Science and Society' (CROSS) 2019 Program, as part of the project 'Leveraging on-device smartphone inference to address resistance to participate in social surveys (LOIS)'; by FORS, the Swiss Centre of Expertise in the Social Sciences; and by the Faculty of Social and Political Sciences at the University of Lausanne. We would like to thank Oliver Bornet (Idiap) for technical discussions about the mobile platform. We also thank all participants in the study. The study design and analysis was not preregistered.

Statement of significance

Previous research focusing on hypothetical willingness to complete mobile data collection tasks suggests that privacy concerns may represent an important barrier to the successful implementation of app-based surveys. Yet research into people's use of apps and other online services finds it is not always consistent with expressed privacy concerns. Our findings from a study testing a research app in the context of a probability-based general

population survey suggest the influence of privacy concerns may also be weaker when it comes to actual participation decisions in app-based surveys and may be overridden by more proximate considerations about how comfortable participants feel about sharing particular types of data with researchers. We discuss the implications of our findings for the design of future app-based surveys.

1. ABSTRACT

Smartphones present many interesting opportunities for survey research, particularly through the use of mobile data collection applications (apps). There is still much to learn, however, about how to integrate apps in general population surveys. Recent studies investigating hypothetical willingness to complete mobile data collection tasks via an app suggest there may be substantial resistance, in particular, due to concerns around data privacy. There is not much evidence, however, about how privacy concerns influence actual decisions to participate in app-based surveys. Theoretical approaches to understanding privacy concerns and survey participation decisions would suggest that the influence of the former over the latter is likely to vary situationally. In this paper, we present results from a methodological experiment conducted in the context of a three-wave probability-based panel survey of the general population, testing different ways of recruiting participants to an app. At wave 1, half the sample was assigned to a browser-based survey and at wave 2, they were invited to switch to the app. Questions included at wave 1 about online data privacy concerns and comfort sharing different types of data with academic researchers allow us to assess their impact on both hypothetical and actual willingness to download a survey app for completing questionnaires, to take and share photos, and to share the smartphone's GPS location. Our findings confirm those of previous studies, indicating that privacy concerns do influence *hypothetical* willingness to complete mobile data collection tasks, but may be overridden by how comfortable people feel about sharing specific types of data with researchers. When it comes to *actual* compliance with task requests, however, neither privacy concerns nor comfort sharing data were significant. We conclude with recommendations for exploring these relationships further in future app-based studies.

2. INTRODUCTION

The rapid uptake of smartphones during the past decade has fundamentally changed human behavior across multiple domains, transforming not only the subject matter of social research, but also the range of methods and sources of data available. Survey researchers, in particular, have been responding to developments in mobile technology (Link et al. 2014) and are increasingly eager to benefit from the range of data collection opportunities they offer (Jäckle, Gaia, and Benzeval 2017). Increased internet penetration means that web surveys incorporating mobile respondents now offer better coverage rates than ever before (Couper, Antoun and Mavletova 2017) and as more and more people become dependent on smartphones for accessing the internet (Pew Research Center 2019a), optimizing the design of surveys for completion on a mobile and encouraging mobile response have become priorities for survey methodology.

Among the available options for mobile survey optimization, applications (apps) – software installed on smartphones and tablets - are of especial interest because of the possibility to gather multimodal (via built-in device sensors) and in-the-moment data, expanding the research possibilities of traditional survey designs (Link et al. 2014), and potentially offering improved measurement quality, reduced burden and better participant engagement (Struminskaya et al. 2021; Jäckle et al. 2019; Keusch et al. 2019; Elevelt et al. 2019; Wenz, Jäckle and Couper 2019; Toepoel, Lugtig and Schouten 2020; Toepoel and Elevelt 2020). Despite their promise, however, early studies testing the feasibility and utility of integrating apps in surveys have encountered difficulties around people's willingness to use their smartphones to take part in research, whether for so-called 'active' data collection tasks (like completing questionnaires or taking photographs) or consenting to the 'passive'

capture of sensor data like GPS location (Jäckle et al. 2019; Wenz et al. 2019; Keusch et al. 2019; Kreuter et al. 2020; Revilla et al. 2019; Toepoel, Lugtig and Schouten 2020; Struminskaya et al. 2021; Keusch et al. 2021). Three categories of factors have been identified as relevant to understanding the barriers (Wenz et al. 2019; Keusch et al. 2019): (1) task characteristics (e.g. the perceived burden and intrusiveness of the information to be provided); (2) respondent socio-demographic and behavioral characteristics (e.g. relating to smartphone usage); and (3) respondent attitudes, including, among others, concerns relating to data privacy and trust in the security of the data provided.

The use of smartphones has brought to the fore a range of complex ethical challenges relating to the collection and protection of personal data (Bouwman et al. 2013), leaving many uncomfortable about the potential consequences of sharing information online. This has important implications not only for the implementation of mobile web surveys, but also for people's willingness to participate, and consequently, for the accuracy of the data gathered. Privacy concerns have emerged in a number of recent studies as an important driver of hypothetical resistance to take part in mobile data collection tasks in surveys (e.g. Wenz et al. 2019; Keusch et al. 2019; Revilla et al. 2019; Struminskaya et al. 2021), yet there have so far been few attempts to investigate this finding in detail (although see Keusch et al. 2021). As a result, not much is known about the extent and nature of people's privacy concerns, nor about how they vary across different population subgroups. This is of interest because of its implications for survey error, given that privacy concerns correlate with the propensity to respond to a survey and consent to sharing different types of personal information (Couper et al. 2008; 2010), and also with response quality, particularly for sensitive measures (e.g. Rasinski et al. 1999). Because fewer studies have focused on actual willingness to complete mobile data collection tasks, it is also not clear how

prohibitive privacy concerns are likely to be in future app-based studies. The relation between expressed attitudes and actual behavior is notoriously unstable (Crano and Prislin, 2006), and in the digital realm, it is well-established that people's online behaviors and use of different technologies do not always align with their worries about sharing their personal data (Barnes 2006; Kokolakis 2017). Furthermore, privacy concerns represent just one consideration among many that may influence survey participation decisions (Groves et al. 2000), so it is relevant to investigate the conditions under which they exert an influence on actual willingness to complete different types of mobile data collection task, if these new technologies are to be successfully integrated in high quality surveys.

In this article, we present the findings of a probability-based three-wave panel survey of the general population that investigated data privacy concerns as a source of resistance to completing different mobile data collection tasks. The study involved an experiment comparing different recruitment strategies to a data collection app, enabling us to draw conclusions about the relative importance of concerns about online data privacy and comfort sharing different types of data as drivers of both hypothetical and actual willingness to complete three mobile data collection tasks: (1) downloading an app to complete survey questionnaires; (2) taking photographs; and (3) sharing the GPS location of the smartphone. The following research questions are addressed:

- **RQ1:** How prevalent are concerns in the general population about online data privacy and to what extent do they vary systematically as a function of respondent characteristics?
- **RQ2:** How do concerns about online data privacy relate to people's comfort sharing different types of data with academic researchers?

- **RQ3:** To what extent do concerns about online data privacy and comfort sharing different types of data influence hypothetical and actual willingness to download a survey app and complete active and passive mobile data collection tasks?

In the following, we review the literature relating to willingness to participate in mobile data collection tasks and the role of privacy concerns and consider the theoretical reasons why this may be less predictable when it comes to actual compliance with mobile data collection task requests.

2.1 Resistance to participate in mobile data collection and the role of data privacy concerns

While the integration of mobile data collection in surveys and, in particular, the use of apps appears to offer numerous advantages, the potential to capitalize on them depends crucially on people's willingness to download and install the app, and to complete different types of response tasks implied by multimodal data gathering (Revilla et al. 2019; Wenz et al. 2019). Understanding the mechanisms underpinning willingness to participate is, therefore, key. Research conducted to date reveals considerable variation in willingness to complete different mobile data collection tasks, across task type and data collection context. It also highlights some noteworthy correlates of willingness. These studies can be divided into those that have investigated people's *hypothetical* willingness to complete different types of mobile data collection task, and those that have investigated *actual* willingness by attempting to gather data using a mobile app.

Hypothetical willingness

Studies of hypothetical willingness to complete mobile data collection tasks generally finds lower levels of willingness for passive data collection requests than for those requiring the active participation of respondents (Keusch et al., 2021). For example, Revilla and her colleagues (2016) found willingness (among members of an access panel in seven countries) to share GPS position was generally lower than willingness to complete questionnaires, install an application or take photographs on a mobile device. Revilla, Couper and Ochua (2019) found variations in willingness across tasks involving requests to share different types of personal data (not all with mobile technologies). The panelists reported greater comfort with tasks over which they had some level of control over the type of information shared, compared with those involving passive behavior tracking. Wenz, Jäckle and Couper (2019) similarly found variation in willingness across tasks, as a function of the perceived intrusiveness of the data request, as opposed to other task characteristics like the degree of burden placed on the respondent or the mobile device. Willingness was lowest for downloading and using a tracking app (28%), agreeing to GPS tracking (39%), or downloading a survey app to complete questionnaires (47%), and highest for taking and uploading photographs or scanning barcodes (65%) (*ibid.*, p.9).

Besides task characteristics, observational studies have investigated other correlates of hypothetical willingness to complete mobile data collection tasks, finding variation as a function of respondent socio-demographic and behavioral characteristics relating to smartphone use (e.g. device used, frequency of device use, familiarity with and degree of comfort using the device, self-reported smartphone skills, and the number and types of activities they use it for) and attitudinal characteristics (such as attitudes towards surveys in general, trust in the survey sponsor, and concerns about privacy and data security) (e.g. Pinter 2015; Wenz et al. 2019; Revilla et al. 2019; Mulder and de Bruijne, 2019). For

example, Wenz and his colleagues found differences in stated willingness across types of device and as a function of respondent sex, education and frequency of smartphone use. Lower levels of willingness were observed among those expressing higher levels of concern about the security of providing data using mobile devices. Revilla and her colleagues (2019) found that trust in the anonymity of the data was a significant predictor of willingness, while privacy concerns were mentioned in open-ended answers about reasons for unwillingness (Revilla et al. 2019, p. 243). Mulder and de Bruijne (2019) found differences in willingness by panel participants' age and sex, as well as by research topic, with concerns over topic intrusiveness overriding potential gains in the reduction of response burden offered by mobile data collection techniques (p.10).

Keusch and his colleagues (2019) investigated willingness to agree to passive data collection using a research app in a German non-probability panel sample and Struminskaya and her colleagues (2021) investigated willingness to share smartphone sensor data (GPS, camera and wearables) in a Dutch probability-based panel. Both studies used vignette experiments, enabling them to assess the relative importance of different correlates of willingness, while varying features of the hypothetical study design (e.g. sponsor, incentives, framing of the request, degree of control offered over the data collection). Consistent with the findings of the observational studies (see also Couper et al. 2008; 2010; Couper and Singer 2013), privacy and trust considerations were found in both studies to be strong predictors of being unwilling to agree to passive data collection/ share smartphone sensor data and were cited in open-ended answers as reasons for unwillingness (Keusch et al., 2019; Struminskaya et al., 2021). Keusch and his colleagues (2021) investigated concerns about the security of providing different information on a smartphone for research purposes further in four studies involving different samples drawn from

different panels. Across all samples, the highest levels of concern were found for downloading an app that gathers data on how respondents use their phones, followed by GPS tracking. Concerns were lower for frequent users of smartphones and those with lower general privacy concerns and inconsistent effects of age, gender and education were observed across samples and data type (*ibid.*, p.666-70).

Actual participation in app-based studies

As with research on hypothetical willingness to complete mobile data collection tasks, studies that have actually tested apps in surveys have mostly done so in the context of ongoing panel or cohort studies, meaning evidence relating to actual cooperation in newly drawn general population samples is lacking. These studies have integrated app-based data collection for different purposes and with mixed results. For example, McGeeney and Weisel (2015) report results from a diary study, in which a sample of smartphone users in the American Trends Panel were invited to an app version of an experience sampling study; 61% of those invited participated, but app response rates were lower than in the mobile browser-based alternative. Scherpenzeel (2017) reported two app-based studies conducted with smartphone users in the Dutch LISS panel for a mobility study and a time-use study (see also Elevelt, Lugtig, and Toepoel 2019). In both, around 75% of those invited completed the invitation survey, of which 37% reported being willing to participate; 81% of these participated in the mobility study and 68% of these participated in the time-use study, using the app at least once (Scherpenzeel 2017). Participation rates in the time-use study were higher among younger panellists who already owned and were familiar with using a smartphone. Privacy-related concerns predicted earlier expressed willingness to participate in future smartphone studies, but not actual participation when invited (Elevelt et al. 2019). Two other time use studies involving apps (see also Sonck and Fernee 2013)

have been implemented with teenaged participants in the context of ongoing youth cohort studies in the UK (Gilbert, Calderwood and Fitzsimmons 2019) and the Netherlands (Elevit et al. 2019). In both cases, participation rates were encouraging, providing further evidence that younger panel participants may be more comfortable with mobile data collection tasks using an app than older participants.

Jäckle and her colleagues (2019) invited panel members to download a smartphone app for the purpose of scanning and transmitting shopping receipts for a household expenditure study. Of those invited, only 12.8% used the app at least once and 10.2% used the app at least once a week throughout the month-long study period. Reasons for not participating related to lack of time availability and inability to use the technology, and concerns around sharing spending data and data security (p.31). Meanwhile, Kreuter and her colleagues (2020) reported on willingness to download and use an app among members of the German Labour Market and Social Security panel, for the purpose of passive collection of digital trace and sensor data. Only 16% of those panel members invited installed the app (although 91% of those consented to at least one of five passive data collection requests and 71% consented to all five) (p.9). The fact that participants did not change their consent settings suggested they were mostly comfortable about their data privacy when using the app (*ibid.*; p.13). Sugie (2018) reported an app-based study involving GPS tracking and experience sampling among a group of 135 parolees in prison re-entry study, of which 89% participated. Participant trust and privacy concerns were cited as potential barriers to participation but were successfully addressed at the recruitment stage by reassuring participants of the security measures in place to protect their data privacy (p.482).

Few studies have attempted to recruit a freshly drawn sample directly to a study involving app-based data collection, where participation rates might be expected to be lower than among existing, already cooperative panel survey members (Lawes et al., 2021). Miller and her colleagues (2018) tested an app for administering questionnaires to college students in a three-wave panel survey, in which 23% of those invited downloaded the app and consented to participate (p.4), but by wave 3, only 28% of these remained. Smeets, Lugtig and Schouten (2019) report on the use of an app in in the context of a travel survey for Statistics Netherlands designed to collect location data chronologically for every trip taken over the course of a week. Of the sample members invited, around 35% downloaded the app and 27% provided data for at least 7 days (see also Armoogum et al. (2013) and Biler, Šenk and Winklerová (2013) for tests of apps in the context of travel surveys). Meanwhile, Lawes and his colleagues (2021) report participation rates in the app-based German Job Search Panel. Response rates were below 10%, with higher rates of participation among younger participants, females and those with an academic degree (p. 10). To our knowledge, none of these studies investigated privacy concerns as a reason for nonparticipation, thus, there is still much to learn about their role in actual willingness to complete different mobile data collection tasks via an app.

2.2 Understanding public concerns about data privacy and implications for behavior

While accumulating evidence finds privacy concerns to be an important predictor of hypothetical unwillingness to complete mobile data collection tasks in surveys (Keusch et al., 2021), empirical research and theory relating to the nature of public concerns about data privacy suggest their role in actual responses to task requests may not be so clear cut – a factor that may also account for the mixed success of previous app-based studies. Public

opinion research in the US and Europe provides some insights (e.g. Pew Research Centre 2019b; European Commission 2010; 2015), painting a picture of widespread discontent - despite many lacking awareness of some of the privacy risks involved (Tozzi and Coppola 2020). For example, Eurobarometer surveys on the topic of data protection (European Commission, 2010; 2015), revealed high levels of public concern about the (mis-)use of personal data provided online (69%), the level of control people have over their personal data (67%), about the recording of their everyday activities on the internet (45%) and via mobile phone use or mobile applications (55%). These, and other studies, also reveal variation in concerns across population subgroups – notably, age groups (see also Friedewald and Pohoryles 2013; Presthus and Sørum 2018. Regan et al. 2013; Kezer et al. 2016), which, in turn, correlate with technical skills and experience, and frequency and types of internet use.

Despite high levels of expressed concern, however, sharing data is viewed by citizens as a pragmatic response to the demands of the information society and accepted as necessary in order to benefit from online services (European Commission, 2015; Pew Research Center, 2019b). As a result, expressed concerns do not necessarily translate into actions directed at protecting or mitigating risks to personal privacy online, or limiting use of desired services (*ibid.*). This apparent contradiction between reported attitudes and behaviors relating to internet use has been referred to as the 'privacy paradox' (e.g. Barnes 2006; Hargittai and Marwick, 2016; Barth and de Jong 2017; Kokolakis 2017). Widely documented in the context of social networking and e-commerce activities, recent studies suggest the phenomenon also applies to smartphone behavior and the download and use of mobile applications (e.g. Deuker 2010; Zafeiropoulou et al. 2013). Research by Barth and her colleagues (2019) suggests that as with use of social network sites and other online

services, even technically skilled users are willing to risk potential privacy intrusions relating to app use, despite awareness of potential risks. When deciding how to select and whether to download an app, the authors concluded that 'functionality, app design, and costs appeared to outweigh privacy concerns' (p.55), leading people to use apps that involved divulging and relinquishing control over their personal data, despite their apparent concerns about the latter. Thus, privacy concerns per se are not prohibitive of app use across the board.

Several explanations account for the apparent contradiction between people's reported privacy concerns and their actual behavior (see Kokolakis 2017 for a review). Barth and her colleagues (2019) argue that it is a result of rational decision-making processes. People use a risk-benefit analysis to decide which online services to use and how, in which concerns relating to data security may be overridden by other factors, such as the desirability of the app and the rewards to be gained from downloading it. Nevertheless, time constraints (Barth and De Jong 2017), a lack of technical literacy (Liccardi et al., 2014) or apathy (Hargittai and Marwick 2016) may lead users to make use of online services despite having concerns about privacy or security issues. Whatever the explanation, the paradox implies that a more nuanced analysis is needed of how and when privacy concerns and actual behavior concur in the context of actual requests to participate in surveys using a data collection app.

Two other theoretical approaches support the conclusion that the observed correlation between privacy concerns and hypothetical willingness to complete mobile data collection tasks might not hold consistently in practice. Firstly, according to Nissenbaum's (2010) framework for understanding how privacy concerns contribute to resistance to a given

technological innovation (e.g., a new data collection method), we should expect their effect to vary across contexts, as a function of whether and how much the integrity of 'context-relative informational norms' is breached by the innovation in question (p.140). This implies taking into account contextual variables such as the attributes of the technology in question (e.g., demands it places on users and the types of information that will be shared); the roles and activities of different actors involved, and the principles that govern the 'transmission, communication, transfer, distribution and dissemination' of personal information from one party to another (p. 141). Secondly, theories of survey participation, such as the Leverage-Salience Theory (Groves et al. 2000) and Social Exchange Theory (Dillman et al. 2014) highlight the variety of considerations that contribute to the decision to agree to a survey data collection request, and the potential to offset concerns relating to privacy by emphasizing the benefits and rewards to be gained from participation in the survey invitation.

In summary, the preceding review highlights the need to gain a better understanding of the nature of people's concerns about data privacy, about how they vary across population subgroups, and about how they may vary as a function of different features of the data collection context. It also points to a need to research factors that may moderate the influence of privacy concerns in actual participation decisions. The present study addresses some of these research gaps.

3. METHODS

The data come from the three-wave online panel study 'Selects-Civique', carried out in the context of the 2019 Swiss Electoral Studies ('Selects' – see Tresch et al., 2021) in the months prior to and immediately following the federal elections (which took place in

October). A random probability-based sample of 2,183 Swiss adult residents (aged 18 and older) in French-speaking municipalities, was drawn by the Swiss Federal Statistical Office from their SRPH sampling frame based on population registers, which provides some auxiliary sociodemographic data (see Roberts et al. 2013). Selects-Civique was designed to investigate willingness to participate in an election study using a mobile device, and in particular, to download a data collection app to complete all or part of the survey on a smartphone. The sample was randomly assigned in equal parts to two treatment groups. At wave 1 (fielded in May 2019), group 1 was invited to participate in a regular browser-based web survey (programmed in Qualtrics) to be completed on the respondent's device of choice (a QR code in the invitation facilitated mobile response). Group 2 was first invited to download an app ('Civique.org') to their mobile device and to complete the survey within the app (later, reminder letters provided the URL for sample members preferring a browser-based option). Civique.org is a multimodal data collection application (for Android and iOS operating systems) (first developed in 2015 and updated over time by D. Gatica-Perez, J.-I. Biel, O. Bornet, P. Abbet, and D. Santani at Idiap Research Institute, Switzerland), and originally intended as a citizen science platform for mobile data collection initiatives designed to inform local civic causes. Selects-Civique was the first attempt to use the app in the context of a probability-based sample survey of the general population.

At wave 2 (fielded in August 2019), group 1 participants were also invited to download and participate via the Civique app (though the option to complete via the browser remained open). At wave 3 (fielded in October 2019), both groups 1 and 2 were re-invited to use the app (and the browser alternative remained available). The overall (AAPOR RR2) response rate at wave 1 for group 1 was 30.5% and for group 2 was 26.5%. At wave 2, for group 1 it

was 51.3% and for group 2, 50.6%. At wave 3, 59.2% of the group 1 sample participated, while 57.1% of the group 2 participated.

3.1.Measures

As an election study, the questionnaires primarily addressed topics relating to political opinion and behavior, media use and electoral campaign attention. To address the study's methodological motivations, wave 1 included measures relating to internet and smartphone usage, willingness to complete mobile data collection tasks, and privacy concerns – specifically addressing 'online data privacy' and perceptions of the sensitivity of different types of data (some categorized as such under GDPR legislation). In the app, the questionnaires were organized in short thematic modules, which respondents could complete in whatever order they liked. Details of question wording is available in the Appendix.

Concerns about online data privacy: Five measures addressed concerns about online data privacy concerns. Three were general measures. The first asked how *worried* respondents were that websites and apps collect personal information; the second was an attitude statement asking to what extent they *agreed* the Internet poses a threat to privacy; and the third asked how *concerned* they were that their data would go to third parties. The other questions asked how *concerned* respondents were about specific negative consequences of disclosing personal information via websites or apps: including a) that data will be used to send targeted ads; and b) that their identity might be stolen.

Perceptions of the sensitivity of different types of data: Respondents were asked how comfortable they felt about university researchers having access to eight different types of

personal information for academic research purposes, intended to assess perceptions of the sensitivity of different data forms and provide more context-specific measures of privacy concerns relevant to willingness to participate in the Selects-Civique (data gathered in the study are shown in italics). Data types included: *administrative data from population registers about socio-demographic characteristics* (including nationality and country of birth), health data, data on religious beliefs, *data on political opinions*, data from criminal records, data about their sex life, data about income and tax records, and *data about how they use their smartphone or tablet*. Note that the order of the data types was not randomized and so ratings of each may be subject to order of presentation effects (Struminskaya et al. 2021).

Willingness to complete mobile data collection tasks: The wave 1 questionnaire included questions assessing respondents' *hypothetical willingness* to complete different mobile data collection tasks in the context of an academic survey. Here we analyze three relating to data collection tasks respondents were actually asked to complete in Selects-Civique: (1) willingness to download a survey app to respond to questionnaires, (2) to take and share photographs, and (3) to agree to GPS tracking. *Actual willingness* is, therefore, measured by compliance with the task requests among group 1 respondents - installing the app at wave 2, taking photographs (at wave 2), and activating location services within their smartphone's privacy settings to allow the passive capture of GPS coordinates when the app was in use (at wave 3). Consent to the collection of all data was obtained within the app by participants agreeing to a general data confidentiality statement when they first logged in (approved by the EPFL Human Research Ethics Committee, which adheres to Swiss data protection laws).

Respondent Characteristics: Auxiliary data from the register-based sampling frame provides a number of socio-demographic measures (including sex, age, residential area, marital status, and household size). These are supplemented by questionnaire measures of highest educational qualification, main occupational activity, and interest in politics. In addition, the questionnaire included several measures of behavioral and attitudinal characteristics relating to internet and smartphone usage, including frequency of internet and smartphone use; devices used to access internet; number and types of activities respondents use their smartphone for; and operating system. Two subjective measures of internet skills (agreeing that it is exciting to try out new technologies, and that they are capable of solving technical problems when using the internet) were also included.

3.2.Analytic Approach

As a preliminary step to assess the prevalence of concerns about online data privacy and how data privacy concerns vary as a function of respondent characteristics (RQ1), we produced descriptive statistics for the online data privacy measures (presented in Appendix Tables 1 and 2) for different sample subgroups defined by a) socio-demographic characteristics; and b) behavioral characteristics relating to smartphone and internet use. Differences in distributions across categories between subgroups for each of the five data privacy concern measures were tested using Chi-square tests of association (after first verifying there were no device-related measurement differences – see Supplementary Material). The analytic sample included all wave 1 respondents providing complete data for all five of the data privacy measures (n=644 of the total n=687 responding[1]).

[1] There were differences across devices in the proportion of respondents with missing data on the five data privacy measures. A total of 7.6 percent of app respondents skipped questions in this module compared with 4% of PC users (p=0.056). Missing rates for mobile browser respondents were comparable with the app group at 7.5% (significantly different compared with the PC group: $X^2(1) = 3.53$; $p<0.05$).

To address the three research questions, we estimated the parameter coefficients of a series of logistic regression equations predicting the probability of 1) reporting being highly concerned about online data privacy (RQ1); 2) reporting feeling discomfort about university researchers having access to each of the different data types (RQ2); 3) being *hypothetically* willing to a) install a survey app to complete questionnaires, b) take photos (specifically, of political posters during the election campaign), and c) share the GPS location of their smartphone; and 4) (for respondents assigned to group 1) *actually* a) installing the app to participate in wave 2, b) actually taking and uploading (any) photograph in wave 2, and c) in wave 3, activating location services for the app in the phone's privacy settings (to allow the passive capture of its GPS location when the app was in use) (RQ3). The first set of models were fitted for (1) all wave 1 respondents with complete data (n=644); (2) respondents assigned to group 1 only (n=344). The base for second set of models was all wave 1 respondents (n=644), except for comfort sharing data about mobile phone use, which included wave 1 smartphone users only (n=570). The base for the third set of models was group 1 wave 1 respondents with a smartphone (n=289).

Covariates in all models included the socio-demographic and internet usage measures described above. The second set included the composite indicator of data privacy concerns, and the third set introduced at a second step a composite measure of comfort sharing data. Categories of covariates included in the estimation were coded as a series of binary indicators, with the exception of the number of smartphone activities. Full details of coding are available in the Supplementary Material.

4. RESULTS

4.1 Prevalence of and variation in general concerns about online data privacy (RQ1)

Descriptive analyses of the distribution of responses to the five measures indicated a relatively high level of concern among the majority of respondents about online data privacy and the consequences of sharing data online (presented in the Supplementary Material). Only three of the five measures (worry about website and apps collecting personal information and concerns about targeted advertising and identity theft) discriminated between respondent subgroups defined by socio-demographic and Internet usage characteristics, but variation in the proportions of respondents in these groups who reported being 'very concerned' was evident (see Appendix Tables 1, 2 and 3).

For the total sample of wave 1 respondents, three covariates were statistically significant predictors of the probability of reporting a high level of concern about online data privacy: respondent sex, level of education, and interest in politics (see column 1, table 1, which shows beta coefficients and odds ratios (Exponent B) from the logistic regression analyses). Women were significantly more likely to be concerned about data privacy, as were those reporting a higher level of interest in politics. Meanwhile, those with a tertiary level educational qualification were significantly less likely to report being concerned. The effects of three other covariates approached significance ($p<0.1$) and are considered here as small sample sizes likely affected the power of the analyses to detect effects of interest: age, household size and using a smartphone to access the internet. Older people (aged 56 and older) and those living with one other person in the household (compared to those living alone) were more likely to be concerned about data privacy. Finally, using a smartphone to access the internet (compared to not using a smartphone) decreased the likelihood of reporting concern about online data privacy. No other internet usage covariates had statistically significant associations with the probability of being very

concerned about online data privacy. Note that the overall fit of the model was significant ($\chi^2(16) = 41.23$, p=0.001; Hosmer and Lemeshow's test was non-significant ($\chi^2(8) = 6.46$, p=0.596), also indicating good model fit), and, based on Nagelkerke's pseudo R^2, the model accounted for only 9% of the variation in the probability of reporting high levels of concern about data privacy.

Table 1. Logistic regression analyses predicting probability of reporting a high level of concern about online data privacy and the consequences of sharing data online

	(1) All Wave 1 Respondents				(2) Group 1 Wave 1 Respondents			
	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*
Sociodemographic variables:								
Female	.619	**	.195	1.858	1.035	***	.293	2.816
Age[1]: 31-55 years	.351		.273	1.420	.238		.386	1.269
Age: 56+ years	.656	†	.350	1.927	.963	†	.518	2.619
Married	-.014		.246	.986	-.266		.346	.766
Household size[2]: 2 members	.518	†	.312	1.679	.837	†	.436	2.310
Household size: 3 members or more	.291		.306	1.337	.418		.410	1.519
Urban residence	-.133		.212	.876	-.966	**	.343	.381
Tertiary education qualification	-.456	*	.196	.634	-.528	†	.286	.590
Main activity[4]: In paid work	.140		.215	1.150	.360		.304	1.434
Interested in politics	.389	*	.193	1.475	.595	*	.282	1.813
Internet usage variables:								
Uses Internet several times a day	-.247		.250	.781	-.269		.361	.764
Has more than 4 devices	.345		.227	1.412	.321		.338	1.379
Excited to try new devices	-.294		.210	.745	-.192		.319	.825
Able to solve problems with devices	.330		.207	1.391	.639	*	.319	1.894
Uses a smartphone to access internet	-.683	†	.350	.505	-.984	†	.517	.374
Assigned to treatment group 1	.211		.185	1.234				
Constant	.514		.548	1.671	1.024		.765	2.783
Nagelkerke R^2			.090				.172	
Observations			644				344	

Notes. [1]Age (ref. 18-30 years old); [2] Household size (ref. single persons); [3] Highest level of education (ref. All non-tertiary); [4] Main activity (ref. not in paid work). $\hat{\beta}$ = unstandardized beta coefficient; Exp B = Exponent B. p= p-value, † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Focusing on the Group 1 wave 1 respondents only (i.e., those assigned to and responding via the web browser), the results were similar to those for the full response sample, with a

few exceptions (see column 2 of table 1). The significant main effects for sex and interest in politics were comparable, as were those of age, household size and using a smartphone (which approached significance). The effect of having a tertiary level of education was weaker for this subgroup, and only approached significance at the 10% level. One other internet usage covariate was a significant predictor of reporting high levels of concern about online data privacy: the self-report measure of ability to solve problems with internet-enabled devices. Those with greater confidence in their technological skills were more likely to be concerned about data privacy than those with less confidence. Overall model fit was better than for the full sample and was significant ($\chi^2(15) = 42.83$, $p<0.001$). Hosmer and Lemeshow's ($\chi^2(8) = 9.52$, $p=0.301$) test was non-significant, again indicating good model fit, and, based on Nagelkerke's pseudo R^2, the model accounted for 17.2% of the variation in the probability of reporting high levels of concern about data privacy.

4.2 Data privacy concerns and comfort sharing different data types with academic researchers (RQ2)

For all eight data types considered, general concerns about online data privacy were positively and significantly predictive of the probability of reporting being uncomfortable about university researchers having access to personal information (regression coefficients from all 8 models are shown in Appendix Table 4 and in Table 2 for the three data types requested in Selects-Civique). The more concerned people are about online data privacy in general, the more uncomfortable they are about researchers accessing their personal data. However, the effect of the other covariates varied across data type, with overall comfort levels varying accordingly (see Appendix Table 1 for descriptive statistics). The data types for which the highest proportions of respondents reported discomfort were: 1) data about sex life (60.6%); 2) data about income and tax records (56.7%); and 3) health data (56.2%);

and 4) data about criminal records (41.8%). Meanwhile, the data types for which the lowest proportions of respondents reported being uncomfortable about sharing were 1) data on political opinions (26.2%); 2) data about how their smartphone or tablet is used (32.5%), and 3) data on religious beliefs (33.4%) and 4) administrative data from population registers (36.8%).

In the case of the three data types of interest in the present survey, besides general data privacy concerns, the following covariates were also significant predictors of discomfort about sharing. For administrative data (column 1, table 2), those who were interested in politics were significantly less likely to feel uncomfortable sharing, as were those who agreed that they found it exciting to try new technological devices. Those living in an urban area were more likely to feel uncomfortable sharing administrative data ($p<0.1$).

For sharing data on political opinions (column 2), living in an urban area significantly increased the probability of feeling uncomfortable. The effects of interest in politics and being excited to try new devices were weaker than for administrative data ($p<0.1$), but similarly lowered levels of discomfort. In addition, respondents with more than 4 devices were significantly more likely to feel uncomfortable sharing political opinions than those with fewer devices. For sharing data about mobile device use (column 3), the results were similar. Living in an urban area again significantly increased the probability of reporting feeling uncomfortable, as did using multiple devices to access the internet, while being excited to try new technologies decreased the probability of being uncomfortable. Interest in politics was not a significant covariate for this data type.

Table 2. Logistic regression coefficients for models predicting probability of reporting feeling uncomfortable about sharing different data types with University researchers

	(1) Admin Data				(2) Data on political opinions				(3) Data about mobile device use[5]			
	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*
Sociodemographic variables:												
Female	-.171		.181	.843	.152		.195	1.164	-.004		.201	.996
Age[1]: 31-55 years	.095		.273	1.100	-.022		.304	.978	.087		.297	1.091
Age: 56+ years	.028		.333	1.029	.185		.358	1.204	-.169		.360	.845
Married	-.090		.229	.914	.213		.247	1.237	.134		.250	1.143
Household size[2]: 2 members	-.007		.296	.993	-.079		.315	.924	-.079		.327	.924
Household size: 3 members or more	.140		.301	1.150	-.205		.322	.815	-.304		.329	.738
Urban residence	.351	†	.201	1.420	.475	*	.223	1.608	1.003	***	.243	2.727
Tertiary education qualification	-.282		.182	.754	.075		.197	1.078	-.304		.202	.738
Main activity[4]: In paid work	.125		.201	1.133	.182		.218	1.199	.261		.227	1.298
Interested in politics	-.369	*	.186	.691	-.334	†	.200	.716	-.238		.207	.789
Internet usage variables:												
Uses Internet several times a day	.059		.226	1.061	.079		.241	1.082	-.174		.259	.841
Has more than 4 devices	.238		.211	1.269	.629	**	.220	1.875	.404	†	.216	1.498
Excited to try new devices	-.732	***	.206	.481	-.413	†	.221	.661	-.497	*	.222	.608
Able to solve problems with	.208		.196	1.231	.003		.212	1.003	.049		.213	1.050
Uses a smartphone to access	.037		.290	1.037	.178		.316	1.195				
Assigned to treatment group 1	-.031		.176	.969	-.123		.189	.884	-.057		.195	.945
Data privacy concerns (mean)	.738	***	.112	2.092	.557	***	.121	1.746	.795	***	.128	2.214
Constant	-3.310	***	.658	.037	-3.802	***	.721	.022	-4.312	***	.696	.013
Model X^2 (17)	74.97***				49.60***				80.31***			
Hosmer & Lemeshow X^2 (8)	7.17				6.70				9.03			
Nagelkerke R^2	.15				.11				.18			

Notes. [1]Age (ref. 18-30 years old); [2] Household size (ref. single persons); [3] Highest level of education (ref. All non-tertiary); [4] Main activity (ref. not in paid work); [5]Smartphone and tablet users only (n=570). $\hat{\beta}$ = unstandardized beta coefficient; Exp B = Exponent B. p= p-value, † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.3. Data privacy concerns, comfort sharing data, and stated versus actual willingness to participate in mobile data collection (RQ3)

Table 3 shows descriptive statistics for stated and actual willingness to complete the three mobile data collection activities considered among smartphone users in group 1. Stated willingness was highest for downloading an app to complete questionnaires (41.7%) and lowest for taking and sharing photos (23.1%) and sharing GPS location (15.9%). Although the proportion of wave 1 group 1 respondents who actually downloaded the survey app at wave 2 to complete questionnaires (28.3%) was substantially lower than that stating willingness to do so at wave 1, the proportions of the same sample who actually took and shared at least one photo at wave 2 was slightly higher (26.9%); as was the proportion who actually activated location services in the app when prompted to do so in wave 3 (20.3% of the wave 1 group 1 respondents).

Table 3. Stated vs. actual willingness to participate in mobile data collection for a scientific study

	Group 1 Smartphone Respondents	**Hypothetical Willingness**		**Actual Willingness**	
	n	**% (SE)**	***n***	**% (SE)**	***n***
Download a survey app to complete questionnaires	Wave 1: n=290	41.7 (2.9)	121	28.3 (2.7)	82
Take and share photos (of a political poster[1])	Wave 2: n=147	23.1 (2.5)	67	26.9 (2.6)	78
Share GPS location of smartphone	Wave 3: n=141	15.9 (2.2)	46	20.3 (2.4)	59

Notes. Base includes all Wave 1 group 1 respondents with a smartphone. [1]Respondents were asked in W1 about hypothetical willingness to take photos of a political poster but in W2 were asked to take and share 3 photos, one of which was of a political poster or other campaign material, the other of their immediate surroundings while completing the questionnaire – the number who actually took a photo of a political poster was considerably lower than for other photo tasks (n=20).

Finally, we address the question of whether data privacy concerns are more or less influential than comfort sharing specific data types with academic researchers when it comes to

hypothetical and actual willingness to complete mobile data collection tasks. Some general patterns are noteworthy. Firstly, in the case of *hypothetical* willingness to install a survey app (left-hand side of table 4) and stated willingness to share GPS location (left-hand side of table 6), general data privacy concerns are significant predictors, but this effect is no longer significant when the 'comfort' measure is included in the models. In the case of hypothetical willingness to take and share photos, comfort was also a significant predictor, but the general data privacy measure was not (in either model). Secondly, in the case of *actual* willingness (right-hand sides of tables 4, 5 and 6), in all three cases, neither the general data privacy concern measure, nor the measure of comfort sharing relevant data types is a significant predictor of completion of the mobile data collection task. Instead, weak effects are observed in all models for just two covariates a) being aged 31-55 compared to being aged 18-30 (for all three types of data collection, the older respondents are less likely to actually participate); and b) having a smartphone with the Android (or Windows) operating system compared to iOS (again, for all three types of data collection task, being an Android user reduced the probability of actually participating).

Some more specific observations can be made in the case of predictors of *hypothetical* willingness. The more uncomfortable respondents were about sharing data, the less likely they were to report being willing to install a survey app, to take and share photographs or to share the GPS location of their smartphone. In relation to installing an app, as for comfort sharing data, living in an urban area decreased the probability of being hypothetically willing. By contrast, being interested in politics and being female increased the probability of reporting being willing, as did having more advanced technological skills and using a smartphone for a larger number of smartphone activities. In relation to taking photos, being interested in politics also significantly increased the probability of being willing, while being

aged 31 years and older significantly reduced it compared to those aged 18-30 years. Finally, in relation to sharing the GPS location of the smartphone, having a tertiary level qualification reduced the probability of reporting being willing and a weak positive effect (approaching significance) was observed for those agreeing they are excited to try new technologies.

Table 4. Logistic regression coefficients predicting hypothetical vs. actual willingness to install a survey app to complete questionnaires

	Hypothetical willingness to install a survey app (W1)								Actually participated via the survey app (W2)							
	(1)				(2)				(3)				(4)			
	$\hat{\beta}$	p	SE	*Exp B*	$\hat{\beta}$	p	SE	*Exp B*	$\hat{\beta}$	p	SE	*Exp B*	$\hat{\beta}$	p	SE	*Exp B*
Sociodemographic variables:																
Female	.527	†	.297	1.695	.524	†	.308	1.688	.125		.295	1.134	.123		.295	1.131
Age[1]: 31-55 years	.179		.388	1.196	.280		.405	1.323	-.856	*	.388	.425	-.846	*	.390	.429
Age: 56+ years	.426		.524	1.532	.566		.546	1.760	-.547		.517	.579	-.534		.519	.586
Married	.534		.347	1.705	.461		.364	1.585	.173		.355	1.189	.162		.358	1.176
Household size[2]: 2 members	-.204		.444	.815	-.215		.462	.807	.092		.492	1.096	.090		.492	1.094
Household size: 3 members or more	-.574		.432	.563	-.422		.451	.656	.525		.464	1.690	.537		.466	1.710
Urban residence	-.915	**	.331	.401	-.852	*	.346	.426	.012		.329	1.012	.024		.331	1.024
Tertiary education qualification	-.001		.293	.999	-.088		.305	.916	.493		.301	1.638	.485		.302	1.624
Main activity[4]: In paid work	.059		.317	1.060	.095		.332	1.100	.290		.326	1.336	.292		.326	1.339
Interested in politics	.963	**	.311	2.619	.913	**	.322	2.491	-.065		.305	.937	-.071		.306	.931
Internet usage variables:																
Uses Internet several times a day	-.129		.412	.879	-.150		.424	.861	-.646		.457	.524	-.644		.457	.525
Has more than 4 devices	-.109		.317	.897	.027		.333	1.027	-.096		.325	.909	-.088		.326	.916
Excited to try new devices	.453		.312	1.573	.474		.327	1.606	.414		.326	1.514	.410		.326	1.507
Able to solve problems with devices	.710	*	.309	2.034	.759	*	.322	2.137	.206		.311	1.229	.210		.311	1.234
Number of smartphone activities	.309	***	.065	1.361	.307	***	.067	1.359	-.045		.059	.956	-.045		.059	.956
Has an Android phone	-.187		.285	.829	-.210		.298	.811	-.493	†	.290	.611	-.494	†	.290	.610
Data privacy concerns (mean)	-.364	*	.152	.695	-.083		.168	.920	-.162		.148	.851	-.142		.163	.868
Comfort sharing data (mean)	-		-	-	-.658	***	.159	.518	-		-	-	-.042		.145	.959
Constant	-2.398	*	1.048	.091	-1.761	*	1.082	.172	-.307		1.012	.736	-.266		1.022	.766
Model X^2 (17/18)	70.87***				89.75***				27.50†				27.60†			
Hosmer & Lemeshow X^2 (8)	2.37				7.14				5.58				5.65			
Nagelkerke R^2	.29				.36				.13				.13			

Notes. Base includes all Wave 1 group 1 respondents with a smartphone (n=289). [1]Age (ref. 18-30 years old); [2] Household size (ref. single persons); [3] Highest level of education (ref. All non-tertiary); [4] Main activity (ref. not in paid work); $\hat{\beta}$ = unstandardized beta coefficient; Exp B = Exponent B. p= p-value, † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5. Logistic regression coefficients predicting hypothetical vs. actual willingness to take and share photos

	Hypothetical willingness to take photos (W1)								Actually shared a photo (W2)							
	(1)				(2)				(3)				(4)			
	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*
Sociodemographic variables:																
Female	.367		.314	1.443	.396		.321	1.485	.125		.300	1.133	.127		.300	1.135
Age[1]: 31-55 years	-.784	†	.404	.456	-.703	†	.409	.495	-.784	*	.397	.456	-.792	*	.398	.453
Age: 56+ years	-1.415	*	.585	.243	-1.358	*	.597	.257	-.465		.529	.628	-.474		.530	.623
Married	.612		.386	1.843	.500		.395	1.648	.151		.360	1.163	.158		.362	1.172
Household size[2]: 2 members	-.437		.510	.646	-.440		.519	.644	.368		.521	1.445	.369		.521	1.447
Household size: 3 members or more	-.336		.473	.715	-.199		.482	.820	.842		.492	2.320	.834		.494	2.302
Urban residence	-.129		.361	.879	-.020		.367	.980	-.110		.331	.896	-.118		.334	.888
Tertiary education qualification	-.024		.323	.976	-.089		.329	.915	.369		.305	1.446	.375		.306	1.455
Main activity[4]: In paid work	-.303		.340	.739	-.288		.347	.749	.280		.331	1.323	.279		.331	1.322
Interested in politics	1.183	**	.360	3.265	1.150	**	.365	3.159	-.161		.308	.851	-.157		.309	.855
Internet usage variables:																
Uses Internet several times a day	-.188		.491	.829	-.174		.493	.841	-.582		.460	.559	-.583		.461	.558
Has more than 4 devices	-.036		.357	.965	.064		.364	1.066	-.193		.332	.825	-.198		.333	.821
Excited to try new devices	.088		.350	1.092	.089		.356	1.093	.494		.332	1.639	.497		.333	1.644
Able to solve problems with devices	.045		.340	1.046	.072		.349	1.075	.141		.317	1.151	.138		.317	1.148
Number of smartphone activities	.117		.068	1.124	.099		.068	1.104	-.045		.060	.956	-.044		.060	.957
Has an Android phone	-.206		.312	.814	-.218		.316	.804	-.634	*	.298	.530	-.634	*	.298	.531
Data privacy concerns (mean)	-.014		.161	.987	.187		.177	1.206	-.193		.151	.825	-.206		.167	.813
Comfort sharing data (mean)	-		-	-	-.467	**	.165	.627	-		-	-	.029		.148	1.029
Constant	-2.126	†	1.165	.119	-1.607		1.176	.200	-.258		1.031	.773	-.286		1.041	.751
Model X^2 (17/18)	33.35*				41.73**				28.39*				28.43†			
Hosmer & Lemeshow X^2 (8)	5.57				2.92				7.73				6.70			
Nagelkerke R^2	.17				.20				.14				.14			

Notes. [1]Age (ref. 18-30 years old); [2] Household size (ref. single persons); [3] Highest level of education (ref. All non-tertiary); [4] Main activity (ref. not in paid work); $\hat{\beta}$ = unstandardized beta coefficient; Exp B = Exponent B. p= p-value, † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 6. Logistic regression coefficients predicting hypothetical vs. actual willingness to share GPS location

	Hypothetical willingness to share GPS location of smartphone (W1)								Actually activated location services (W3)							
	(1)				(2)				(3)				(4)			
	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*
Sociodemographic variables:																
Female	.343		.361	1.409	.336		.368	1.399	-.067		.322	.935	-.086		.324	.918
Age[1]: 31-55 years	-.179		.467	.836	-.071		.472	.931	-.832	†	.433	.435	-.797	†	.434	.451
Age: 56+ years	-.526		.657	.591	-.422		.671	.656	-.150		.557	.861	-.097		.561	.907
Married	.427		.435	1.533	.324		.444	1.383	.553		.393	1.738	.503		.396	1.653
Household size[2]: 2 members	-.376		.540	.686	-.388		.551	.678	-.048		.536	.953	-.048		.537	.953
Household size: 3 members or more	-.814		.520	.443	-.664		.529	.515	.224		.506	1.251	.279		.509	1.322
Urban residence	-.186		.404	.830	-.111		.409	.895	.198		.368	1.219	.246		.371	1.279
Tertiary education qualification	-.747	*	.362	.474	-.849	*	.370	.428	.549		.336	1.731	.511		.337	1.668
Main activity[4]: In paid work	-.062		.392	.940	-.011		.398	.989	.543		.366	1.721	.558		.368	1.747
Interested in politics	.157		.371	1.170	.108		.377	1.114	-.297		.336	.743	-.332		.338	.717
Internet usage variables:																
Uses Internet several times a day	-.613		.575	.541	-.601		.576	.548	-.482		.489	.618	-.471		.489	.624
Has more than 4 devices	-.027		.387	.973	.057		.395	1.058	-.509		.368	.601	-.480		.371	.619
Excited to try new devices	.725	†	.401	2.065	.685	†	.406	1.983	.134		.358	1.143	.118		.358	1.126
Able to solve problems with devices	.121		.396	1.129	.142		.405	1.153	-.047		.344	.954	-.032		.346	.968
Number of smartphone activities	.028		.074	1.028	.015		.074	1.015	.036		.066	1.037	.033		.065	1.033
Has an Android phone	-.497		.362	.608	-.495		.366	.610	-.547	†	.321	.579	-.558	†	.322	.572
Data privacy concerns (mean)	-.474	**	.175	.622	-.282		.191	.754	-.014		.168	.986	.077		.183	1.080
Comfort sharing data (mean)	-		-	-	-.467	*	.189	.627					-.195		.158	.823
Constant	.582		1.174	1.789	1.119		1.202	3.061	-1.840		1.151	.159	-1.639		1.158	.194
Model X^2 (17/18)	24.65				30.98*				18.81				20.35			
Hosmer & Lemeshow X^2 (8)	8.84				11.73				13.28				5.31			
Nagelkerke R^2	.14				.17				.10				.11			

Notes. [1]Age (ref. 18-30 years old); [2] Household size (ref. single persons); [3] Highest level of education (ref. All non-tertiary); [4] Main activity (ref. not in paid work); $\hat{\beta}$ = unstandardized beta coefficient; Exp B = Exponent B. p= p-value, † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5. DISCUSSION

Recent research into people's hypothetical willingness to complete mobile data collection tasks via a smartphone app suggests that concerns around data privacy and the security of providing data on a mobile device may represent significant barriers to gaining the cooperation of research participants (e.g. Struminskaya et al. 2021; Keusch et al. 2021; Keusch et al. 2019; Revilla et al. 2019; Wenz et al. 2019). There is not much evidence, however, about whether and how privacy concerns affect *actual* decisions to participate in research using apps. Given that people's use of other apps and online services are not always consistent with their expressed concerns about privacy, it is of interest to investigate how prohibitive they really are, especially given the risk they present for survey data quality. We were able to investigate this in a general population app-based panel study, in which participants were invited to install an app, to take and share photographs and activate location services to allow the passive capture of their smartphone's GPS location.

5.1 Summary of main findings

Firstly, with respect to the prevalence of general concerns about online data privacy and the consequences of sharing data online (RQ1), a clear and consistent finding was that the majority of participants in the study reported being concerned about the use of data gathered as a result of their internet use. This fits with the findings of public opinion research (e.g., Pew Research Center 2019b; European Commission 2015). Given our estimates of prevalence are likely to be biased by the fact that those with the strongest concerns about data privacy did not participate in our online survey, we assume that our results underestimate the true prevalence of concern in the target population. Out of the five measures considered, two failed to discriminate at all between subgroups that might be expected to differ as a function of sociodemographic characteristics or internet usage characteristics. Nevertheless, on the

remaining indicators, some variation across population subgroups was evident, giving clues as to the potential risk of bias if privacy concerns translate into non-participation in app-based studies. Women, older people (aged 56 years and older), those living in two person households, and those most interested in politics, had a higher predicted probability of reporting high levels of concern about online data privacy and the consequences of sharing data. Meanwhile, having a tertiary level educational qualification and using a smartphone to access the Internet was associated with a reduced probability of reporting higher levels of concern.

Secondly, we observed considerable variation in comfort levels according to the type of data. Differences in comfort levels were also observed across respondent subgroups. For example, older people were more concerned about sharing their health data, those with a tertiary level education were more likely to be uncomfortable sharing data about criminal records, women were more concerned about sharing data about their sex life, while smartphone users were more concerned about sharing data on income and tax records. For the three data types relevant to the Selects-Civique study, being interested in politics reduced the probability of being uncomfortable sharing admin data, and to a lesser extent data on political opinions, but was not predictive of discomfort sharing smartphone data (though people interested in politics were more likely to report higher levels of concern about data privacy). Living in an urban area increased the probability of being uncomfortable sharing all three types of data (in particular, data on mobile phone use). Meanwhile, being excited to try new technologies reduced the probability of feeling uncomfortable sharing administrative sociodemographic data, data on political opinions and data about mobile phone use - though paradoxically, actually using multiple devices increased the probability of feeling uncomfortable about sharing these data types. These patterns aside, by far the strongest predictor of being

uncomfortable sharing any of the data types considered was respondents' general data privacy concerns, which was significant across the board (RQ2).

Thirdly, in keeping with the findings of other recent studies (e.g. Wenz et al. 2019; Revilla et al. 2019; Keusch et al. 2019; Struminskaya et al. 2021), we found that data privacy concerns were a strong, negative and significant predictor of hypothetical willingness to download an app and agree to passive capture of GPS data. However, privacy concerns were not a significant predictor of willingness to take and share photographs. Furthermore, once the composite measure of comfort sharing data relevant to the study (administrative data, political opinions and mobile phone data) was included in the models, the effect of the privacy measure was negated. Given the correlation between the two composite measures we derived, it is perhaps not surprising that they did not have independent effects on willingness. Nevertheless, given the theoretical interest of considering them separately, the finding is informative. General online data privacy concerns do not appear *per se* to be a barrier to all types of mobile data collection and should be considered as less relevant than the more proximate consideration of how comfortable respondents feel sharing particular types of data requested in a given study (RQ3).

In the models predicting actual compliance with task requests, neither the general data privacy concerns measure nor comfort sharing data with researchers were statistically significant. Instead, just two variables were significant: age and the operating system of the respondent's smartphone. Older participants were significantly less likely to complete the three tasks, as were Android users. The finding relating to age fits with those of other studies that find digital natives to be more at ease with using their smartphones to provide data (Keusch et al. 2021). However, it is not clear why Android users in particular were less willing to take part. We are not aware of any technical reasons why this was the case – it would be important to

investigate this further and especially, possible interactions with privacy concerns (Reinfelder et al. 2014).

The fact that most people reported being concerned but that concerns were not predictive of actual compliance with task requests highlights the normative nature of data privacy attitudes nowadays. Another explanation could be that, lacking a clear understanding of what data are divulged, stored and analyzed as a result of their online activity, people express, when asked, a kind of 'nonattitude' (Converse 1964) based on perceptions of majority opinion (Chung and Rimal, 2016). Such an explanation could account for the so-called 'privacy paradox' (e.g., Barth and De Jong 2017; Kokolakis 2017) that describes the commonly observed disjuncture between expressed privacy concerns and actual online behavior, for which we find some evidence in our data. Measures of general attitudes are not always strong predictors of how people will act (Crano and Prislin 2006) and there is still much to learn about their correspondence in the context of app-based surveys.

5.2 Limitations

Low response rates at wave 1 of this study, combined with the inevitable impact of between-wave attrition, meant that the sample size available for analysis was small, and reduced our power to detect statistical effects and relationships between variables of interest. The sample was further reduced by the need to focus on the group 1 respondents who completed wave 1 on a browser, in order to assess the effect of privacy concerns on hypothetical willingness (measured at wave 1) and actual willingness to use the app at subsequent waves. Furthermore, due to budgetary constraints it was not possible to start with a larger sample or extend the study across all three linguistic regions of the country because of the additional costs involved of fielding the survey in multiple languages (and offering incentives to the

sample members). This limits the generalizability of the study to the wider Swiss population, but the study nevertheless offers an important contribution to the literature on willingness to complete mobile data collection tasks, which for the most part has tested apps in the context of ongoing panel studies, and not always with (probability-based) samples of the general population (Keusch et al. 2019; Kreuter et al. 2020; Lawes et al. 2021). Our study provides rare insight into actual willingness to use a survey app in a freshly drawn random probability sample, also shedding light on optimal recruitment strategies.

Small samples also impinged on our ability to extend the analysis to consider possible moderators of the effect of general data privacy concerns and/ or comfort sharing specific data types, which represent an important avenue for future research. According to the theoretical framework provided by the Leverage-Salience Theory of survey nonresponse (Groves et al., 2000) and Social Exchange Theory (Dillman et al., 2014), data privacy considerations represent just one of a number of salient considerations when responding to a request to participate in an app-based survey. Drawing attention to the positive benefits of responding via an app, and other relevant social norms (like the value of contributing to science – Struminskaya et al. 2021) may help to offset any negative leverage attached to privacy concerns. Measures of what motivated respondents to take part in the survey were included in Selects-Civique, but due to the limited number of observations it was not possible to explore their interaction with the privacy concern measures in any detail. This highlights another general limitation of observational analyses of this kind, as it is hard to disentangle the multiple confounded influences on respondent behavior. Experimental designs offer a clear advantage in this regard, though the ecological validity that is lost when focusing solely on hypothetical willingness to complete mobile data collection tasks may be more problematic for advancing understanding in this field.

5.3 Conclusions

Though the power of our analyses of actual compliance with task requests may have been compromised, the good news from our findings is that neither general privacy concerns, nor variation in the level of comfort felt about sharing requested data appear to be definitive in the decision as to whether to participate in an app-based survey and complete mobile data collection tasks. This lends further support to the conclusion – as the privacy paradox implies – that expressed concerns about data privacy are poor predictors of actual behavior when responding to a survey (task) request. Nevertheless, privacy concerns are clearly part of the story that need to be addressed to help reassure research participants and guarantee the success of future app-based surveys. Given the many benefits to be gained, there is a need for further (ideally experimental) research exploring the role of more context-specific privacy concerns in actual decisions to complete mobile data collection tasks, and how they may be counterbalanced by other survey design features. The multimodal data collection capabilities of apps so attractive to researchers (and hoped also to appeal to respondents) mean that app-based surveys encompass multiple contexts. A more detailed assessment (e.g., following Nissenbaum's (2010) framework) of how data privacy norms relevant to these contexts may be breached by this multimodality may also help in identifying ways to address this challenge in survey methodology.

6. REFERENCES

Armoogum, J., Roux, S., & Pham, T. H. T. (2013). Total nonresponse of a GPS-based travel survey. Paper presented at the conference on New Techniques and Technologies for Statistics, Brussels, Belgium, March 5–7.

Barnes, S. (2006). A privacy paradox: Social networking in the United States. First Monday, 11(9). Retrieved from http://firstmonday.org/article/view/1394/1312

Barth, S., & de Jong, M. D. T. (2017). The privacy paradox – investigating discrepancies between expressed privacy concerns and actual online behavior – a systematic literature review. Telemat. Inf. 34, 1038–1058.

Barth, S., de Jong, M. D. T., Junger, M., Hartel, P. H., & Roppelt, J. C. (2019). Putting the privacy paradox to the test: Online privacy and security behaviours among users with technical knowledge, privacy awareness and financial resources. *Telematics and Informatics*, 41: 55-69.

Biler, S., Šenk, P., & Winklerová, L. (2013). Willingness of individuals to participate in a travel behavior survey using GPS devices. Paper presented at the conference on new techniques and technologies for statistics, Brussels, Belgium, March 5–7.

Bouwman, H., de Reuver, M., Heerschap, N., & Verkasalo, H. (2013). Opportunities and problems with automated data collection via smartphones. *Mobile Media and Communication*, 1(1): 63-68. DOI: https://doi.org/10.1177/2050157912464492

Chung, A., & Rimal, R. N. (2016). Social norms: a review. *Review of Communication Research,* 4, 1–29. doi: 10.12840/issn.2255-4165.2016.04.01.008

Converse, P. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), Ideology and discontent. New York: Free Press.

Couper, M. P., Antoun, C., & Mavletova, A. (2017). Mobile web surveys. In P.P. Biemer, E. Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, & B. T. West (Eds.), *Total survey error in practice* (pp. 133-154). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781119041702.ch7

Couper, M. P., & Singer, E. (2013). "Informed Consent for Web Paradata Use." *Survey Research Methods* 7:57–67.

Couper, M. P., Singer, E., Conrad, F. G., & Groves, R. M. (2008). "Risk of Disclosure, Perceptions of Risk, and Concerns about Privacy and Confidentiality as Factors in Survey Participation." *Journal of Official Statistics* 24:255–75.

Couper, M. P., Singer, E., Conrad, F. G., & Groves, R. M. (2010). "Experimental Studies of Disclosure Risk, Disclosure Harm, Topic Sensitivity, and Survey Participation." *Journal of Official Statistics* 26:287–300.

Crano, W. D., & Prislin, R. (2006). *Attitudes and Persuasion. Annual Review of Psychology, 57(1), 345–374.* doi:10.1146/annurev.psych.57.102904.190034

Deuker, A. (2010). Addressing the privacy paradox by expanded privacy awareness – the example of context-aware services. In Bezzi, M., Duquenoy, P., Fischer-Hüber, S., Hansen, M., Zhang, G. (Eds.), *Privacy and Identity Management for Life*. Springer-Verlag, Berlin, Heidelberg, pp. 275–283.

Dillman, D., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method.* Hoboken, NJ: John Wiley & Sons.

Elevelt, A., Bernasco, W., Lugtig, P., Ruiter, S., & Toepoel, V. (2019). Where you at? Using GPS location in an electronic Time Use diary study to derive functional locations. *Social Science Computer Review, 1-18*. DOI:https://doi.org/10.1177/0894439319877872

Elevelt, A., Lugtig, P., & Toepoel, V. (2019). Doing a Time Use Survey on Smartphones Only: What Factors Predict Nonresponse at Different Stages of the Survey Process?. *Survey Research Methods*, *13*(2), 195-213. DOI: https://doi.org/10.18148/srm/2019.v13i2.7385

European Commission (2011). *TNS Opinion & Social: Special Eurobarometer 359 / Wave 74.3: Attitudes on Data Protection and Electronic Identity in the European Union.* Survey requested by the European Commission Directorate General Information Society and Media (INFSO), the Directorate-General Justice (JUST) and the Directorate-General JRC and co-ordinated by the Directorate-General Communication ('Research and Speechwriting' Unit)., Brussels, June 2011.

European Commission (2015). Special Eurobarometer 431. *Data protection – Report*. Survey conducted by TNS opinion & social at the request of the European Commission, Directorate-General for Justice and Consumers (DG JUST). Survey coordinated by the Directorate-General Communication ´Strategy, Corporate Communication Actions and Eurobarometer´ Unit. Brussels, June 2015. dx.doi.org/10.2838/552336.

Friedewald, M., & Pohoryles, R. J. (2013). Technology and privacy, *Innovation: The European Journal of Social Science Research*, 26:1-2, 1-6, DOI: https://10.1080/13511610.2013.768011

Gilbert, E., Calderwood, L., & Fitzsimmons, E. (2019). *Does mode matter? The impact of using a mixed mode time use diary on data quality and measurement in the Age 14 Survey of the Millennium Cohort Study.* London: Centre for Longitudinal Studies, Institute of Education.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation. *Public Opinion Quarterly, 64*(3), 299–308. DOI: https://doi.org/10.1086/317990

Hargittai, E., & Marwick, A. (2016). "What can I really do?" Explaining the Privacy Paradox with online apathy. *International Journal of Communication*, 10: 3737-3757.

Jäckle, A., Burton, J., Couper, M. P., & Lessof, C. (2019). "Participation in a Mobile App Survey to Collect Expenditure Data as Part of a Large-Scale Probability Household Panel: Coverage and Participation Rates and Biases." Survey Research Methods 13:23–44.

Jäckle, A., Gaia, A., & Benzeval, M. (2017). *The Use of New Technologies to Measure Socio-economic and Environmental Concepts in Longitudinal Studies*. Colchester: University of Essex.

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). "Willingness to Participate in Passive Mobile Data Collection." Public Opinion Quarterly 83:210–35.

Keusch, F., Struminskaya, B., Kreuter, F., & Weichbold, M. (2021). "Combining Active and Passive Mobile Data Collection: A Survey of Concerns." In *Big Data Meets Survey Science*, edited by Craig A. Hill, Paul P. Biemer, Trent Buskirk, Lilli Japec, Antje Kirchner, Stas Kolenikov, and Lars E. Lyberg, 657–82. Hoboken, NJ: Wiley.

Kezer, M., Sevi, B., Cemalcilar, Z., & Baruh, L. (2016). Age differences in privacy attitudes, literacy and privacy management on Facebook. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 10(1), article 2. doi: 10.5817/CP2016-1-2

Kokolakis, S. (2017). "Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon." *Computers & Security*, 64, 122–134. doi:10.1016/j.cose.2015.07.002.

Kreuter, F., Haas, G.-C., Keusch, F., Bähr, S., & Trappman, M. (2020). "Collecting Survey and Smartphone Sensor Data with an App: Opportunities and Challenges around Privacy and Informed Consent." *Social Science Computer Review* 38: 533–49.

Lawes, M., Hetschko, C., Sakshaug, J. W., & Griessemer, S. (2021). Contact modes and participation in app-based smartphone surveys: Evidence from a large-scale experiment. *Social Science Computer Review*, 1-18. DOI: 10.1177/0894439321993832

Liccardi, I., Pato, J., Weitzner, D. J., Abelson, H., & de Roure, D. (2014). No technical understanding required: Helping users make informed choices about access to their personal data. *MOBIQUITOUS*, 2014, London, Great Britain, 140-150.

Link, M. W., Murphy, J., Schober, M. F., Buskirk, T. D., Hunter Childs, J., & Langer Tesfaye, C. (2014). "Mobile Technologies for Conducting, Augmenting and Potentially Replacing Surveys: Executive Summary of the AAPOR Task Force on Emerging Technologies in Public Opinion Research." *Public Opinion Quarterly* 78:779–87.

McGeeney, K., & Weisel, R. (2015). App vs. web for surveys of smartphone users experimenting with mobile apps for signal-contingent experience sampling method surveys, Pew Research Center report, Washington, DC. http://www.pewresearch.org/2015/04/01/app-vs-web-for-surveys-of- smartphone-users/ (accessed May 6, 2016).

Miller, Y., DiCiccio, C., Lavista, J., Gore-Felton, C., Acle, C., Hancock, J., Richardson, A., Nelson, L., Palesh, O., & Oakley-Girvan, I. (2018). Smart(phone) approaches to mobile app data collection. *Survey Practice*, 11(2): 1-23. DOI: https://10.29115/SP-2018-0030

Mulder, J., & de Bruijne, M. (2019). Willingness of online respondents to participate in alternative modes. Survey Practice, 12(1): 1-11. DOI: https://10.29115/SP-2019-0001

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford University Press.

Pew Research Center. (2019a). “Mobile Technology and Home Broadband 2019”. Retrieved from https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/06/PI_2019.06.13_Mobile-Technology-and-Home-Broadband_FINAL2.pdf

Pew Research Center. (2019b). *“Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information”.* Retrieved from https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/11/Pew-Research-Center_PI_2019.11.15_Privacy_FINAL.pdf

Pinter, R. (2015). “Willingness of Online Access Panel Members to Participate in Smartphone Application-Based Research.” In *Mobile Research Methods*, edited by Daniele Toninelli, Robert Pinter, and Pablo de Pedraza, 141–56. London: Ubiquity Press.

Presthus, W., & Sørum, H. (2018). Are consumers concerned about privacy? An online survey emphasizing the General Data Protection Regulation. *Procedia Computer Science*, 138: 603-611.

Rasinski, K. A., Willis, G. B., Baldwin, A. K., Yeh, W. C., & Lee, L. (1999). Methods of data collection, perceptions of risks and losses, and motivation give truthful answers to sensitive survey questions. Appl. Cogn. Psychol. 13, 465–484.

Regan, P. M., Fitzgerald, G., & Balint, P. (2013). Generational views of information privacy. *Innovation: The European Journal of Social Science Research*, 26(1-2): 81-99. DOI: https://doi.org/10.1080/13511610.2013.747650

Reinfelder, L., & Gassmann, F. (2014). Differences between Android and iPhone Users in Their Security and Privacy Awareness. In C. Eckert, S. K. Katsikas, G. Pernul (Eds.), *Trust, Privacy, and Security in Digital Business. 11th International Conference*, TrustBus 2014, Munich, Germany, September 2–3, 2014 Proceedings (pp.156-167). Publisher: Springer. DOI: https://doi.org/10.1007/978-3-319-09770-1

Revilla, M., Couper, M. P., & Ochoa, C. (2019). "Willingness of Online Panelists to Perform Additional Tasks." *Methods, Data, Analyses*, 13:223–52.

Revilla, M., Toninelli, D., Ochoa, C., & Loewe, G. (2016). "Do Online Access Panels Need to Adapt Surveys for Mobile Devices?" Internet Research 26:1209–27.

Roberts, C., Lipps, O., & Kissau, K. (2013). Using the Swiss population register for research into survey methodology. *FORS Working Paper Series*, 2013-01. Retrieved from https://forscenter.ch/wp-content/uploads/2018/07/fors_wps_2013-01_roberts.pdf

Scherpenzeel, A. (2017). "Mixing Online Panel Data Collection with Innovative Methods." In *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung* [Methodological Problems of Mixed-Mode Approaches in Survey Research], edited by Stefanie Eifler and Frank Faulbaum, 27–49. Wiesbaden: Springer.

Smeets, L. M. S., Lugtig, P., & Schouten, J. G. (2019). Automatic Travel Mode Prediction in a National Travel Survey. *CBS Discussion Paper*.

Sonck, N., & Fernee, H. (2013). *Using Smartphones in Survey Research: A Multifunctional Tool. Implementation of a Time Use App. A Feasibility Study*. The Hague, Netherlands: The Netherlands Institute for Social Research. Retrieved from http://www.websm.org/uploadi/editor/1388746254Sonck_Fernee_2013_Using_smartphones_in_survey_research.pdf

Struminskaya, B., Toepoel, V., Lutgtig, P., Haan, M., Luiten, A., & Schouten, B. (2021). Understanding willingness to share smartphone-sensor data. *Public Opinion Quarterly*, DOI:https://doi.org/10.1093/poq/nfaa044

Sugie, N. F. (2018). "Utilizing Smartphones to Study Disadvantaged and Hard-to-Reach Groups." *Sociological Methods & Research* 47(3):458–91.

Toepoel, V., & Elevelt, A. (2020). Mobile and Sensor Data Collection. In *Mobile and Sensor Data collection* (SAGE Research Methods Foundations). DOI:https://doi.org/10.4135/9781526421036913576

Toepoel, V., Lugtig, P. J., & Schouten, J. G. (2020). Active and Passive Measurement in Mobile Surveys. *The Survey Statistician*, *82*, 14-26.

Tozzi, A., & Coppola, G. (2020). *Digital footprint awareness. A European survey to analyse EU citizens' understanding of digital footprint*. Milan: European Digital Learning Network. ISBN: 9788894501544

Tresch, A., Lauener, L., Bernhard, L., & Scaperrotta, L. (2021). *Selects: Panel Survey (waves 1-4) - 2019-2020 [Dataset].* Distributed by FORS, Lausanne, 2021. www.selects.ch https://doi.org/10.23662/FORS-DS-1184-2

Wenz, A., Jäckle, A., & Couper, M. P. (2019). "Willingness to Use Mobile Technologies for Data Collection in a Probability Household Panel." Survey Research Methods 13:1–22.

Zafeiropoulou, A. M., Millard, D. E., Webber, C., & O'Hara, K. (2013). *Unpicking the privacy paradox: can structuration theory help to explain location-based privacy decision?* New Your, USA In: WebSci '13 Proceeding of the 5th Annual ACM Web Science Conference, pp. 463–472.

7. APPENDIX

Question Wording

Concerns about online data privacy:

- How concerned are you by the fact that websites and apps collect your personal information?
    - (Not at all, a little, moderately, very, extremely)
- To what extent do you agree or disagree that the Internet poses a threat to privacy?
    - (Agree strongly, agree, neither agree nor disagree, disagree, disagree strongly)
- When you use websites and/ or apps, how concerned are you…
- that your data will be shared with third parties without your permission?
- that your data will be used to send you targeted advertising?
- that your identity could be stolen online?
    - (Not at all, a little, moderately, very, extremely)

Perceptions of the sensitivity of different types of data:

- To what extent do you feel comfortable with the idea of university researchers having access to the following personal information about you?
- Data from the local authority (e.g., your name, address, sex and date of birth)
- Data about your health
- Data about your religious beliefs
- Data about your political opinions
- Data relating to your criminal records
- Data about your sex life
- Data about your income and your tax records

- Data about how you use your smartphone or tablet
    - (Completely comfortable, quite comfortable, moderately comfortable, not very comfortable, not at all comfortable)

***Willingness to complete mobile data collection tasks*:**

- How willing would you be to download an application on your mobile phone to fill out a questionnaire for a scientific study?
- How willing would you be to take photos of political posters in your community and share them with researchers for a scientific study?
- How willing would you be to share the GPS position of your smartphone for a scientific study?
    - (Completely willing, mostly willing, mostly not willing, not willing at all)

Interest in politics:

- In general, how interested are you in politics?
    - (Very interested, mostly interested, mostly not interested, not at all interested)

Frequency of internet and smartphone use:

- How often do you use the internet for personal purposes?
- How often do you use your smartphone for activities besides phone calls and texts?
    - (Several times an hour, several times a day, once a day, several times a week, several times a month, once a month or less)

Devices used to access internet:

- Which of the following devices do you use to connect to the internet? (Check all).
    - (Desktop computer, laptop computer, smartphone, tablet, basic mobile phone, e-reader, smart watch, other)

Types of activities respondents use their smartphone for:

- Do you use your smartphone for the following activities? (Yes/ no)
    - Consulting web sites
    - Writing or reading emails
    - Taking photos
    - Consulting social media content (e.g. on Facebook, Twitter, Instagram, Snapchat)
    - Posting content on social media (e.g. on Facebook, Twitter, Instagram, Snapchat)
    - Making purchases (e.g. reserving train tickets, buying clothes, ordering food)
    - For banking transactions (e.g. consulting the balance of your account, transferring money),
    - To install new applications (e.g. from ITunes or the Google Play Store)
    - To use geo-localization/ GPS applications (e.g. Google Maps, Foursquare, Yelp)
    - To connect to other electronic devices via Bluetooth (e.g. smart watches, fitness devices)
    - To play games
    - To watch videos or listen to music;
    - Other activities

Internet skills:

- To what extent do you agree or disagree with the following statements?
- It is exciting for me to try new technologies and devices.

- I am able to resolve problems with devices if they arise when using the internet
    - (Agree strongly, agree, neither agree nor disagree, disagree, disagree strongly)

Data privacy concerns as a source of resistance to complete mobile data collection tasks via a smartphone app

Supplementary Material

1. Methods: Coding of variables in logistic regressions

Variables	Coding
Sociodemographic variables from the sampling frame:	
Respondent sex	Female (1) Male (0)
Age group	18-30 (0) 31-55 (1) 56 years and older (1)
Marital status	Single, never married, divorced, widowed or separated (0) Married (1)
Household size	Single person household (0) 2 members (1) 3 or more members (1)
Urbanicity of residential area	Rural (0) Urban (1)
Self-report measures of respondent characteristics:	
Tertiary level education	Has a tertiary level educational qualification (1) Completed secondary-level or equivalent qualification or less (0)
Main occupational activity	In full-time or part-time paid work (1) Student/ apprentice/ in training; not in paid work (retired, unemployed, home-maker) (0)
Interest in politics	Very or somewhat interested in politics (1) Not at all or rather not interested (0)[2].
Internet usage variables:	
Frequency of internet use	Uses internet less than once a day (1) Uses internet more than once a day (0)
Has more than 4 devices	Accesses internet from more than 4 devices (1) Fewer (0)
Excited to try new devices	Agrees strongly or agrees it is exciting for me to try new technologies and devices (1) Neither agrees nor disagrees, disagrees, disagrees strongly (0)
Able to solve problems with devices	Agrees strongly or agrees I am able to resolve problems with devices if they arise when using the internet (1) Neither agrees nor disagrees, disagrees, disagrees strongly (0)
Uses a smartphone to access internet	Uses a smartphone (1) Does not use a smartphone (0)
Number of smartphone activities	Count variable giving total out of 13 activities checked
Has an Android phone	Uses Android or Windows operating system (1) Uses iOS (0)
Assigned to treatment group 1	Assigned to treatment group 1 (1) Assigned to treatment group 0 (0)

Data privacy concerns (mean)	Mean of four measures: • How concerned are you by the fact that websites and apps collect your personal information? • When you use websites and/ or apps, how concerned are you… - that your data will be shared with third parties without your permission? - that your data will be used to send you targeted advertising? - that your identity could be stolen online? • Not at all, (2) a little, (3) moderately, (4) very, (5) extremely
Data privacy concerns (dichotomized)	Mean as above but dichotomized: scores 3.5 and greater = 1 (concerned), else=0 (less concerned)
Comfort sharing data	Mean of three measures: • To what extent do you feel comfortable with the idea of university researchers having access to the following personal information about you? - Data from the local authority (e.g., your name, address, sex and date of birth) - Data about your political opinions - Data about how you use your smartphone or tablet (1) Completely comfortable, (2) quite comfortable, (3) moderately comfortable, (4) not very comfortable, (5) not at all comfortable

2. Methods: Preliminary Analyses

Despite the random assignment of sample members to the two treatment groups, respondents could choose which device/ software to use to respond (in group 1, between a PC or (at a later follow-up stage) mobile browser; and in group 2, between the app and a PC or mobile browser). Because characteristics of the response device could potentially affect how respondents answer questions, we first assessed evidence for measurement differences. Comparisons across device groups were weighted by a propensity score weight controlling for observed differences in the composition of the samples responding using different devices and software (based on a logistic regression model predicting the probability of responding with one device compared to its alternative, given a number of observed characteristics (the frame variables described above). There were differences across devices in the proportion of respondents with missing data on the five data privacy measures. A total of 7.6 percent of app respondents skipped questions in this module compared with 4% of PC users (p=0.056).

Missing rates for mobile browser respondents were comparable with the app group at 7.5% (significantly different compared with the PC group: $X^2(1) = 3.53$; $p<0.05$). No other differences were observed.

To assess systematic variations in online data privacy concerns across the respondent characteristics of interest (RQ1), the dependent variable was derived from a composite measure based on the four items that discriminated best between sample subgroups, including the general measure of concern that websites and apps collect personal information and that personal information will be shared with third parties, and the two more specific measures relating to targeted advertising and identity theft. The mean score (ranging from 1-5) for the four items was dichotomized, with those scoring greater than 3.5 coded as 1, and those scoring below 3.5 coded as 0).

To investigate how people's general concerns about online data privacy relate to their level of comfort sharing different types of data with academic researchers (RQ2), the dependent variables were dichotomised such that those reporting they were hardly or not at all comfortable with researchers accessing their personal information coded 1 and those who felt more comfortable coded 0). The covariates were the same as for the previous analyses, plus the composite measure of data privacy concerns - the mean score ranging from 1-5 (not dichotomized).

Finally, to address the question of whether and how privacy concerns and comfort sharing personal data with academic researchers influence willingness to complete mobile data collection tasks (RQ3), the dependent variables were binary indicators of stated willingness (coded 1 if the respondent reported being completely or somewhat willing to complete the task, and 0 if not).

The same covariates were included as for the previous analyses, with the addition – at a second step - of a composite measure of respondents' overall comfort with academic researchers having access to their administrative data, data about their political opinions and data about how they use their smartphone (the three data types most relevant to the survey). This was computed as the mean of the three comfort measures, yielding a score from 1-5, where 5 indicated being not at all comfortable about sharing the data types requested in the Selects-Civique study.

3. Results of descriptive analyses

There were no statistically significant differences in reported attitudes as a function of the experimental treatment group (browser vs. app) or the device used to respond at wave 1. Overall, the majority - around 60 percent - of the respondents reported feeling worried that websites and apps collect their personal information and agreed that the Internet poses a threat to privacy. Similarly, around two thirds of respondents reported that they felt very or extremely concerned about the consequences of sharing their data online (67.5 and 66.5 percent respectively for the possibility that data will go to third parties or be used to send targeted advertisements), rising to 77% of respondents reporting they felt very or extremely concerned about the possibility that their identity might be stolen. Some variation across population subgroups was evident for the measure of worry about website and apps collecting personal information and the two measures of concerns about the consequences of sharing data, but not for the other two items (see Appendix Table 2). For example, levels of concern were generally lowest among the youngest and the unmarried respondents. Concerns about targeted advertising and identity theft varied with household size, education and income, with larger proportions of those living alone, with primary or secondary education only or with lower household incomes expressing concern. Statistical differences between subgroups

defined by characteristics relating to internet and smartphone usage were observed for the same three indicators (see Appendix Table 3). For example, differences were observed as a function of frequency of Internet and smartphone use, devices used, the number and type of smartphone activities, with frequent users, completing more activities on their phone being less concerned about data privacy.

Appendix Table 1. Descriptive statistics for data privacy concern measures by treatment group and response device (unweighted)

	(1) All n=644		(2) Group 1 All n=344		(3) Group 2 All n=300		(4) Group 1 PC n=220		(5) Group 1 Mobile n=124		(6) Group 2 App n=219		(7) Group 2 Browser n=81		
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE) p^1	%	(SE)	%	(SE)	p^2
General concerns about data privacy:															
Worried websites & apps collect personal info	58.7	(1.9)	59.3	(2.7)	58.0	(2.9)	61.8	(3.3)	54.8	(4.5)	56.6	(3.4)	61.7	(5.4)	
Agrees Internet poses a threat to privacy	61.8	(1.9)	62.8	(2.6)	60.7	(2.8)	62.7	(3.3)	62.9	(4.4)	59.4	(3.3)	64.2	(5.4)	
Concerns about the consequences of sharing data online:															
Concerned data will go to 3rd parties	67.5	(1.8)	68.0	(2.5)	67.0	(2.7)	68.6	(3.1)	66.9	(4.2)	65.8	(3.2)	70.4	(5.1)	
Concerned data will be used to send targeted ads	66.5	(1.9)	68.3	(2.5)	64.3	(2.8)	69.1	(3.1)	66.9	(4.2)	62.1	(3.3)	70.4	(5.1)	
Concerned identity might be stolen	76.9	(1.7)	76.7	(2.3)	77.0	(2.4)	75.0	(2.9)	79.8	(3.6)	77.2	(2.8)	76.5	(4.7)	
Not comfortable sharing following types of personal information with academic researchers:															
Administrative data from population registers	36.8	(1.9)	36.6	(2.6)	37.0	(2.8)	35.9	(3.2)	37.9	(4.4)	33.8	(3.2)	45.7	(5.6)	
Health data	56.2	(2.0)	54.7	(2.7)	58.0	(2.9)	55.5	(3.4)	53.2	(4.5)	54.3	(3.4)	67.9	(5.2)	
Data on religious beliefs	33.4	(1.9)	31.4	(2.5)	35.7	(2.8)	29.5	(3.1)	34.7	(4.3)†	31.5	(3.1)	46.9	(5.6)	
Data on political opinions	26.2	(1.7)	25.3	(2.3)	27.3	(2.6)	23.2	(2.9)	29.0	(4.1)†	24.2	(2.9)	35.8	(5.4)	
Data relating to criminal record	41.8	(1.9)	41.6	(2.7)	42.0	(2.9)	39.5	(3.3)	45.2	(4.5)†	40.2	(3.3)	46.9	(5.6)	
Data about sex life	60.6	(1.9)	59.9	(2.6)	61.3	(2.8)	56.8	(3.3)	65.3	(4.3)	58.4	(3.3)	69.1	(5.2)	
Data about income and tax records	56.7	(2.0)	57.3	(2.7)	56.0	(2.9)	53.6	(3.4)	63.7	(4.3)†	53.4	(3.4)	63.0	(5.4)	
Data about how smartphone or tablet is used	32.5	(1.8)	31.7	(2.5)	33.3	(2.7)	30.0	(3.1)	34.7	(4.3)	28.8	(3.1)	45.7	(5.6)†	

Notes. [1.]Propensity-score weighted comparison between (4) and (5). [2.]Comparison between (6) and (7). SE = robust standard errors; † p<0.1, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Appendix Table 2. Descriptive statistics for data privacy concern measures by socio-demographics.

Respondent characteristics	n=644	General concerns about online privacy (1) Worried websites & apps collect personal info % (SE)	General concerns about online privacy (2) Agrees Internet poses a threat to privacy % (SE)	Concerns about the consequences of sharing data online (3) Concerned data will go to 3rd parties % (SE)	Concerns about the consequences of sharing data online (4) Concerned data will be used to send targeted ads % (SE)	Concerns about the consequences of sharing data online (5) Concerned identity might be stolen % (SE)
Socio-demographics (from register):						
Male	317	56.2 (2.8)	59.6 (2.8)	66.6 (2.7)	63.4 (2.7)	74.1 (2.5) †
Female	327	61.2 (2.7)	63.9 (2.7)	68.5 (2.6)	69.4 (2.6)†	79.5 (2.2)
Age						
Aged 18-30	138	47.8 (4.3)**	60.1 (4.2)	63.0 (4.1)	54.3 (4.3) ***	68.8 (4.0) ***
Aged 31-55	300	58.3 (2.9)	59.0 (2.8)	66.0 (2.7)	66.0 (2.7)	74.0 (2.5)
Aged 56+	206	66.5 (3.3)	67.0 (3.3)	72.8 (3.1)	75.2 (3.0)	86.4 (2.4)
Marital status						
Married or in partnership	338	63.0 (2.6)*	60.9 (2.7)	69.8 (2.5)	68.0 (2.5)	80.8 (2.1)**
Single/Divorced/ Separated/ Widowed	306	53.9 (2.9)	62.7 (2.8)	65.0 (2.7)	64.7 (2.7)	72.5 (2.6)
Household size						
1 person	102	54.9 (5.0)	65.7 (4.7)	62.7 (4.8)	71.6 (4.5)*	72.5 (4.4)*
2 persons	202	62.9 (3.4)	65.8 (3.3)	69.8 (3.2)	72.3 (3.2)	83.7 (2.6)
3 persons or more	340	57.4 (2.7)	58.2 (2.7)	67.6 (2.5)	61.5 (2.6)	74.1 (2.4)
Urbanicity						
City or town centre/suburban area	176	58.0 (3.7)	66.5 (3.6)†	68.8 (3.5)	63.6 (3.6)	79.0 (3.1)
Village or in the countryside	468	59.0 (2.3)	60.0 (2.3)	67.1 (2.2)	67.5 (2.2)	76.1 (2.0)
Socio-demographics (from survey):						
Education						
Primary or Secondary	291	60.1 (2.9)	62.9 (2.8)	69.4 (2.7)	69.8 (2.7)†	82.8 (2.2)**
Tertiary	353	57.5 (2.6)	60.9 (2.6)	66.0 (2.5)	63.7 (2.6)	72.0 (2.4)
Occupation						
In paid work	390	58.2 (2.5)	60.5 (2.5)	65.9 (2.4)	65.1 (2.4)	75.4 (2.2)
Not in paid work or in training	168	59.4 (3.1)	63.8 (3.0)	70.1 (2.9)	68.5 (2.9)	79.1 (2.6)
Household income						
Less than 10,000 CHF per month	333	59.8 (2.7)	64.6 (2.6)*	67.6 (2.6)	69.4 (2.5)*	80.2 (2.2)**
10,001CHF per month or more	168	56.0 (3.8)	56.0 (3.8)	65.5 (3.7)	59.5 (3.8)	69.6 (3.6)
Reported income	501	58.5 (2.2)	61.7 (2.2)	66.9 (2.1)	66.1 (2.1)	76.6 (1.9)
Did not report income	143	59.4 (4.1)	62.2 (4.1)	69.9 (3.8)	67.8 (3.9)	77.6 (3.5)

Notes. SE = standard errors; † p<0.1, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Appendix Table 3. Descriptive statistics for data privacy concern measures by smartphone usage variables.

		General concerns about online privacy		Concerns about the consequences of sharing data online		
Respondent characteristics	**n**	**(1) Worried websites & apps collect personal info % (SE)**	**(2) Agrees Internet poses a threat to privacy % (SE)**	**(3) Concerned data will go to 3rd parties % (SE)**	**(4) Concerned data will be used to send targeted ads % (SE)**	**(5) Concerned identity might be stolen % (SE)**
Internet usage variables:	**644**					
Uses several times a day/hour	480	57.9 (2.3)	59.0 (2.2)**	67.3 (2.1)	65.0 (2.2)	75.0 (2.0)*
Several times per week or less	164	61.0 (3.8)	70.1 (3.6)	68.3 (3.6)	70.7 (3.6)	82.3 (3.0)
Uses fewer than 3 devices	267	59.9 (3.0)	62.9 (3.0)	67.4 (2.9)	70.4 (2.8)*	74.5 (2.7)
Uses 3 or more devices	377	57.8 (2.5)	61.0 (2.5)	67.6 (2.4)	63.7 (2.5)	78.5 (2.1)
Devices used to access internet						
Desktop and/or laptop	574	58.0 (2.1)	62.0 (2.0)	67.9 (1.9)	65.0 (2.0)*	75.6 (1.8)*
No desktop/laptop	70	64.3 (5.8)	60.0 (5.9)	64.3 (5.8)	78.6 (4.9)	87.1 (4.0)
Smartphone	547	57.8 (2.1)	60.5 (2.1)†	66.5 (2.0)	64.2 (2.1)**	75.3 (1.8)*
No smartphone	97	63.9 (4.9)	69.1 (4.7)	73.2 (4.5)	79.4 (4.1)	85.6 (3.6)
Tablet	258	59.7 (3.1)	64.7 (3.0)	68.6 (2.9)	66.7 (2.9)	81.0 (2.4)*
No tablet	386	58.0 (2.5)	59.8 (2.5)	66.8 (2.4)	66.3 (2.4)	74.1 (2.2
Other devices used	157	63.7 (3.9)†	63.7 (3.9)	74.5 (3.5)*	73.2 (3.5)*	84.7 (2.9)**
No other devices used	487	57.1 (2.2)	61.2 (2.2)	65.3 (2.2)	64.3 (2.2)	74.3 (2.0)
Exciting to try out new tech	245	52.2 (3.2)**	54.7 (3.2)**	67.3 (3.0)	62.0 (3.1)*	72.7 (2.9)*
Does not agree	399	62.7 (2.4)	66.2 (2.4)	67.7 (2.3)	69.2 (2.3)	79.4 (2.0)
Capable of solving tech problems	264	57.2 (3.1)	58.7 (3.0)	68.2 (2.9)	67.4 (2.9)	72.7 (2.7)*
Does not agree	380	59.7 (2.5)	63.9 (2.5)	67.1 (2.4)	65.2 (2.4)	79.7 (2.1)
Smartphone usage variables:	**547**					
Uses SP several times a day/hour	496	56.3 (2.2)*	59.1 (2.2)*	66.3 (2.1)	63.3 (2.2)	75.4 (1.9)
Uses SP once a day or less often	48	72.9 (6.5)	77.1 (6.1)	68.8 (6.8)	72.9 (6.5)	75.0 (6.3)
Smartphone activities…*						
Browse social media	352	53.7 (2.7)**	61.9 (2.6)	64.8 (2.5)	60.8 (2.6)*	75.0 (2.3)
Does not browse social media	192	65.1 (3.4)	58.3 (3.6)	69.8 (3.3)	70.3 (3.3)	76.0 (3.1)
Post on social media	271	53.5 (3.0)*	63.1 (2.9)	66.1 (2.9)	61.3 (3.0)†	76.8 (2.6)
Does not post on social media	273	61.9 (2.9)	58.2 (3.0)	67.0 (2.9)	67.0 (2.9)	74.0 (2.7)
Make purchases	338	54.4 (2.7)*	58.6 (2.7)	67.5 (2.6)	62.7 (2.6)*	74.3 (2.4)
Does not make purchases	206	63.1 (3.4)	64.1 (3.4)	65.0 (3.3)	66.5 (3.3)	77.2 (2.9)
Online banking	260	55.8 (3.1)	58.1 (3.1)	68.1 (2.9)	64.6 (3.0)*	76.9 (2.6)
Does not do online banking	284	59.5 (2.9)	63.0 (2.9)	65.1 (2.8)	63.7 (2.9)	73.9 (2.6)
Installing apps	418	55.7 (2.4)†	59.6 (2.4)	67.5 (2.3)	61.7 (2.4)*	76.1 (2.1)
Does not install apps	126	64.3 (4.3)	64.3 (4.3)	63.5 (4.3)	72.2 (4.0)	73.0 (4.0)
Use apps with location services	459	55.3 (2.3)**	60.3 (2.3)	65.6 (2.2)	63.2 (2.3)**	75.4 (2.0)
Does not use apps with location	85	70.6 (5.0)	62.4 (5.3)	71.8 (4.9)	69.4 (5.0)	75.3 (4.7)
Connect to Bluetooth devices	273	52.7 (3.0)*	59.0 (3.0)	66.7 (2.9)	58.2 (3.0)**	74.7 (2.6)
Does not connect to Bluetooth	271	62.7 (2.9)	62.4 (2.9)	66.4 (2.9)	70.1 (2.8)	76.0 (2.6)
Play games	232	56.0 (3.3)	56.9 (3.3)†	66.8 (3.1)	64.2 (3.2)*	75.9 (2.8)
Does not play games	312	59.0 (2.8)	63.5 (2.7)	66.3 (2.7)	64.1 (2.7)	75.0 (2.5)
Listen to music/ watch videos	435	54.9 (2.4)**	59.3 (2.4)	65.3 (2.3)	61.8 (2.3)*	74.0 (2.1)†
Does not listen to music/videos	109	68.8 (4.5)	66.1 (4.6)	71.6 (4.3)	73.4 (4.3)	80.7 (3.8)
Other activities	185	51.9 (3.7)*	58.9 (3.6)	68.1 (3.4)	61.1 (3.6)	70.8 (3.4)*
No other activities	359	60.7 (2.6)	61.6 (2.6)	65.7 (2.6)	65.7 (2.5)	77.7 (2.2)
Mean no. activities (concerned)	-	8.4 (0.16)**	8.6 (0.18)	8.7 (0.14)	8.5 (0.15)*	8.7 (0.14)
Mean activities (not concerned)	-	9.2 (0.17)	8.9 (0.15)	8.8 (0.20)	9.1 (0.19)	8.8 (0.24)
IOS operating system	308	55.8 (2.8)	61.4 (2.8)	68.8 (2.6)	66.6 (2.7)	75.6 (2.4)
Android or Windows OS[1]	239	60.3 (3.2)	59.4 (3.2)	63.6 (3.1)	61.1 (3.2)	74.9 (2.8)
Smartphone skills (group 1 only)	290					
Beginner or intermediate	115	53.0 (4.7)	61.7 (4.6)	63.5 (4.5)	59.1 (4.6)†	74.8 (4.1)
Advanced	198	60.6 (3.7)	60.6 (3.7)	69.7 (3.5)	68.6 (3.5)	73.7 (3.3)

Notes. [1] only 6 cases with windows OS; SE = robust standard errors; † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix Table 4. Logistic regression coefficients for models predicting probability of reporting feeling uncomfortable about sharing different data types with University researchers (Base: All wave 1 respondents)

	(1) Admin Data				(2) Health Data				(3) Data on Religious Beliefs				(4) Data on political opinions			
	$\widehat{\beta}$	p	SE	*Exp B*	$\widehat{\beta}$	p	SE	*Exp B*	$\widehat{\beta}$	p	SE	*Exp B*	$\widehat{\beta}$	p	SE	*Exp B*
Sociodemographic variables:																
Female	-.171		.181	.843	.114		.177	1.120	.122		.185	1.130	.152		.195	1.164
Age[1]: 31-55 years	.095		.273	1.100	.355		.261	1.426	.255		.291	1.290	-.022		.304	.978
Age: 56+ years	.028		.333	1.029	.685	*	.325	1.984	.669	†	.343	1.953	.185		.358	1.204
Married	-.090		.229	.914	.107		.225	1.113	.076		.235	1.079	.213		.247	1.237
Household size[2]: 2 members	-.007		.296	.993	-.311		.291	.733	-.087		.303	.917	-.079		.315	.924
Household size: 3 members or more	.140		.301	1.150	-.226		.293	.798	.006		.309	1.006	-.205		.322	.815
Urban residence	.351	†	.201	1.420	-.148		.194	.862	.202		.205	1.224	.475	*	.223	1.608
Tertiary education qualification	-.282		.182	.754	.193		.179	1.213	.059		.187	1.061	.075		.197	1.078
Main activity[4]: In paid work	.125		.201	1.133	.360	†	.199	1.433	-.053		.206	.949	.182		.218	1.199
Interested in politics	-.369	*	.186	.691	-.140		.183	.869	-.186		.191	.831	-.334	†	.200	.716
Internet usage variables:																
Uses Internet several times a day	.059		.226	1.061	.348		.229	1.416	.197		.227	1.218	.079		.241	1.082
Has more than 4 devices	.238		.211	1.269	.394		.209	1.482	.432	*	.213	1.540	.629	**	.220	1.875
Excited to try new devices	-.732	***	.206	.481	-.309	†	.195	.734	-.441	*	.208	.644	-.413	†	.221	.661
Able to solve problems with	.208		.196	1.231	-.035		.191	.965	-.014		.200	.986	.003		.212	1.003
Uses a smartphone to access	.037		.290	1.037	.195		.290	1.215	.342		.295	1.408	.178		.316	1.195
Assigned to treatment group 1	-.031		.176	.969	-.214		.173	.807	-.230	***	.179	.794	-.123		.189	.884
Data privacy concerns (mean)	.738	***	.112	2.092	.667	***	.101	1.948	.696	***	.117	2.006	.557	***	.121	1.746
Constant	-3.310	***	.658	.037	-2.829	***	.618	.059	-4.086		.690	.017	-3.802	***	.721	.022
Nagelkerke R^2	.150				.167				.147				.108			

Appendix Table 4 ***continued.***

	(5) Data about criminal records				(6) Data about sex life				(7) Data about income and tax records				(8) Data about mobile device use[5]			
	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*	$\hat{\beta}$	*p*	SE	*Exp B*
Sociodemographic variables:																
Female	-.253		.177	.776	.553	**	.175	1.739	.084		.174	1.087	-.004		.201	.996
Age[1]: 31-55 years	-.122		.263	.885	-.158		.259	.854	.222		.256	1.249	.087		.297	1.091
Age: 56+ years	.098		.320	1.103	.087		.323	1.091	.099		.317	1.104	-.169		.360	.845
Married	.177		.222	1.194	.139		.222	1.149	.175		.221	1.191	.134		.250	1.143
Household size[2]: 2 members	.009		.287	1.009	-.144		.288	.866	-.049		.284	.952	-.079		.327	.924
Household size: 3 members or more	-.082		.291	.921	-.090		.289	.914	-.127		.286	.881	-.304		.329	.738
Urban residence	.121		.193	1.129	.055		.190	1.057	.206		.189	1.228	1.003	***	.243	2.727
Tertiary education qualification	.432	*	.179	1.540	-.019		.176	.981	-.131		.175	.877	-.304		.202	.738
Main activity[4]: In paid work	.204		.198	1.226	.047		.195	1.048	.042		.193	1.043	.261		.227	1.298
Interested in politics	.017		.181	1.017	.010		.179	1.010	-.120		.178	.887	-.238		.207	.789
Internet usage variables:																
Uses Internet several times a day	.240		.221	1.272	.277		.228	1.319	.329		.225	1.389	-.174		.259	.841
Has more than 4 devices	.240		.204	1.272	.164		.207	1.178	.316		.205	1.371	.404	†	.216	1.498
Excited to try new devices	-.628	**	.197	.534	-.181		.192	.834	.006		.192	1.007	-.497	*	.222	.608
Able to solve problems with	-.148		.190	.862	.009		.189	1.009	-.205		.187	.814	.049		.213	1.050
Uses a smartphone to access	.810	**	.292	2.248	.662		.287	1.938	.689	*	.283	1.991				
Assigned to treatment group 1	-.060		.170	.942	-.076	*	.170	.927	.051		.168	1.052	-.057		.195	.945
Data privacy concerns (mean)	.569	***	.104	1.767	.438	***	.094	1.549	.530	***	.095	1.700	.795	***	.128	2.214
Constant	-3.413	***	.633	.033	-2.080	***	.594	.125	-2.672	***	.596	.069	-4.312	***	.696	.013
Nagelkerke R^2	.132				.098				.184				.181			

Notes. [1]Age (ref. 18-30 years old); [2] Household size (ref. single persons); [3] Highest level of education (ref. All non-tertiary); [4] Main activity (ref. not in paid work); [5]Smartphone and tablet users only (n=570).

$\hat{\beta}$ = unstandardized beta coefficient; Exp B = Exponent B. p= p-value, † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

APPENDIX Table 5. Logistic regression analyses predicting concerns about data privacy and the consequences of sharing data online

	General concerns about online data privacy										Specific concerns about the consequences of sharing data							
	(1) Worried websites & apps collect personal info				(2) Agrees Internet poses a threat to privacy				(3) Concerned data will go to 3rd parties				(4) Concerned data will be used to send targeted ads				(5) Concerned identity might be stolen			
	$\hat{\beta}$	p	SE	*Exp B*	$\hat{\beta}$	p	SE	*Exp B*	$\hat{\beta}$	p	SE	*Exp B*	$\hat{\beta}$	p	SE	*Exp B*	$\hat{\beta}$	p	SE	*Exp B*
Sociodemographic variables:																				
Female	0.26		0.17	1.30	0.11		0.17	1.11	0.16		0.18	1.17	0.35	†	0.18	1.42	0.34		0.20	1.40
Age[1]: 31-55 years	0.38		0.25	1.46	-0.13		0.26	0.88	0.22		0.26	1.24	0.60	*	0.26	1.81	0.07		0.28	1.07
Age: 56+ years	0.67	*	0.31	1.95	0.01		0.32	1.01	0.47		0.33	1.60	0.80	*	0.33	2.22	0.71	†	0.37	2.03
Married	0.05		0.22	1.05	-0.15		0.22	0.87	-0.05		0.23	0.95	-0.12		0.23	0.89	0.17		0.26	1.19
Household size[2]: 2 members	0.34		0.28	1.40	0.06		0.29	1.06	0.33		0.29	1.39	0.15		0.30	1.17	0.60	†	0.33	1.83
Household size: 3 members or more	0.31		0.28	1.36	-0.12		0.28	0.89	0.41		0.29	1.51	-0.15		0.30	0.86	0.28		0.31	1.33
Urban residence	0.03		0.19	1.03	-0.29		0.19	0.75	-0.12		0.20	0.89	0.19		0.19	1.21	-0.13		0.22	0.88
Tertiary education qualification	-0.19		0.17	0.83	-0.02		0.17	0.98	-0.22		0.18	0.81	-0.31	†	0.18	0.73	-0.59	**	0.21	0.55
Main activity[4]: In paid work	0.03		0.19	1.03	0.02		0.19	1.02	-0.13		0.20	0.88	0.00		0.20	1.00	0.05		0.23	1.05
Interested in politics	0.57	***	0.17	1.77	0.11		0.18	1.12	0.40	**	0.18	1.49	0.39	*	0.18	1.48	0.03		0.21	1.03
Internet usage variables:																				
Uses Internet several times a day	-0.15		0.22	0.86	0.35		0.22	1.41	-0.12		0.23	0.89	-0.21		0.23	0.81	-0.05		0.27	0.95
Has more than 4 devices	0.17		0.20	1.18	0.22		0.20	1.25	0.25		0.21	1.29	-0.12		0.21	0.88	0.26		0.24	1.30
Excited to try new devices	-0.41	*	0.19	0.67	-0.41	*	0.19	0.66	0.00		0.20	1.00	-0.16		0.20	0.86	-0.20		0.22	0.82
Able to solve problems with devices	0.19		0.19	1.21	0.01		0.19	1.01	0.17		0.19	1.19	0.25		0.19	1.28	-0.06		0.21	0.94
Uses a smartphone to access internet	-0.03		0.28	0.97	-0.11		0.29	0.90	-0.25		0.30	0.78	-0.53		0.32	0.59	-0.25		0.37	0.78
Assigned to treatment group 1	0.09		0.17	1.10	0.09		0.17	1.09	0.08		0.17	1.08	0.19		0.17	1.21	0.04		0.20	1.04
Constant	-0.69		0.48	0.50	0.81		0.49	2.24	0.23		0.50	1.25	0.29		0.52	1.34	1.05	†	0.58	2.85
Nagelkerke R^2	0.07				0.04				0.04				0.08				0.09			
Observations	644				644				644				644				644			

Notes. Base: All wave 1 respondents. Notes. [1]Age (ref. 18-30 years old); [2] Household size (ref. single persons); [3] Highest level of education (ref. All non-tertiary); [4] Main activity (ref. not in paid work) $\hat{\beta}$ = unstandardized beta coefficient; Exp B = Exponent B. p= p-value, † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.