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

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

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. This 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 online panel survey of the general population, as part of the 2019 Swiss Election Study ('Selects'), testing different ways of recruiting participants to an 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 willingness to download a survey app for completing questionnaires, to take and share photos, and to share the smartphone's GPS location and actual completion of these tasks. Our findings confirm that general concerns about online data privacy 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 seem to matter. 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, transforming not only the subject matter of social research, but also the range of methods and sources of data available. Survey researchers have been responding to these developments (Link et al. 2014) and are increasingly eager to benefit from the range of opportunities they offer (Jäckle, Gaia, and Benzeval 2017). Web surveys incorporating mobile respondents now provide better coverage rates than previously (Couper, Antoun and Mavletova 2017) and as people become increasingly dependent on smartphones for accessing the internet (Pew Research Center 2019a), optimizing the design of surveys for mobile response has become a priority for survey methodology.

Among the available options for mobile survey optimization, apps – software installed on smartphones and tablets - are of special 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). Apps also potentially offer 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 et al. 2019; Toepoel, Lugtig and Schouten 2020; Toepoel and Elevelt 2020). Despite their promise, however, early studies testing their feasibility have encountered difficulties around people's willingness to use their smartphone to complete mobile data collection tasks (MDCT) for research purposes. MDCT include activities in which participants play an active role (like completing questionnaires or taking photographs) and those in which participants play a passive role, after consenting to and activating the capture of smartphone sensor data (like GPS location) (Wenz et al. 2017). Several factors have been identified as relevant to understanding people's resistance to MDCT, but one that has emerged as key are concerns

around the privacy and security of personal data (e.g. Jäckle et al. 2017; Wenz et al. 2019; Keusch et al. 2019; Revilla et al. 2019; Struminskaya et al. 2021).

Privacy concerns have been shown to correlate with the propensity to respond to surveys generally and consent to the collection of different types of personal information (Couper et al. 2008; 2010; Singer and Couper 2011; Couper and Singler 2013), as well as with response quality (e.g. Rasinski et al. 1999). As such, they represent an important barrier to the successful integration of new data collection technologies in surveys and a risk to the accuracy of the data gathered (Struminskaya et al. 2021). While previous research has established a broad link between privacy concerns and *hypothetical* willingness to complete MDCT, less is known about how different types of privacy concern affect *actual* MDCT completion, and the extent to which they prohibit the successful implementation of smartphone and app-based surveys.

In this article, we investigate the relationship between two categories of concern about data privacy (general concerns about sharing data online, and discomfort sharing different types of data with researchers) and willingness to complete MDCT. We use data from a probability-based survey of the general population (called 'Selects-Civique'), in which participants were invited to download and respond via a survey app. The design of the study allows us to investigate how general data privacy concerns and discomfort sharing data influence both hypothetical willingness to complete, and actual completion of, three MDCT: (1) downloading an app to respond to survey questionnaires; (2) taking and sharing photographs; and (3) sharing the GPS location of the smartphone. The following research questions are addressed:

- **RQ1:** How do general concerns about online data privacy relate to people's level of discomfort sharing different types of personal data with researchers in the context of a mobile web survey?
- **RQ2:** To what extent do general concerns about online data privacy and discomfort sharing different types of data influence hypothetical willingness to download a survey app and complete MDCT?
- **RQ3:** How do general concerns about online data privacy, discomfort sharing different types of data and hypothetical willingness to complete MDCT influence actual task completion?

Willingness to complete MDCT

A number of studies have investigated willingness to complete MDCT in surveys and its correlates, either by asking respondents about their hypothetical willingness to complete tasks (e.g., Keusch et al. 2021, 2019; Mulder and de Bruijne 2019; Revilla et al. 2019, 2016; Scherpenzeel, 2017; Struminskaya et al. 2021, 2020; Wenz et al. 2019), or by actually implementing MDCT in smartphone or app-based surveys and assessing participation rates (e.g. Scherpenzeel 2017; Kreuter et al. 2020; Jäckle et al. 2019; Elevelt et al. 2019, 2021). A consistent finding in the literature on hypothetical willingness is that willingness varies depending on the nature of the MDCT and the type of data requested (Struminskaya et al. 2021). People are more hesitant about the capture of sensor data from their smartphones than they are about actively completing tasks (Keusch et al. 2021, 2019; Revilla et al. 2019, 2016; Wenz et al. 2019). Two explanations for this include respondents' sense of control over what data are shared (Revilla et al. 2019; Keusch et al. 2021) and the perceived intrusiveness of the data request or sensitivity of the data type (Keusch et al. 2021; Wenz et al. 2019;

Struminskaya et al. 2021). Both appear to outweigh more positive considerations of potential task benefits – e.g., reduced response burden (Mulder and de Bruijne 2019; Keusch et al. 2021).

Besides task characteristics, respondents' attitudes appear to be equally important for explaining variation in hypothetical willingness to complete MDCT (Wenz et al. 2019; Keusch et al. 2019; 2021). These include attitudes towards surveys in general, trust in the survey sponsor and data collection organization, and more general concerns about privacy and data security (e.g. Pinter 2015; Jäckle et al. 2019; Wenz et al. 2019; Mulder and de Bruijne 2019; Keusch et al. 2019; Keusch et al. 2021; Revilla et al. 2019; Struminskaya et al. 2020; 2021). For example, Keusch and his colleagues (2021) found that higher levels of concern about data security reduced hypothetical willingness to complete MDCT, and that levels of concern and willingness varied by task type (see also Struminskaya et al. 2021).

To date, relatively few studies have implemented MDCT in the context of smartphone or app-based surveys (though apps have rapidly gained popularity in health research – e.g., Drew et al. 2020). Those that have, have done so for a variety of purposes, with mixed success. As with survey response more generally (Groves and Couper 1998), numerous factors influence MDCT completion decisions (Wenz et al. 2019; Keusch et al. 2019), making comparisons of task completion rates across studies problematic. Furthermore, because most studies to date have been conducted among existing members of panel surveys (e.g., Jäckle et al. 2019; Kreuter et al. 2020; Scherpenzeel et al. 2017) or, where fresh samples have been drawn, mostly on special populations (e.g., Sugie 2018; Miller et al. 2018; Wang et al. 2014; Lawes et al. 2021), it is difficult to extrapolate conclusions about willingness to complete MDCT in

surveys of the general population (Struminskaya et al. 2021) - or about how this may be influenced by privacy concerns (Keusch et al. 2021).

As with hypothetical willingness, the type of data collected appears to play a part. For example, smartphones seem well-suited to collecting data for surveys in which respondents have to provide regular and frequent reports over an extended period, such as time-use studies involving daily diaries (e.g. Scherpenzeel 2017; Elevelt et al. 2019; Gilbert, Calderwood and Fitzsimmons 2019) and travel/ mobility studies where GPS tracking can replace logs of journeys taken (e.g., Scherpenzeel 2017; Elevelt et al. 2019; Smeets, Lugtig and Schouten 2019). However, other studies pursuing these benefits have seen lower app participation rates, attributed partly to the type of data requested (e.g. scanning shopping receipts (Jäckle et al. 2019); passive collection of digital trace and sensor data (Kreuter et al. 2020); and momentary assessments of health and wellbeing using experience sampling (Lawes et al. 2021).

Less is known about whether privacy concerns also explain some of the variation in actual participation in MDCT. In Jäckle and her colleagues' (2019) study, nonparticipants mentioned concerns around sharing spending data and data security as a reason for not taking part. However, Elevelt and her colleagues (2019) found that while privacy concerns measured at a prior panel wave significantly reduced hypothetical willingness to take part in a smartphone study, they were not predictive of actual participation later on. A similar result was found in an app-free smartphone survey requesting GPS data (Struminskaya et al. 2021). To complicate matters, self-reports of hypothetical willingness do not map directly on to actual participation in MDCT either (though measures of attitudes generally do not predict behaviors well (Ajzen and Fishbein 1977)). Hence, there is still much to learn about the

mechanism by which decisions to participate in MDCT are influenced by different types of privacy concern.

Concerns about data privacy

Smartphone use 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 (e.g. Pew Research Centre 2019b; European Commission 2011, 2015). Expressed concerns do not necessarily translate into actions directed at protecting or mitigating risks to personal privacy online or limiting use of desired services, however. Some lack awareness of privacy risks involved in using the internet (Tozzi and Coppola 2020), while others see data sharing as a pragmatic response to the demands of the information society and necessary to benefit from online services (European Commission 2015).

The 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). It has been documented in the context of social networking and e-commerce activities (e.g. Deuker 2010; Zafeiropoulou et al. 2013), as well as, more recently, when selecting and deciding whether to download mobile apps (Barth et al. 2019). Barth and her colleagues (2019) found that 'functionality, app design, and costs appeared to outweigh privacy concerns' (p.55), even for technically skilled users aware of risks, leading them to use apps that involved divulging and relinquishing control over personal data. Diverse explanations for this paradox have been proposed (Kokolakis 2017), including e.g., time constraints (Barth and De Jong 2017), lack of technical literacy (Liccardi et al. 2014); apathy (Hargittai and Marwick 2016); or a rational risk-benefit analysis in which

concerns relating to data security may be overridden by the perceived benefits of downloading it (Barth et al. 2019). Irrespective, it implies that a more nuanced analysis is needed of when and how privacy concerns and actual behavior concur in the context of requests to complete MDCT in smartphone and app-based surveys.

In summary, actual completion of MDCT appears to depend on a variety of different factors, which may be quite specific to the data-sharing request and its context (Struminskaya et al. 2021), as well as respondent concerns and other characteristics (Keusch et al. 2021). As a result, conclusions drawn about the likely success of smartphone and app-based research based on studies of hypothetical willingness may be misleading.

3. METHODS

3.1. Data

The data come from a three-wave online panel study - 'Selects-Civique' - carried out in the context of the 2019 Swiss Election Study ('Selects' – see Tresch et al. 2021) in the months prior to and immediately following the federal elections (in October). A random probability-based sample of 2,183 adult residents (aged 18 and older) in the French-speaking municipalities of Switzerland, was drawn by the Federal Statistical Office from their sampling frame based on population registers. Selects-Civique was designed to investigate willingness to participate in an election study using a mobile device, and partcularly, to download and complete survey tasks using an app. The sample was randomly assigned in equal parts to two treatment groups to investigate the effect of invitation timing (wave 1 vs. wave 2) on app participation. At wave 1 (fielded in May 2019), group 1 (referred to as the browser group) was invited to participate in a regular browser-based web survey (programmed in Qualtrics®) to be completed on the respondent's device of choice. Group 2 (referred to as the app group)

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 link for sample members preferring a browser-based option. Civique.org is a multimodal data collection application for Android and iOS operating systems (developed by D. Gatica-Perez, J.-I. Biel, O. Bornet, P. Abbet, and D. Santani at Idiap Research Institute, Switzerland), designed as a citizen science platform for mobile data collection initiatives to inform local civic causes.

At wave 2 (fielded in August 2019), the browser group participants were also invited to download and participate via the app (the option to complete via the browser remained open). At wave 3 (fielded in October 2019), both the browser and app groups were re-invited to use the app (and the browser alternative remained available). In total, 687 sample members participated in wave 1 of the survey (AAPOR RR2 = 31.6%). The overall (AAPOR RR2) response rate at wave 1 for the browser group was 33.6% (n=366) and for the app group was 29.5% (n=321), of which 237 (73.8%) responded via the app. Only 358 respondents participated in wave 2 (52.1% of those participating at wave 1). In the browser group, the proportion of respondents who participated in wave 2 was 50.3% (n=184, of which 89 [48.4%] responded via the app) and in the app group, it was 54.2% (n=174, of which 109 [62.6%] used the app). At wave 3, 96.7% (n=178) of the browser group wave 2 sample participated (of which 71 (39.9%) responded via the app), while 94.3% (n=164) of the app group wave 2 sample participated (of which 103 [62.8%] responded via the app). The analytic sample includes all respondents providing complete data for the data privacy measures asked at wave 1 (n=644 of the total n=687 responding[1]). This includes 344 cases from the browser group (31.6% of the gross sample) and 300 cases from the app group (27.6% of the gross

[1] There were differences across devices in the proportion of respondents with missing data on the five data privacy measures. 7.6% 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: $X^2(1) = 3.53$; $p<0.05$).

sample). To assess the impact of privacy concerns on both hypothetical willingness and actual participation, we focus mainly on respondents in the browser group.

3.2. Measures

Concerns about online data privacy: Four measures of which two were general, asking 1) how concerned respondents were that websites and apps collect their personal information; and 2) how concerned they were that their data would go to third parties (see Appendix Table 1 for question wording). The other two asked how concerned respondents were about specific negative consequences of disclosing personal information online: 3) that data will be used to send targeted ads; and 4) that their identity might be stolen. Respondents gave their answers to all items on a 5-point, fully labelled, unipolar response scale (1 = Not at all, and 5 = Extremely). The four measures were highly correlated, so we derived a composite measure of concerns about online data privacy, taking the mean of the four items (giving equal weight to each). Scores ranged from 1 to 5, where 5 indicated the highest level of concern. Those with a mean score above 3.5 were coded 1 to produce a dichotomous indicator (see Appendix A in the Online Supplementary Material (OSM) for justification).

Perceptions of the sensitivity of different types of data (discomfort): Respondents were asked how 'comfortable' (1=Totally comfortable, and 5=Not at all comfortable) they felt about university researchers having access to eight different types of personal information, three of which were actually requested in the study: *administrative data from population registers*, *data on political opinions*, and *data about how you use your smartphone or tablet*. To participate, respondents had to consent to the collection of all three types of data, so we derived a composite indicator of how 'uncomfortable' respondents felt about sharing these three data types (the mean of the three items) (see Appendix A of the OSM). For the three

data types, those selecting 4 and 5 on the scale were coded 1 to create dichotomous indicators. For the composite score, those with a mean score greater than 3.0 were coded 1 (see Appendices A – C of the OSM).

Willingness to complete mobile data collection tasks: The wave 1 questionnaire (for respondents in the browser group who reported using a smartphone to access the internet) included questions assessing *hypothetical willingness* to complete different MDCT in the context of an academic survey. Here, we analyze three measures relevant to tasks respondents were actually asked to complete in Selects-Civique: willingness (1) to download a survey app to respond to questionnaires, (2) to take and share photographs (specifically, of political posters during the election campaign), and (3) to agree to GPS tracking. Respondents were asked how willing they would be to complete each task and gave their answer on a four-point scale (1= Completely willing, and 4= Not willing at all). We derived dichotomous measures by collapsing categories 1 and 2 (1= willing) and 3 and 4 (0= not willing).

Actual participation in the three tasks is measured by compliance with the task requests among browser group smartphone respondents (who responded to the hypothetical willingness items) – 1) installing the app[2] to participate at either wave 2 or wave 3; 2) taking photographs (at wave 2); and 3) activating location services within their privacy settings to allow the passive capture of GPS coordinates when the app was in use (at wave 3). For details of how the tasks were presented to respondents see Appendix Table 2.

Respondent Characteristics: Auxiliary data from the sampling frame provides the following

[2] For technical reasons, we do not have an indicator of how many participants accessed the app stores or downloaded the app but did not proceed further – only of those who downloaded and actually participated in the study. For this reason, our single measure of compliance is the proportion of the group 1 sample from wave 1 that responded via the app at wave 2.

socio-demographic variables: 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 measures of behavioral characteristics, 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 (see Appendix Table 3 for descriptive statistics).

3.3. Analytic Approach

Despite the random assignment of sample members to the two treatment groups at wave 1, respondents could choose which device/ software to use to respond (in the browser group, between a PC or mobile browser; and in the app group, between the app and a PC or mobile browser). Because characteristics of the response device could potentially affect how respondents answer questions (Lugtig and Toepoel 2016), or respondents with different levels of concern about online data privacy and comfort sharing data might choose to respond to the survey via different access options, we first assessed whether there were differences in the composition of the samples responding in each experimental group and on different devices within groups. We found small statistically significant differences for which we adjusted using propensity score weights (procedures used are described in Appendix D of the OSM). To assess whether the experimental treatment and/or the chosen response device were associated with the measures of interest, we used Chi-square tests of association to test differences between subgroups across categories for each of the dichotomized privacy concern measures and Independent Samples T-tests to compare means on the composite measures.

For the remainder of our analyses, we estimated the parameter coefficients of a series of logistic regression equations predicting the probability of 1) reporting feeling discomfort about university researchers having access to each of the data types requested (RQ1); 2) being *hypothetically* willing to a) install a survey app to complete questionnaires, b) take photos , and c) share their GPS location of their smartphone (RQ2); and 3) *actually* a) installing the app to participate in wave 2, b) 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 (RQ3). The first set of models were fitted for (1) all wave 1 respondents with complete data (n=644); and the second and third set were fitted only for respondents assigned to the browser group who reported using a smartphone to access the Internet (n=290), with complete data for the questions measuring hypothetical willingness (n=289). For the models predicting willingness, we use a sequential approach starting with the privacy concerns indicator, then adding the discomfort indicator, then finally, for actual completion, we add the indicator of hypothetical willingness to perform the task in question. Covariates in all models included the socio-demographic, internet usage measures and for (1) a control for the experimental design.

4. RESULTS

Preliminary analyses confirmed there were no statistically significant differences in data privacy concerns and discomfort as a function of the experimental treatment group (browser vs. app) (see Appendix Table 4 for descriptive statistics for the sample as a whole and by experimental and device group). In the browser group, there were no differences due to the device used (PC vs. mobile browser) at wave 1 in either the unweighted or weighted estimates. In the app group, there were differences by device on some of the unweighted estimates, which were no longer significant once the propensity score weight was applied.

4.1 General concerns about online data privacy and discomfort sharing different types of data (RQ1)

Perception of the sensitivity of the data varied as a function of data type (see lower half of Appendix Table 4, column 1). The proportions of the full sample reporting discomfort about university researchers having access to their data were lowest for the data types actually requested in the Selects-Civique study: 1) administrative data from population registers (36.8%); 2) data about how their smartphone or tablet is used (32.5%); and 3) data on political opinions (26.2%). Coefficients from logistic regression models predicting discomfort sharing these data types are shown in Table 1, alongside those for the model predicting overall discomfort (see Appendix E of the OSM for models predicting other data types). For all data types considered, general concerns about online data privacy were positively and significantly predictive of the probability of reporting discomfort about university researchers having access to personal information. The more concerned people are about online data privacy in general, the more uncomfortable they are about researchers accessing their personal data. Focusing on the model predicting overall discomfort in the full sample, the effect of the general measure of data privacy concerns was positive and highly statistically significant (b = .811, Wald X^2 (1) = 47.78, p<0.001). The probability of reporting discomfort was 2.5 times greater for each unit increase in the mean privacy concern measure (Exp(B) = 2.510, $CI_{.95}$= [1.948, 3.234]). Other significant covariates were living in an urban area, which was positively associated with discomfort (b = .629, Wald X^2 (1) = 9.15, p<.01); as was using more than four devices to access the Internet (b = .647, Wald X^2 (1) = 9.74, p<.01); and agreeing that it is exciting to try out new technologies, which was significantly negatively associated with discomfort (b = -5.96, Wald X^2 (1) = 8.16, p<.01). For the other models, the effect of the covariates varied by data type.

4.2 Hypothetical willingness to complete and actual completion of MDCT (RQ2 and RQ3)

Table 2 shows descriptive statistics for hypothetical willingness and actual task completion among smartphone users in the browser group. Hypothetical 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%). The proportion of browser group smartphone users who actually downloaded the survey app at wave 2 to complete questionnaires (32.1%) was substantially lower than that for hypothetical willingness, resulting in small sample sizes for the subsequent analyses. However, 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 at wave 3 (20.3%). As completion of the photo tasks and activating location services in the app was conditional on participation via the app at waves 2 and 3 respectively, Table 2 also presents the proportion who completed these tasks among app respondents at each wave. Levels of compliance with task requests are high, despite the low levels of hypothetical willingness expressed at wave 1. At wave 2, 95% (78 out of 82) of the browser group app respondents completed at least one photo task. At wave 3, 84.3% (59 out of 70) activated location services in the app and provided GPS data.

The discrepancy between hypothetical willingness to complete MDCTs and actual rates of participation is further evidenced in Table 3, which cross-tabulates hypothetical willingness and actual completion for all wave 1 browser group smartphone users. For all three tasks, the proportion who participated among those who reported being hypothetically willing to do so was higher than it was among those who reported not being willing to do so. This association

was statistically significant for downloading an app ($X^2(1)=4.37$; $p<0.05$) and taking photographs ($X^2(1)=9.83$; $p<0.01$), but not so for activating location services to share GPS data. Nevertheless, the proportion of unwilling respondents who actually completed tasks was noteworthy. For downloading an app, 38.8% of those who said they would be willing to do so actually did, compared with 27.2% of those who said they were not willing to do so. For taking photographs, 41.8% of those who said they were willing actually did take and share a photo, compared with 22.4% of those who said they were not willing. Finally, for activating location services to share GPS data, 26.1% of those who said they were willing to do so

Table 1. Logistic regression coefficients for models predicting probability of reporting discomfort sharing different data types with university researchers

	(1) Admin Data				(2) Data on political opinions				(3) Data about mobile device use				(4) Overall discomfort			
	$\hat{\beta}$	p	SE	*Exp B*	$\hat{\beta}$	p	SE	*Exp B*	$\hat{\beta}$	p	SE	*Exp B*	$\hat{\beta}$	p	SE	*Exp B*
Data privacy concerns (mean)	.738	***	.112	2.092	.557	***	.121	1.746	.795	***	.128	2.214	.920	***	.129	2.510
Internet usage variables:																
Uses Internet once a day or less	.059		.226	1.061	.079		.241	1.082	-.174		.259	.841	-.052		.253	.950
Has more than 4 devices	.238		.211	1.269	.629	**	.220	1.875	.404	†	.216	1.498	.687	**	.214	1.988
Excited to try new devices	-.732	***	.206	.481	-.413	†	.221	.661	-.497	*	.222	.608	-.631	**	.219	.532
Able to solve problems with	.208		.196	1.231	.003		.212	1.003	.049		.213	1.050	.130		.210	1.139
Uses a smartphone to access	.037		.290	1.037	.178		.316	1.195	-		-	-	-		-	-
Assigned to Browser Group	-.031		.176	.969	-.123		.189	.884	-.057		.195	.945	-.073		.192	.929
Sociodemographic variables:																
Female	-.171		.181	.843	.152		.195	1.164	-.004		.201	.996	.057		.198	1.059
Age[1]: 31-55 years	.095		.273	1.100	-.022		.304	.978	.087		.297	1.091	.109		.294	1.115
Age: 56+ years	.028		.333	1.029	.185		.358	1.204	-.169		.360	.845	.091		.353	1.096
Married	-.090		.229	.914	.213		.247	1.237	.134		.250	1.143	.112		.247	1.119
Household size[2]: 2 members	-.007		.296	.993	-.079		.315	.924	-.079		.327	.924	.230		.328	1.259
Household size: 3 members or more	.140		.301	1.150	-.205		.322	.815	-.304		.329	.738	.020		.330	1.020
Urban residence	.351	†	.201	1.420	.475	*	.223	1.608	1.003	***	.243	2.727	.801	***	.228	2.227
Tertiary education qualification	-.282		.182	.754	.075		.197	1.078	-.304		.202	.738	-.133		.200	.875
Main activity[4]: In paid work	.125		.201	1.133	.182		.218	1.199	.261		.227	1.298	.348		.223	1.416
Interested in politics	-.369	*	.186	.691	-.334	†	.200	.716	-.238		.207	.789	-.415	*	.204	.660
Constant	-3.310	***	.658	.037	-3.802	***	.721	.022	-4.312	***	.696	.013	-4.92		.703	.007
Model X^2 (17/16)	74.97***				49.60***				80.31***				102.90***			
Hosmer & Lemeshow X^2 (8)	7.17				6.71				9.03				12.29			
Nagelkerke R^2	.15				.11				.18				.22			
Observations (n)	644				644				580[5]				580[5]			

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=580). $\hat{\beta}$ = unstandardized beta coefficient; Exp B = Exponent B. p= p-value, † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2. Hypothetical willingness and actual participation in mobile data collection tasks by wave of data collection: Browser group respondents, smartphone users and app users

	All W1 Browser Group Respondents	Hypothetically willing to download a survey app to complete questionnaires		Hypothetically willing to take and share photos (of a political poster[1])		Hypothetically willing to share GPS location of smartphone		Actually downloaded the app & completed questionnaires (W2 or W3)		Actually took and shared at least one photo (W2)		Actually activated location services and shared GPS data (W3)	
	n	***n***	**% (SE)**	***n***	**% (SE)**	***n***	**% (SE)**	***n***	**% (SE)**	***n***	**% (SE)**	***n***	**% (SE)**
Wave 1													
W1 Browser Group Smartphone Users	290	121	41.7 (2.9)	67	23.1 (2.5)	46	15.9 (2.2)	93	32.1 (2.8)	78	26.9 (2.6)	59	20.3 (2.4)
Wave 2													
W2 Browser Group Smartphone Users	147	69	46.9 (4.1)	45	30.6 (3.8)	24	16.3 (3.1)	87	59.2 (4.1)	78	53.1 (4.1)	54	36.7 (4.0)
W2 Browser Group Smartphone Users using the App	82	39	47.6 (5.6)	28	34.1 (5.3)	18	22.0 (4.6)	-	-	78	95.1 (2.4)	49	59.8 (5.5)
Wave 3													
W3 Browser Group Smartphone Users	141	67	47.5 (4.2)	37	26.2 (3.7)	26	18.4 (3.3)	79	56.0 (4.2)	64	45.4 (4.2)	59	41.8 (4.2)
All W3 Browser Group Smartphone Users using the App	70	37	52.9 (6.0)	24	34.3 (4.8)	14	20.0 (4.3)	-	-	55	78.6 (4.9)	59	84.3 (4.4)

Notes. [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).

Table 3. Hypothetical willingness vs. actual participation in mobile data collection tasks among browser group smartphone users for research purposes

			Downloaded the app & completed questionnaires			
			No	Yes	Total	
Hypothetically willing to download a survey app to complete questionnaires	No	Count	123	46	169	$X^2(1)=4.37$*
		%	72.8	27.2	100.0	
	Yes	Count	74	47	121	
		%	61.2	38.8	100.0	
	Total	Count	197	93	290	
		%	67.9	32.1	100.0	
			Took and shared at least one photo			
			No	Yes	Total	
Hypothetically willing to take and share photos (of a political poster[2])	No	Count	173	50	223	$X^2(1)=9.83$**
		%	77.6	22.4	100.0	
	Yes	Count	39	28	67	
		%	58.2	41.8	100.0	
	Total	Count	212	78	290	
		%	73.1	26.9	100.0	
			Activated location services and shared GPS data			
			No	Yes	Total	
Hypothetically willing to share GPS location of smartphone	No	Count	197	47	244	$X^2(1)=1.11$
		%	80.7	19.3	100.0	
	Yes	Count	34	12	46	
		%	73.9	26.1	100.0	
	Total	Count	231	59	290	
		%	79.7	20.3	100.0	

Notes. Base = All Wave 1 Browser Group smartphone users. Chi-square Tests of Association (X^2=Pearson's Chi-square). † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

actually did, compared with 19.3% of those who said they were not willing. Thus, while overall participation rates are low, being hypothetically unwilling to complete MDCT does not necessarily translate into non-participation, and once responding via the app, most respondents cooperated with task requests.

Relation to data privacy concerns and discomfort sharing requested data types

Finally, we address the question of how general concerns about online data privacy and discomfort sharing specific data types with academic researchers relate to hypothetical and

actual willingness to complete MDCT. Firstly, in the case of *hypothetical* willingness to install a survey app (left-hand side of table 4), the composite measure of concerns about data privacy significantly and negatively predicts the probability being willing to download a survey app to complete questionnaires (b = -.364, Wald X^2 (1) = 5.74, p<0.05). However, this effect is no longer significant when the 'discomfort' measure is included in the models. Instead, discomfort sharing data is a significant, negative predictor of the probability of being willing to download a survey app (b = -.658, Wald X^2 (1) = 17.10, p<0.01). The greater the level of discomfort, the lower the likelihood of being hypothetically willing to download an app. Similarly, in the case of GPS location (left-hand side of table 6), the composite measure of data privacy concerns significantly and negatively predicts willingness (b = -.474, Wald X^2 (1) = 7.36, p<0.01). Once again, when the indicator of discomfort is entered into the model, the effect of data privacy concerns is no longer significant, but is outweighed by the negative, statistically significant effect of discomfort sharing the data types requested in the study (b = -.467, Wald X^2 (1) = 6.08, p<0.05). Finally, in the case of hypothetical willingness to take and share photos (left-hand side of table 5), the general data privacy concern measure is *not* a significant predictor in either model. However, the measure of discomfort is once again, a significant negative predictor of hypothetical willingness to take photographs (b = -.467, Wald X^2 (1) = 8.03; p<0.01).

Alongside the negative effects of privacy concerns and discomfort sharing data on hypothetical willingness, some other observations can be made about covariates. In relation to installing an app, as for comfort sharing data, living in an urban area decreased the probability of being hypothetically willing (b = -.852, Wald X^2 (1) = 6.06; p<0.05). By contrast, being

Table 4. Logistic regression coefficients from models predicting hypothetical willingness to install a survey app to complete questionnaires and actual participation

	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*
Data privacy concerns (mean)	-.364	*	.152	.695	-.083		.168	.920	-.042		.158	.959	-.032		.159	.969
Discomfort sharing data (mean)	-		-	-	-.658	***	.159	.518	-.150		.140	.861	-.112		.143	.894
Hypothetical willingness	-		-	-	-		-	-	-		-	-	.346		.312	1.413
Internet usage variables:																
Uses Internet once a day or less	-.129		.412	.879	-.150		.424	.861	-.935	*	.452	.393	-.931	*	.453	.394
Has more than 4 devices	-.109		.317	.897	.027		.333	1.027	-.403		.319	.668	-.408		.321	.665
Excited to try new devices	.453		.312	1.573	.474		.327	1.606	.410		.314	1.506	.383		.315	1.467
Able to solve problems with devices	.710	*	.309	2.034	.759	*	.322	2.137	.272		.302	1.312	.215		.306	1.240
Number of smartphone activities	.309	***	.065	1.361	.307	***	.067	1.359	-.009		.057	.991	-.029		.060	.972
Has an Android phone	-.187		.285	.829	-.210		.298	.811	-.493	†	.281	.611	-.481	†	.281	.618
Sociodemographic variables:																
Female	.527	†	.297	1.695	.524	†	.308	1.688	.063		.285	1.065	.024		.288	1.024
Age[1]: 31-55 years	.179		.388	1.196	.280		.405	1.323	-.842	*	.379	.431	-.861	*	.380	.423
Age: 56+ years	.426		.524	1.532	.566		.546	1.760	-.303		.501	.739	-.340		.506	.712
Married	.534		.347	1.705	.461		.364	1.585	.396		.350	1.485	.370		.352	1.447
Household size[2]: 2 members	-.204		.444	.815	-.215		.462	.807	-.106		.464	.899	-.083		.465	.920
Household size: 3 members or more	-.574		.432	.563	-.422		.451	.656	.185		.443	1.203	.222		.445	1.248
Urban residence	-.915	**	.331	.401	-.852	*	.346	.426	.025		.321	1.025	.077		.325	1.080
Tertiary education qualification	-.001		.293	.999	-.088		.305	.916	.334		.290	1.397	.346		.290	1.413
Main activity[4]: In paid work	.059		.317	1.060	.095		.332	1.100	.292		.314	1.339	.285		.314	1.329
Interested in politics	.963	**	.311	2.619	.913	**	.322	2.491	-.165		.297	.848	-.222		.303	.801
Constant	-2.398	*	1.048	.091	-1.761	*	1.082	.172	-.087		.990	.917	-.161		.993	.851
Model X^2 (17/18/19)	70.87***				89.75***				29.90*				32.28*			
Hosmer & Lemeshow X^2 (8)	2.37				7.14				15.35†				9.96			
Nagelkerke R^2	.29				.36				.14				.15			
Observations (*n*)	289				289				289				289			

Notes. Base includes Wave 1 Browser Group 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 willingness to take and share photos and actual participation

	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*
Data privacy concerns (mean)	-.014		.161	.987	.187		.177	1.206	-.190		.166	.827	-.222		.169	.801
Discomfort sharing data (mean)	-		-	-	-.467	**	.165	.627	-.067		.149	.936	-.001		.153	.999
Hypothetical willingness to take photos	-		-	-	-		-	-	-		-	-	.811	*	.339	2.250
Internet usage variables:																
Uses Internet once a day or less	-.188		.491	.829	-.174		.493	.841	-.423		.461	.655	-.396		.467	.673
Has more than 4 devices	-.036		.357	.965	.064		.364	1.066	-.002		.331	.998	.002		.335	1.002
Excited to try new devices	.088		.350	1.092	.089		.356	1.093	.438		.332	1.550	.434		.339	1.544
Able to solve problems with devices	.045		.340	1.046	.072		.349	1.075	.307		.317	1.359	.299		.322	1.349
Number of smartphone activities	.117		.068	1.124	.099		.068	1.104	-.021		.060	.980	-.033		.062	.968
Has an Android phone	-.206		.312	.814	-.218		.316	.804	-.567	†	.298	.567	-.558	†	.303	.572
Sociodemographic variables:																
Female	.367		.314	1.443	.396		.321	1.485	.182		.302	1.200	.151		.306	1.164
Age[1]: 31-55 years	-.784	†	.404	.456	-.703	†	.409	.495	-.929	*	.396	.395	-.820	*	.405	.440
Age: 56+ years	-1.415	*	.585	.243	-1.358	*	.597	.257	-.678		.532	.508	-.498		.546	.607
Married	.612		.386	1.843	.500		.395	1.648	.165		.364	1.180	.083		.371	1.087
Household size[2]: 2 members	-.437		.510	.646	-.440		.519	.644	-.025		.503	.976	.096		.512	1.101
Household size: 3 members or more	-.336		.473	.715	-.199		.482	.820	.533		.472	1.703	.619		.482	1.858
Urban residence	-.129		.361	.879	-.020		.367	.980	-.089		.338	.915	-.088		.341	.915
Tertiary education qualification	-.024		.323	.976	-.089		.329	.915	.463		.308	1.590	.498		.313	1.645
Main activity[4]: In paid work	-.303		.340	.739	-.288		.347	.749	.406		.337	1.500	.445		.339	1.561
Interested in politics	1.183	**	.360	3.265	1.150	**	.365	3.159	-.027		.312	.973	-.194		.324	.824
Constant	-2.126	†	1.165	.119	-1.607		1.176	.200	-.325		1.041	.722	-.564		1.057	.569
Model X^2 (17/18/19)	33.35*				41.73**				32.03*				37.73**			
Hosmer & Lemeshow X^2 (8)	5.57				2.92				6.19				3.17			
Nagelkerke R^2	.17				.20				.15				.18			
Observations (n)	289				289				289				289			

Notes. Base includes Wave 1 Browser Group 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 6. Logistic regression coefficients predicting hypothetical willingness to share GPS location and actual participation

	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*
Data privacy concerns (mean)	-.474	**	.175	.622	-.282		.191	.754	.077		.183	1.080	.096		.186	1.101
Discomfort sharing data (mean)	-		-	-	-.467	*	.189	.627	-.195		.158	.823	-.178		.159	.837
Hypothetical willingness to share GPS	-		-	-	-		-	-	-		-	-	.340		.413	1.405
Internet usage variables:																
Uses Internet once a day or less	-.613		.575	.541	-.601		.576	.548	-.471		.489	.624	-.448		.490	.639
Has more than 4 devices	-.027		.387	.973	.057		.395	1.058	-.480		.371	.619	-.482		.372	.617
Excited to try new devices	.725	†	.401	2.065	.685	†	.406	1.983	.118		.358	1.126	.089		.361	1.093
Able to solve problems with devices	.121		.396	1.129	.142		.405	1.153	-.032		.346	.968	-.044		.347	.957
Number of smartphone activities	.028		.074	1.028	.015		.074	1.015	.033		.065	1.033	.035		.065	1.035
Has an Android phone	-.497		.362	.608	-.495		.366	.610	-.558	†	.322	.572	-.536	†	.323	.585
Sociodemographic variables:																
Female	.343		.361	1.409	.336		.368	1.399	-.086		.324	.918	-.102		.325	.903
Age[1]: 31-55 years	-.179		.467	.836	-.071		.472	.931	-.797	†	.434	.451	-.791	†	.435	.453
Age: 56+ years	-.526		.657	.591	-.422		.671	.656	-.097		.561	.907	-.068		.564	.934
Married	.427		.435	1.533	.324		.444	1.383	.503		.396	1.653	.485		.397	1.624
Household size[2]: 2 members	-.376		.540	.686	-.388		.551	.678	-.048		.537	.953	-.014		.537	.986
Household size: 3 members or more	-.814		.520	.443	-.664		.529	.515	.279		.509	1.322	.319		.512	1.376
Urban residence	-.186		.404	.830	-.111		.409	.895	.246		.371	1.279	.245		.371	1.278
Tertiary education qualification	-.747	*	.362	.474	-.849	*	.370	.428	.511		.337	1.668	.551		.342	1.736
Main activity[4]: In paid work	-.062		.392	.940	-.011		.398	.989	.558		.368	1.747	.573		.370	1.774
Interested in politics	.157		.371	1.170	.108		.377	1.114	-.332		.338	.717	-.346		.339	.708
Constant	.582		1.174	1.789	1.119		1.202	3.061	-1.639		1.158	.194	-1.882		1.195	.152
Model X^2 (17/18/19)	24.65				30.98*				20.35				21.01			
Hosmer & Lemeshow X^2 (8)	8.84				11.73				5.31				6.20			
Nagelkerke R^2	.14				.17				.11				.11			
Observations (*n*)	289				289				289				289			

Notes. Base includes Wave 1 Browser Group 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, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

interested in politics increased the probability of being willing to download an app (b = .913, Wald X^2 (1) = 8.01; p<0.01), as did having more advanced technological skills (b = .759, Wald X^2 (1) = 5.57; p<0.05) and using a smartphone for a larger number of activities besides calls and texts (b = .307, Wald X^2 (1) = 21.21; p<0.001). For hypothetical willingness to take photos, being interested in politics was also a significant, positive predictor (b = 1.150, Wald X^2 (1) = 9.91; p<0.01). Being aged 55 and older significantly reduced the probability of being willing to take photographs compared to those aged 18-30 years (b = -1.358, Wald X^2 (1) = 5.181; p<0.05) and being 31 years and older also reduced the probability of being willing though the effect only approached significance (b = -.703, Wald X^2 (1) = 2.955; p<0.1). Finally, in relation to sharing the GPS location, having a tertiary level qualification significantly reduced the probability of reporting being willing (b = -.849, Wald X^2 (1) = 52.35; p<0.05) and a weak positive effect (approaching significance) was observed for those agreeing they are excited to try new technologies (b = .685, Wald X^2 (1) = 2.845; p<0.1).

For *actual* participation (right-hand sides of tables 4, 5, 6), in all cases, neither the general measure of concerns about online data privacy, nor the measure of discomfort sharing relevant data types are significant predictors of completing the MDCT. Instead, weak effects are observed in all models for just three covariates. Firstly, using the internet less than once a day compared to using it several times an hour or day – less frequent internet users were significantly less likely to download and respond via the app (b = -.935, Wald X^2 (1) = 4.274; p<0.05). They were also less likely to take photos and share their GPS location, but the effects were not significant in these models. Second, weak negative effects approaching significance were observed for respondents using the Android (or Windows) operating system compared to iOS. Android users were less likely to download and respond via the app (b = -.493, Wald X^2 (1) = 3.079; p<0.1), less likely to take and share photos (b = -.567, Wald X^2 (1) = 3.617;

p<0.1) and less likely to activate location services in the app to share GPS data (b = -.558, Wald X^2 (1) = 3.011; p<0.1). Finally, being aged 31-55 compared to being aged 18-30 significantly decreased the likelihood of downloading and responding via the app (b = -842, Wald X^2 (1) = 4.945; p<0.05), taking and sharing a photograph (b = -.929, Wald X^2 (1) = 5.496; p<0.05), and activating location services (b = -.797, Wald X^2 (1) = 3.373; p<0.1). Respondents aged 56 and older were also less likely to complete the tasks than the youngest, but the effect was not significant.

5. DISCUSSION

Research into hypothetical willingness to complete MDCT in surveys (e.g. Struminskaya et al. 2021; Keusch et al. 2021; Keusch et al. 2019; Revilla et al. 2019; Wenz et al. 2019) suggests that concerns about data privacy may pose a barrier to gaining participants' cooperation and hence, a risk to data quality. This study investigated whether and how privacy concerns and discomfort sharing different data types affect *actual* decisions to complete MDCT in a general population app-based panel study, at wave 2 of which, half the participants were invited to install an app to complete questionnaires and share photographs, and, at wave 3, to activate location services to allow the passive capture of their smartphone's GPS location. A summary of the main results of our analyses can be found in Table 7.

We found variation in discomfort about researchers having access to personal data depending on the data type. By far the strongest predictor of discomfort sharing any of the data types considered was respondents' general data privacy concern, which was significant across the board (RQ1). Other variables (e.g., topic interest, living in an urban area, and being excited to try new technologies), were also relevant for understanding variation in discomfort, but their influence varied by data type. In keeping with the findings of other studies (e.g. Wenz et al.

2019; Revilla et al. 2019; Keusch et al. 2019; Struminskaya et al. 2021; Keusch et al. 2021), we also found that data privacy concerns were strong, negative predictors of hypothetical willingness to download an app and agree to passive capture of GPS data (RQ2), but not of hypothetical willingness to take and share photographs. This may have been because respondents were asked about their willingness to take a photograph of a political poster and this was not deemed to be an especially intrusive request (unlike in Struminskaya et al.'s (2021) study).

Once the composite measure of discomfort sharing data requested in the study (administrative data, political opinions and mobile phone data) was included in the models, however, the effect of the privacy measure was negated – even for hypothetical willingness to take photographs, discomfort was a significant negative predictor. 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, the finding is informative because of the theoretical interest of considering them separately. General online data privacy concerns do not appear *per se* to be a barrier to all types of MDCT and should be considered as less relevant than more proximate considerations of the sensitivity of different data types requested in a study, or how comfortable respondents feel sharing them (RQ2).

In the models predicting actual compliance with task requests (RQ3), neither general data privacy concerns nor discomfort sharing data were statistically significant. Instead, just three variables were: respondent's age and the operating system of their smartphone, and, in the case of downloading the app only, how frequently they use the internet. Older respondents were less likely to complete all 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 providing data via a smartphone (Keusch et al. 2021). However, it is not clear why Android users were less

Table 7. Summary of main results

Research Question	Measures	Main Results
RQ1: How do general concerns about online data privacy relate to people's level of discomfort sharing different types of personal data with researchers in the context of a mobile web survey?	- *Dependent variables:* Discomfort sharing administrative data, data on political opinions and data about mobile device use; and a dichotomized composite measure of all three. - *Covariates:* General data privacy concerns, internet usage variables and socio-demographic variables (including interest in politics).	- Discomfort sharing personal data with university researchers varies by data type (see Appendix Table 4). - There is a positive association between general concerns about online data privacy and discomfort sharing data with researchers. The more concerned people are about online data privacy, the more uncomfortable they are about researchers accessing their personal data (see Table 1). - *Significant covariates* (model predicting composite measure of discomfort): Living in an urban area (+), using more than 4 devices to access Internet (+), and agreeing it is exciting to try new technologies (-) (Table 1).

RQ2: To what extent do general concerns about online data privacy and discomfort sharing different types of data influence hypothetical willingness to download a survey app and complete mobile data collection tasks (MDCT)?	- *Dependent variables:* Hypothetical willingness to complete MDCT: download an a survey app, take photographs and share GPS location.
	- *Covariates:* General data privacy concerns, discomfort sharing data types with researchers, internet usage variables and socio-demographic variables (including interest in politics).
		- Hypothetical willingness to complete MDCT varies by task (see Table 2).
		- There is a significant negative association between general data privacy concerns and hypothetical willingness to install an app and share GPS data, but not for hypothetical willingness to take photos (see Tables 4, 5, and 6).
		- Discomfort sharing data with researchers is significantly and negatively associated with hypothetical willingness to complete all three MDCT.
		- The effect of discomfort sharing data overrides the effect of general data privacy concerns on hypothetical willingness to install an app and share GPS data.
		- *Significant covariates:*
		  - *Willing to install an app:* Living in an urban area (-), interested in politics (+), no. of smartphone activities (+), being able to solve problems with devices (+), and being female (+) (Table 4).
		  - *Willing to take a photo:* Being aged 31 and older (-), being interested in politics (+) (Table 5).
		  - *Willing to share GPS:* Having completed tertiary-level education (-) (Table 6).

RQ3: How do general concerns about online data privacy, discomfort sharing different types of data and hypothetical willingness to complete MDCT influence actual task completion?	- *Dependent variables:* Actual completion of MDCT (downloaded an app, took photographs and shared GPS data).
	- *Covariates:* General data privacy concerns, discomfort sharing data types with researchers, hypothetical willingness to complete MDCT, internet usage variables and socio-demographic variables (including interest in politics).
		- Reporting being hypothetically unwilling to complete MDCT does not necessarily translate into non-participation in MDCT. Once responding via the app, most respondents cooperated (see Tables 2 and 3).
		- Neither general data privacy concerns nor discomfort sharing data are significantly associated with actual completion of MDCT (see Tables 4, 5 and 6).
		- Hypothetical willingness to complete MDCT is not associated with actual completion of MDCT, except in the case of taking a photograph.
		- *Significant covariates:*
		  - *Installed an app:* Using Internet once a day or less (-), being aged 31-55 compared to younger respondents (-), having an Android phone (-) (Table 4).
		  - *Shared a photo:* Being aged 31-55 (-), having an Android phone (-) (Table 5).
		  - *Shared GPS data:* Being aged 31-55 (-), having an Android phone (-) (Table 6).

Notes. RQ = Research Question; MDCT = Mobile Data Collection Tasks; - = negative association; + = positive association.

willing to complete MDCT. We are not aware of any technical reasons why this was the case. Android respondents were significantly younger and less likely to have completed a tertiary-level qualification than iPhone respondents, which may account for their lower engagement. 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 concerns 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 'privacy paradox' mentioned previously (Kokolakis 2017), for which we find some evidence in our data. Measures of general attitudes are not always strong predictors of specific actions (Crano and Prislin 2006; Ajzen and Fishbein 1977). Our findings confirm that more specific, proximal measures of concern that fit more closely the specific data collection context (i.e., discomfort sharing the data types requested) better predict participation in MDCT, and appear to outweigh more general privacy concerns. This finding fits with theory relevant to understanding how privacy concerns contribute to resistance to technological innovation (Nissenbaum 2010), which emphasises their context-dependency and the need to factor in aspects like the attributes of the technology in question (e.g., demands on users and types of information shared) and the roles and activities of different actors involved, (Nissenbaum 2010, p. 141).

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 further reduced by the need to focus on the browser group smartphone users (n=289) to avoid confounding. 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. To assess the impact this may have had on our findings, we conducted a post-hoc power analysis drawing on recommendations by Lakens (2021), using the G*Power calculator (version 3.1.9.7; Faul et al. 2007; 2009). The results indicated that we would have needed to observe odds ratios of around 1.50 in order to detect a significant effect with the available sample size. For the predictors we investigated with odds ratios smaller than this, therefore, we did not have sufficient power to support the analysis. On this basis, we advise caution in drawing firm conclusions on the basis of our findings. Note, however, that in any case, the size of the (significant) effects of privacy concerns on hypothetical willingness observed in this study were small (see Chen et al. (2010) for a discussion of the size of odds ratios in relation to Cohen's *d*). It is not clear whether larger effects of privacy concerns would be more likely to be observed with a larger sample, or whether a larger sample would improve the likelihood of detecting small effects as significant. These remain open questions for future research and studies investigating these relationships further should be conscious of the need to ensure sufficient analytic power, factoring in the risk of sample depletion due to low participation rates.

6. CONCLUSIONS

Though the power of our analyses of actual compliance with task requests was compromised, we tentatively conclude – as the privacy paradox implies – that neither general privacy concerns, nor perceptions of the sensitivity of (discomfort sharing) different types of data

appear to be definitive in the decision to participate in MDCT via an app. Nevertheless, privacy concerns are clearly part of the story that need to be addressed to help reassure research participants and guarantee the future success of app-based surveys. According to the Leverage-Salience Theory of survey nonresponse (Groves et al. 2000) and Social Exchange Theory (Dillman et al. 2014), data privacy concerns represent just one of a number of potentially salient considerations that may influence participation decisions. Further research is needed to investigate the nature of concerns, the strength with which they are held in different contexts, and how they may be counterbalanced by other survey design features. Attention should be paid to the measures of privacy concern used to ensure respondents properly understand them, facilitate comparisons across studies, and maximize their correspondence with measures of behavioral outcomes of interest. Experimental designs offer clear advantages in the attempt to disentangle multiple confounded influences on respondent behavior, but the ecological validity that is lost when focusing solely on hypothetical willingness to complete MDCT may be more problematic for advancing understanding in this field. We welcome future studies that investigate how privacy concerns and perceptions of data sensitivity influence participation decisions in actual smartphone and app-based surveys.

Data Availability

The data analyzed have not yet been archived for public use. Please contact the corresponding author for more information about the availability of the data and software code used in the analyses presented.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Online Supplementary Materials

Supporting analyses and additional tables referred to in the text are available in the online supplemental material.

8. REFERENCES

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*(5), 888–918. DOI : https://doi.org/10.1037/0033-2909.84.5.888

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.

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

Chen, H., Cohen, P., and Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics – Simulation and Computation*, 39: 860-864.

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.

Drew, D.A., Nguyen, L.H., Steves, C.J., Menni, C., Freydin, M., Varsavsky, T., Sudre, C.H., Cardoso, J., Ourselin, S., Wolf, J., Spector, T.D., Chan, A.T., and COPE Consortium. (2020). Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science*, Vol. 368(6497): 1362-1367. DOI: 10.1126/science.abc0473

Elevelt, A., Bernasco, W., Lugtig, P., Ruiter, S., & Toepoel, V. (2021). 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.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Groves, R.M., and Couper, M.P. (1998). *Nonresponse in Household Interview Surveys.* New York: Wiley.

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.

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.

Lakens, Daniel. 2021. "Sample Size Justification." *PsyArXi*v. January 4. oi:10.31234/osf.io/9d3yf.

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.

Lugtig, P. and V. Toepoel. 2016. "The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey: Effects on Survey Measurement Error." *Social Science Computer Review*, 34: 78–94. DOI: https://doi.org/10.1177/0894439315574248.

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.

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.

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.

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.

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

Wang, R., Fanglin C., Zhenyu C., Tianxing L., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., and Campbell, A.T. (2014.) "StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones." *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3–14.

Wenz, A., Jäckle, A., and Couper, M.P. (2017). Willingness to use mobile technologies for data collection in a probability household panel. ISER Working Paper 2017-10. www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/ 2017-10.pdf.

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.

Appendix Tables

Appendix Table 1. Question wording

General 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)
- 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 (discomfort):

- 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)

Appendix Table 2. Presentation and wording of mobile data collection tasks

Task	Location	Details of presentation and wording
Download a survey app	Wave 2 Invitation	Respondents in the browser group from wave 1 were mailed an invitation to take part in the 'second phase of the study' about their experiences of the election campaign. The invitation to download the app was worded as follows: *"In order to make your participation in this survey more enjoyable, easier and more fun, we are now offering you the possibility of completing the questionnaire through the "civique.org" mobile application, which you can download to your smartphone or tablet. The questionnaire is divided into small modules of 2 to 3 minutes allowing you to complete it at your own pace. The total duration of the questionnaire is about fifteen minutes."* As for the app group at wave 1, the letter provided a QR code to link directly to the respondent's app store (depending on the operating system detected). The letter went on to say: *"If the mobile application is not for you, you can still access the questionnaire on your web browser. To do this, enter the address www.selectscivique.ch/survey and your login details."*
Photo tasks	Wave 2 Questionnaire	There were three photo tasks in wave 2, the wording of which is shown below. Each task was initiated within the app, which accessed the camera on the device and allowed respondents to take multiple photos until they were satisfied with the image selected, and then upload the selected photo within the app.
Photo Task 1	Wave 2, Module 1 'Let's go'	*"Before answering the questions for wave 2 of the Selects-Civique study, could you help us better understand in which context you are using the Civique application and completing our questionnaires? This information will help us to improve the design of future scientific studies with smartphone applications.* *Take a photo of your surroundings and upload it to the Civique app to show us where you are now!* *Some important rules: No selfies! Nor photos of people you know personally."*
Photo Task 2	Wave 2, Module 8 'The election campaign: Take a photo!'	*"In order to help us understand the content of the political advertising to which you have been exposed in recent days, could you take a photo, either of an election poster in your vicinity or of an advertisement received by mail.* *Yes/ No, but I can do it later/ No, I'd prefer not to.* *Take a photo, either of an election poster in your area or of an advertisement received by mail."*

Appendix Table 2 Continued.

Photo Task 3	Wave 2, Module 10 'Your evaluation of the survey'	*"As explained at the outset, we are interested in improving the design of future scientific studies with smartphone applications. To do this, we would like to know where our participants filled out the Selects-Civique questionnaires.* *Could you take a picture of your surroundings to show us where you are now?* *Yes/No* *Take a picture of your surroundings.* *Some important rules: No selfies! No photos of people you know personally."*
Activate GPS tracking	Wave 3, Module 1 'To Begin…'	Respondents were already informed in the data privacy and confidentiality statement/ consent form they had to sign when they first used the app that GPS data would be collected in the study. Specifically, the GPS location at which uploaded photographs were taken (though in practice this data were not collected) and the GPS location of their mobile device each time they used the app. For the app to access these data, respondents had to adjust the settings for the Civique.org app on their phone to allow location access. App respondents were shown the following text: *"As explained in the consent form, the Civique app is able to record your location each time you use it. To enable this feature, please allow location services in your phone settings. Once activated, the app will only capture your location when you are completing the survey modules. To continue, please search for "Location" in your phone settings to authorize the service."* Within the app location settings, respondents could choose between the following options: to allow location access 'never', 'ask next time', 'while using the app' or 'always' and were shown the following App explanation: "This app will only track your location when you are filling surveys that require it.". Thus, GPS coordinates were captured by the app only for those respondents who explicitly adjusted their settings as described and only when the app was in use (i.e. there was no continual tracking of GPS location over a specified time period).

Notes. Consent to the collection of photographs and GPS 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) and the same statement was available to browser respondents via the survey log-in page.

Appendix Table 3. Descriptive statistics for respondent characteristics

	All (N=644)		Browser Group Only (N=344)		Browser Group Smartphone Users Only (N=290)	
	%	n	%	n	%	n
Sociodemographic variables:						
Respondent sex						
Female	50.8	327	49.7	173	50.0	145
Male	49.2	317	50.3	171	50.0	145
Respondent age						
Aged 18-30	21.4	138	22.4	77	26.6	77
Aged 31-55 years	46.6	300	44.8	154	49.0	142
Aged 56+ years	32.0	206	32.8	113	24.5	71
Marital Status						
Married or in a legal partnership	52.5	338	50.0	172	47.2	137
Single, divorced, separated, widowed	47.5	306	50.0	172	52.8	153
Household size						
1 member	15.8	102	18.3	63	17.2	50
2 members	31.4	202	30.5	105	28.3	82
3 members or more	52.8	340	51.3	176	54.5	158
Area of residence						
Village or in the countryside	27.3	176	26.7	92	25.9	75
City or town center/suburbs	72.7	468	73.3	252	74.1	215
Level of education						
Primary or Secondary	45.2	291	43.3	149	57.2	166
Tertiary	54.8	353	56.7	195	42.8	124
Main activity						
In paid work	60.6	390	62.8	216	68.6	199
Not in paid work	39.4	254	37.2	128	31.4	91
Political interest						
Interested in politics	62.6	403	63.7	219	61.7	179
Not very/at all interested	37.4	241	36.3	125	38.3	111
Internet usage variables:						
Frequency of internet use						
Uses Internet several times a day	74.5	480	73.8	254	81.0	235
Uses Internet once a day or less	25.5	164	26.2	90	19.0	55
Devices used to access Internet						
Uses a smartphone	84.9	547	84.3	290	100	290
Does not use a smartphone	15.1	97	15.7	54	0	
Uses fewer than 4 devices	73.6	474	74.7	257	70.0	203
Uses 4 or more devices	26.4	170	25.3	87	30.0	87
Smartphone operating system[1]						
IOS	56.3	308	53.4	155	53.4	155
Android or Windows	43.7	239	46.6	135	46.6	135
Internet skills						
Finds it exciting to try out new tech	66.5	428	65.7	226	70.0	203
Does not agree	33.5	216	34.3	118	30.0	87
Feels able to solve tech problems	69.1	445	68.0	234	73.1	212
Does not agree	30.9	199	32.0	110	26.9	78
Number of smartphone activities						
Min		1		-		1
Max		13		-		13
Median		10		-		9
Mode		12		-		10
Mean (SE)		8.74 (.12)		-		8.69 (.17)
N		544		-		289

Appendix Table 4. Descriptive statistics for measures of conc erns about online data privacy and discomfort about sharing different data types by treatment group and response device (unweighted)

	(1) All n=644		(2) Browser Group All n=344		(3) App Group All n=300		(4) Browser Group PC Respondents n=220		(5) Browser Group Mobile Respondents n=124		(6) App Group App Respondents n=219		(7) App Group Browser Respondents n=81	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE) p^1	%	(SE)	%	(SE) p^2
General concerns about online data privacy:														
Concerned 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)
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)
Concerns about consequences of sharing data:														
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)
Concerned about data privacy (combined measure)	73.0	(1.8)	74.4	(2.4)	71.3	(2.6)	72.7	(3.0)	77.4	(3.8)	69.4	(3.1)	76.5	(4.7)
Mean score concerns about online data privacy	3.86	(.04)	3.89	(.05)	3.82	(.05)	3.86	(.06)	3.93	(.09)	3.80	(.06)	3.87	(.10)
Not comfortable sharing data types 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)†
Mean score 'comfort sharing data requested in the study'[4]	2.86	(.04)	2.88	(.06)	2.84	(.07)	2.84	(.07)	2.96	(.07)	2.75	(.08)	3.07	(.13)

Notes. [1] Propensity-score weighted comparison between (4) and (5). [2] Weighted comparison between (6) and (7). SE = robust standard errors; † p<0.1, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; [3] Mean of four measures of data privacy concerns. [4] Mean score of comfort sharing administrative data, data on political opinions and data about smartphone or tablet are used.

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

Supplementary Material

A. **Output of Principal Components Analyses of Data Privacy Measures**
B. **Descriptive Statistics for the Composite Data Privacy Measures**
C. **Coding of Variables in Logistic Regressions**
D. **Preliminary Analyses and Construction of Weights**
E. **Correlates of Data Privacy Concerns – Supplementary Tables**

A. Output of Principal Components Analyses of Data Privacy Measures

1. Concerns about online data privacy

We computed a composite measure based on the mean of four measures of concerns about online data privacy. The decision to scale the items was supported by a Principal Component Analysis (output below), which extracted one component with an eigenvalue of 2.79, accounting for 69.6% of the variance in reported attitudes. All four items loaded strongly and positively on this component. Reliability Analysis for the scale gave a value for Cronbach's alpha of 0.860. Descriptive statistics for the composite measure are available in Appendix B of the online supplement). The mean score for the full sample was 3.86 (SE = 0.04). However, the variable was strongly negatively skewed (see Figure B1 in Appendix B), with the results of the Kolmogorov-Smirnov test confirming that scores were significantly non-normal ($D(644) = 0.133$, $p < .001$). For this reason, the decision was taken to dichotomize the variable for use as a dependent variable in multivariable analyses (described below). Those scoring greater than 3.5 on the composite measure were coded as 1 (indicating high levels of concern – i.e., greater use of categories 4 and 5 on the response scale), and those scoring below 3.5 coded as 0.

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Concerned about websites and apps collecting personal information	3.62	1.043	644
When browsing websites/using apps: concerned info will be shared with a 3rd party	3.84	1.084	644
When browsing websites/using apps: concerned info will be used for targeted ads	3.82	1.166	644
When browsing websites/using apps: concerned that R identity will be stolen	4.15	1.107	644

Correlation Matrix

	Concerned about websites and apps collecting personal information	When browsing websites/using apps: concerned info will be shared with a 3rd party	When browsing websites/using apps: concerned info will be used for targeted ads	When browsing websites/using apps: concerned that R identity will be stolen
Concerned about websites and apps collecting personal information	1.000	.680	.551	.522
When browsing websites/using apps: concerned info will be shared with a 3rd party	.680	1.000	.639	.612
When browsing websites/using apps: concerned info will be used for targeted ads	.551	.639	1.000	.559

When browsing websites/using apps: concerned that R identity will be stolen	.522	.612	.559	1.000

Communalities

	Initial	Extraction
Concerned about websites and apps collecting personal information	1.000	.682
When browsing websites/using apps: concerned info will be shared with a 3rd party	1.000	.782
When browsing websites/using apps: concerned info will be used for targeted ads	1.000	.677
When browsing websites/using apps: concerned that R identity will be stolen	1.000	.644

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.785	69.627	69.627	2.785	69.627	69.627
2	.488	12.209	81.836			
3	.434	10.838	92.674			
4	.293	7.326	100.000			

Component Matrix

	Component 1
Concerned about websites and apps collecting personal information	.826
When browsing websites/using apps: concerned info will be shared with a 3rd party	.884
When browsing websites/using apps: concerned info will be used for targeted ads	.823
When browsing websites/using apps: concerned that R identity will be stolen	.803

Notes. Extraction Method: Principal Component Analysis (1 component extracted).

2. Perceptions of the sensitivity of the data types requested in the study ('discomfort')

Respondents were asked how 'comfortable' (at ease) respondents felt about university researchers having access to different types of personal information for academic research purposes. We derived a composite indicator of 'discomfort' sharing the three data types actually requested in the study: *administrative data from population registers*, *data on political opinions*, and *data about how you use your smartphone or tablet*. The three measures were positively and significantly correlated with one another. A Principal Component Analysis confirmed a single component that accounted for 71.8% of the variance (output below), and Reliability Analysis gave a value for Cronbach's alpha of 0.803. The

mean score on this indicator for the full sample 2.86 (Standard Error [SE]=0.04). The distribution of the variable was not quite normal, with a significantly higher proportion of respondents than would be expected having a mean score of 5 (see Figure B2 in Appendix B). To create a dichotomous measure based on the composite score, we coded those with a mean score greater than 3.0 as 1 (implying greater use of the negative end of the scale for all items), and those with a score of 3.0 or lower as 0.

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Comfort sharing administrative data	3.05	1.271	644
Comfort sharing political opinions	2.68	1.317	644
Comfort sharing smartphone/tablet data	2.84	1.376	644

Correlation Matrix

	Comfort sharing administrative data	Comfort sharing political opinions	Comfort sharing smartphone/tablet data
Comfort sharing administrative data	1.000	.559	.586
Comfort sharing political opinions	.559	1.000	.586
Comfort sharing smartphone/tablet data	.586	.586	1.000

Communalities

	Initial	Extraction
Comfort sharing administrative data	1.000	.710
Comfort sharing political opinions	1.000	.711
Comfort sharing smartphone/tablet data	1.000	.733

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.154	71.803	71.803	2.154	71.803	71.803
2	.441	14.691	86.494			
3	.405	13.506	100.000			

Component Matrix

	Component 1
Comfort sharing administrative data	.843
Comfort sharing political opinions	.843
Comfort sharing smartphone/tablet data	.856

Notes. Extraction Method: Principal Component Analysis (1 component extracted).

B. Descriptive Statistics for the Composite Data Privacy Measures

1. Concerns about online data privacy

Descriptives (Composite measure of data privacy concerns)

		Statistic	Std. Error
Mean		3.8552	.03614
95% Confidence Interval for Mean	Lower Bound	3.7842	
	Upper Bound	3.9262	
Median		4.0000	
Variance		.841	
Std. Deviation		.91721	
Interquartile Range		1.25	
Skewness		-.842	.096
Kurtosis		.161	.192

Tests of Normality (Composite measure of data privacy concerns)

Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Statistic	df	Sig.	Statistic	df	Sig.
.133	644	.000	.924	644	.000

Notes. a. Lilliefors Significance Correction

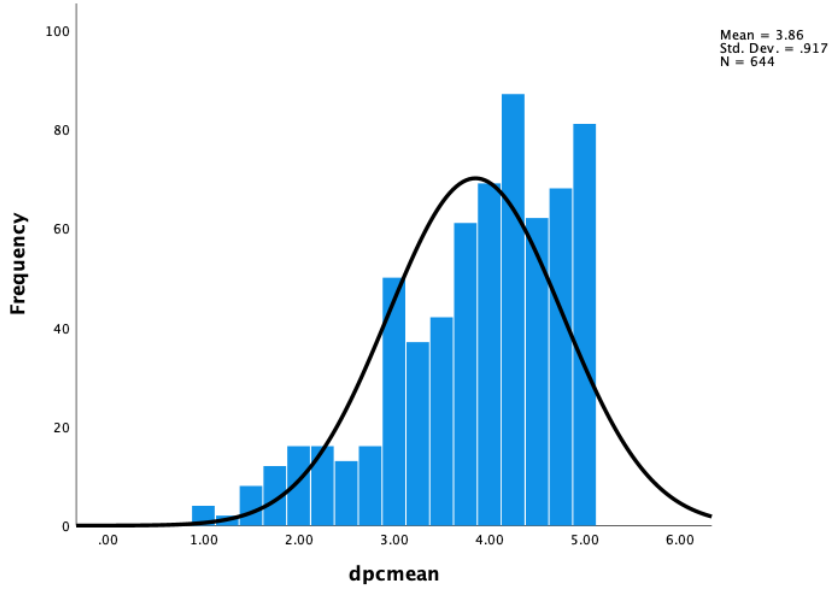

Figure B1: Histogram of composite measure of concerns about online data privacy

2. Discomfort sharing data types requested in the study

Descriptives (Composite measure of discomfort)

		Statistic	Std. Error
Mean		2.8602	.04413
95% Confidence Interval for Mean	Lower Bound	2.7736	
	Upper Bound	2.9469	
Median		2.6667	
Variance		1.254	
Std. Deviation		1.11997	
Interquartile Range		1.67	
Skewness		.235	.096
Kurtosis		-.880	.192

Tests of Normality (Composite measure of discomfort)

Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Statistic	df	Sig.	Statistic	df	Sig.
.102	644	.000	.960	644	.000

Notes. a. Lilliefors Significance Correction

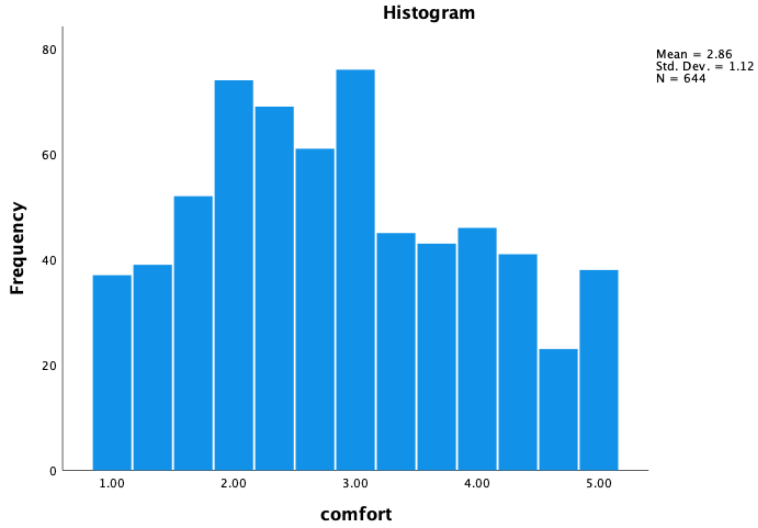

Figure B2: Histogram of composite measure of comfort sharing data types requested in the study

C. Coding of Variables in Logistic Regressions

Appendix Table SM1. 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 once a day or less (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 greater than 3.5 = 1 (concerned), else=0 (less concerned)
Comfort sharing data (mean)	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
Comfort sharing data (dichotomized)	Mean as above but dichotomized: scores greater than 3.0 = 1 (concerned), else=0 (less concerned)

D. Preliminary Analyses and Construction of Weights

The experimental design involved the random assignment of sample members in equal parts to two treatments: Group 1 – the Browser Group – was invited to participate via a web browser, while Group 2 – the App Group – was invited to participate via the Civique.org app. Despite the random assignment of sample members to the two treatment groups at wave 1, respondents could choose which device/ software to use to respond (in Group 1, between a PC or 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 where there was evidence for measurement differences. To do so, we compared answers given to the data privacy measures by respondents on different devices separately in each treatment group, using Chi-square tests for the four individual measures, and Independent Samples t-tests for the composite measures of data privacy concern and comfort sharing the data types collected in the study with university researchers. However, to address differences in the composition of the samples responding on different devices and software (see Appendix Table SM3), we adjusted the comparisons across device/software groups using a propensity score weight.

To compute the weight, we used the estimated response propensities from logistic regression models predicting the probability of responding with one device/software compared to its alternative in each treatment group, given a number of observed characteristics (the sociodemographic variables from the sampling frame described above). In the browser group (group 1), the dependent variable was responding on a mobile browser (coded 1) compared to a PC browser (coded 0). In the app group (group 2), the dependent variable was responding on a PC or mobile browser (coded 1) compared to responding via the app (coded 0). Note that for the app group, we combined PC and mobile browser respondents because there were only 16 respondents in this group who responded on a mobile browser. Coefficients from the two models are shown in Appendix Table 2 and show that in

the browser group, mobile respondents differed significantly from PC browser respondents on two register variables: they were more likely to be female ($p<0.1$), and less likely to be aged 56 or older. In the app group, browser respondents were more likely to be aged 56 or older ($p<0.1$), and less likely to live in households with three or more members.

Within each group, for respondents in the predicted category, the weight is equal to p, while for respondents in the reference category, the weight is equal to

$$p_{ref} = 1 - \Pr(Y_i = 1 | X_i)^{-1} = 1 - \frac{1}{\Pr(Y_i = 1 | X_i)}$$

The weight was applied for all cross-device comparisons within the treatment groups, shown in Appendix Table 3 and was effective at balancing the sub-samples on the covariates analysed.

Note that we also compared respondents using the device/software intended by the experimental design (i.e. respondents using a PC in the browser group and respondents using the app in the app group) using the same approach. The dependent variable in the logistic regression model was coded 1 if the respondent used the app, and 0 if they used a PC. Results are shown in Appendix Table 2, column 3. There were no statistically significant differences between these groups on the register variables analysed, so we did not compute a weight for the purpose of comparing these groups.

Note that there were also differences across groups responding with different devices/ software in the proportion of respondents with missing data on the five data privacy measures (not shown in table). A total of 7.6 percent of app respondents skipped the module in which the questions appeared, compared with 4% of PC users ($p=0.056$). Missing rates for mobile browser respondents were comparable with those for app respondents at 7.5% (significantly different compared with the rate for the PC respondents: $X^2(1) = 3.53$; $p<0.05$).

Appendix Table SM2 – Logistic regression coefficients predicting chosen response device at wave 1 by treatment group.

	(1) Browser group – Responding on a mobile				(2) App group – Responding on a browser				(3) All groups – Responding on the app			
	$\widehat{\beta}$	*p*	SE	*Exp B*	$\widehat{\beta}$	*p*	SE	*Exp B*	$\widehat{\beta}$	*p*	SE	*Exp B*
Sociodemographic variables:												
Female	.462	†	.236	1.587	.040		.274	1.041	.233		.197	1.262
Age[1]: 31-55 years	-.396		.335	.673	.399		.502	1.491	-.006		.324	.994
Age: 56+ years	-1.123	**	.431	.325	1.144	†	.596	3.140	-.601		.392	.548
Married[2]	.020		.330	1.021	-.083		.442	.921	.042		.302	1.042
Divorced/ Separated	.064		.415	1.066	.023		.534	1.023	-.275		.385	.760
Household size[3]: 2 members	-.614		.380	.541	-.673		.449	.510	.425		.358	1.529
Household size: 3 members or more	-.361		.364	.697	-1.010	*	.472	.364	.573		.354	1.774
Urban residence	-.173		.264	.841	.532		.329	1.702	-.216		.221	.860
Constant	.177		.446	1.194	-1.225	*	.598	.294	-.212		.422	.809
Model X^2 (8)			22.53**				25.58***				19.01*	
Hosmer & Lemeshow X^2 (8)			4.67				1.58				4.46	
Nagelkerke R^2			.09				.12				.06	
Observations (*n*)			344				300				439[4]	

Notes. [1]Age (ref. 18-30 years old); [2]Marital status (ref. single/never married) [3]Household size (ref. single persons);). [4]Includes only group 1 respondents on a PC and group 2 respondents on the app. $\hat{\beta}$ = unstandardized beta coefficient; Exp B = Exponent B. p= p-value, † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

E. Correlates of Data Privacy Concerns – Supplementary Tables

Appendix Table SM3. Descriptive statistics for measures of concerns about online data privacy by socio-demographics, internet and smartphone usage variables.

	n=644	(1) Concerned websites & apps collect personal info % (SE)	(2) Concerned data will go to 3rd parties % (SE)	(3) Concerned data will be used to send targeted ads % (SE)	(4) Concerned identity might be stolen % (SE)	(5) Concerned about data privacy (combined) % (SE)
Socio-demographics variables:						
Male	317	56.2 (2.8)	66.6 (2.7)	63.4 (2.7)	74.1 (2.5)†	68.1 (2.6)**
Female	327	61.2 (2.7)	68.5 (2.6)	69.4 (2.6)†	79.5 (2.2)	77.1 (2.3)
Age						
Aged 18-30	138	47.8 (4.3)**	63.0 (4.1)	54.3 (4.3)***	68.8 (4.0)***	63.8 (4.1)**
Aged 31-55	300	58.3 (2.9)	66.0 (2.7)	66.0 (2.7)	74.0 (2.5)	72.0 (2.6)
Aged 56+	206	66.5 (3.3)	72.8 (3.1)	75.2 (3.0)	86.4 (2.4)	80.6 (2.8)
Marital status						
Married or in partnership	338	63.0 (2.6)*	69.8 (2.5)	68.0 (2.5)	80.8 (2.1)*	76.0 (2.3)†
Not married or in a partnership[1]	306	53.9 (2.9)	65.0 (2.7)	64.7 (2.7)	72.5 (2.6)	69.6 (2.6)
Household size						
1 person	102	54.9 (5.0)	62.7 (4.8)	71.6 (4.5)*	72.5 (4.4)*	70.6 (4.5)†
2 persons	202	62.9 (3.4)	69.8 (3.2)	72.3 (3.2)	83.7 (2.6)	78.7 (2.9)
3 persons or more	340	57.4 (2.7)	67.6 (2.5)	61.5 (2.6)	74.1 (2.4)	70.3 (2.5)
Area of residence						
Village or in the countryside	176	58.0 (3.7)	68.8 (3.5)	63.6 (3.6)	79.0 (3.1)	75.0 (3.3)
City or town centre/suburbs	468	59.0 (2.3)	67.1 (2.2)	67.5 (2.2)	76.1 (2.0)	72.2 (2.1)
Education						
Primary or Secondary	291	60.1 (2.9)	69.4 (2.7)	69.8 (2.7)†	82.8 (2.2)**	77.0 (2.5)*
Tertiary	353	57.5 (2.6)	66.0 (2.5)	63.7 (2.6)	72.0 (2.4)	69.7 (2.5)
Main activity						
In paid work	390	58.2 (2.5)	65.9 (2.4)	65.1 (2.4)	75.4 (2.2)	72.3 (2.3)
Not in paid work	168	59.4 (3.1)	70.1 (2.9)	68.5 (2.9)	79.1 (2.6)	74.0 (2.8)
Interested in politics	403	64.0 (2.4)***	70.7 (2.3)*	70.0 (2.3)*	77.2 (2.1)	75.7 (2.1)*
Not interested in politics	241	49.8 (3.2)	62.2 (3.1)	60.6 (3.2)	76.3 (2.7)	68.5 (3.0)
Internet usage variables:	**644**					
Uses Internet several times a day	480	57.9 (2.3)	67.3 (2.1)	65.0 (2.2)	75.0 (2.0)†	71.9 (2.1)
Uses Internet once a day or less	164	61.0 (3.8)	68.3 (3.6)	70.7 (3.6)	82.3 (3.0)	76.2 (3.3)
Uses fewer than 4 devices	474	57.8 (2.3)	66.2 (2.2)	67.3 (2.2)	75.9 (2.0)	71.7 (2.1)
Uses 4 or more devices	170	61.2 (3.7)	71.2 (3.5)	64.1 (3.7)	79.4 (3.1)	76.5 (3.3)
Devices used to access internet						
Desktop and/or laptop	574	58.0 (2.1)	67.9 (1.9)	65.0 (2.0)*	75.6 (1.8)*	72.6 (1.9)
No desktop/laptop	70	64.3 (5.8)	64.3 (5.8)	78.6 (4.9)	87.1 (4.0)	75.7 (5.2)
Smartphone	547	57.8 (2.1)	66.5 (2.0)	64.2 (2.1)**	75.3 (1.8)*	71.1 (1.9)*
No smartphone	97	63.9 (4.9)	73.2 (4.5)	79.4 (4.1)	85.6 (3.6)	83.5 (3.8)
Tablet	258	59.7 (3.1)	68.6 (2.9)	66.7 (2.9)	81.0 (2.4)*	77.1 (2.6)†
No tablet	386	58.0 (2.5)	66.8 (2.4)	66.3 (2.4)	74.1 (2.2	70.2 (2.3)
Other devices used	157	63.7 (3.9)†	74.5 (3.5)*	73.2 (3.5)*	84.7 (2.9)**	80.9 (3.2)*
No other devices used	487	57.1 (2.2)	65.3 (2.2)	64.3 (2.2)	74.3 (2.0)	70.4 (2.1)
Exciting to try out new tech	245	52.2 (3.2)**	67.3 (3.0)	62.0 (3.1)†	72.7 (2.9)*	68.6 (3.0)*
Does not agree	399	62.7 (2.4)	67.7 (2.3)	69.2 (2.3)	79.4 (2.0)	75.7 (2.2)
Capable of solving tech problems	264	57.2 (3.1)	68.2 (2.9)	67.4 (2.9)	72.7 (2.7)*	72.3 (2.8)
Does not agree	380	59.7 (2.5)	67.1 (2.4)	65.2 (2.4)	79.7 (2.1)	73.4 (2.3)

Appendix Table SM3 continued.

Smartphone usage variables:	**547**					
Uses SP several times a day/hour	496	56.3 (2.2)*	66.3 (2.1)	63.3 (2.2)	75.4 (1.9)	70.2 (2.1)
Uses SP once a day or less often	48	72.9 (6.5)	68.8 (6.8)	72.9 (6.5)	75.0 (6.3)	81.3 (5.7)
Smartphone activities…						
Browses social media	352	53.7 (2.7)*	64.8 (2.5)	60.8 (2.6)*	75.0 (2.3)	60.8 (2.5)*
Does not browse social media	192	65.1 (3.4)	69.8 (3.3)	70.3 (3.3)	76.0 (3.1)	70.3 (3.2)
Posts on social media	271	53.5 (3.0)*	66.1 (2.9)	61.3 (3.0)†	76.8 (2.6)	70.8 (2.8)
Does not post on social media	273	61.9 (2.9)	67.0 (2.9)	67.0 (2.9)	74.0 (2.7)	71.4 (2.8)
Makes purchases	338	54.4 (2.7)*	67.5 (2.6)	62.7 (2.6)*	74.3 (2.4)	70.4 (2.5)
Does not make purchases	206	63.1 (3.4)	65.0 (3.3)	66.5 (3.3)	77.2 (2.9)	72.3 (3.1)
Does online banking	260	55.8 (3.1)	68.1 (2.9)	64.6 (3.0)*	76.9 (2.6)	71.9 (2.8)
Does not do online banking	284	59.5 (2.9)	65.1 (2.8)	63.7 (2.9)	73.9 (2.6)	70.4 (2.7)
Installs apps	418	55.7 (2.4)†	67.5 (2.3)	61.7 (2.4)*	76.1 (2.1)	70.8 (2.2)
Does not install apps	126	64.3 (4.3)	63.5 (4.3)	72.2 (4.0)	73.0 (4.0)	72.2 (4.0)
Use apps with location services	459	55.3 (2.3)**	65.6 (2.2)	63.2 (2.3)**	75.4 (2.0)	70.6 (2.1)
Does not use apps with location	85	70.6 (5.0)	71.8 (4.9)	69.4 (5.0)	75.3 (4.7)	74.1 (4.8)
Connects to Bluetooth devices	273	52.7 (3.0)*	66.7 (2.9)	58.2 (3.0)**	74.7 (2.6)	67.4 (2.8)†
Does not connect to Bluetooth	271	62.7 (2.9)	66.4 (2.9)	70.1 (2.8)	76.0 (2.6)	74.9 (2.6)
Plays games	232	56.0 (3.3)	66.8 (3.1)	64.2 (3.2)*	75.9 (2.8)	70.3 (3.0)
Does not play games	312	59.0 (2.8)	66.3 (2.7)	64.1 (2.7)	75.0 (2.5)	71.8 (2.6)
Listens to music/ watch videos	435	54.9 (2.4)**	65.3 (2.3)	61.8 (2.3)*	74.0 (2.1)	69.7 (2.2)
Does not listen to music/videos	109	68.8 (4.5)	71.6 (4.3)	73.4 (4.3)	80.7 (3.8)	77.1 (4.1)
Other activities	185	51.9 (3.7)*	68.1 (3.4)	61.1 (3.6)	70.8 (3.4)†	68.1 (3.4)
No other activities	359	60.7 (2.6)	65.7 (2.6)	65.7 (2.5)	77.7 (2.2)	72.7 (2.4)
IOS operating system	308	55.8 (2.8)	68.8 (2.6)	66.6 (2.7)	75.6 (2.4)	72.7 (2.5)
Android or Windows OS[2]	239	60.3 (3.2)	63.6 (3.1)	61.1 (3.2)	74.9 (2.8)	69.0 (3.0)

Notes. [1]Single/Divorced/ Separated/ Widowed; [2]only 6 cases with windows OS; [3]Browser respondents only (not asked on app). SE = robust standard errors; † p<0.1, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Differences tested using Chi-Square Tests of Association.

Appendix Table SM4. Logistic regression analyses predicting concerns about data privacy and the consequences of sharing data online

	(1) Concerned websites & apps collect personal info				(2) Concerned data will go to 3rd parties				(3) Concerned data will be used to send targeted ads				(4) 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*
Sociodemographic variables:																
Female	0.26		0.17	1.30	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	0.22		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	0.47		0.33	1.60	0.80	*	0.33	2.22	0.71	†	0.37	2.03
Married	0.05		0.22	1.05	-0.05		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	0.33		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	0.41		0.29	1.51	-0.15		0.30	0.86	0.28		0.31	1.33
Urban residence	0.03		0.19	1.03	-0.12		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	-0.22		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	-0.13		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	0.40	**	0.18	1.49	0.39	*	0.18	1.48	0.03		0.21	1.03
Internet usage variables:																
Uses Internet once a day or less	-0.15		0.22	0.86	-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	0.25		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	0.00		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	0.17		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	-0.25		0.30	0.78	-0.53		0.32	0.59	-0.25		0.37	0.78
Assigned to Browser Group	0.09		0.17	1.10	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	0.23		0.50	1.25	0.29		0.52	1.34	1.05	†	0.58	2.85
Nagelkerke R^2	0.07				0.04				0.08				0.09			
Observations	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, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix Table SM5. 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*
Data privacy concerns (mean)	.738	***	.112	2.092	.667	**	.101	1.948	.696	***	.117	2.006	.557	***	.121	1.746
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 once a day or less	.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 Browser Group	-.031		.176	.969	-.214		.173	.807	-.230	***	.179	.794	-.123		.189	.884
Constant	-3.310	***	.658	.037	-2.829	***	.618	.059	-4.086		.690	.017	-3.802	***	.721	.022
Nagelkerke R^2	.150				.167				.147				.108			
Observations	644				644				644				644			

Appendix Table SM5 ***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*
Data privacy concerns (mean)	.569	***	.104	1.767	.438	**	.094	1.549	.530	***	.095	1.700	.795	***	.128	2.214
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 once a day or less	.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 Browser Group	-.060		.170	.942	-.076	*	.170	.927	.051		.168	1.052	-.057		.195	.945
Constant	-3.413	***	.633	.033	-2.080	***	.594	.125	-2.672	***	.596	.069	-4.312	***	.696	.013
Nagelkerke R^2	.132				.098				.184				.181			
Observations	644				644				644				580			

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=580). $\hat{\beta}$ = unstandardized beta coefficient; Exp B = Exponent B. p= p-value, † p<0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.