ABSTRACT

Title of Dissertation: NONPARTICIPATION ISSUES RELATED TO PASSIVE DATA COLLECTION

Alexandra Marie Brown Breslin, Doctor of Philosophy, 2024

Dissertation directed by: Assistant Professor Chris Antoun, Joint Program in Survey Methodology, and
Professor Stanley Presser, Joint Program in Survey Methodology

New passive data collection techniques on smartphones allow for the direct observation of a participant's behavior and environment in place of self-reported information. However, such studies are not appealing to all people, especially those with higher security concerns. The current work explores the mechanisms that impact a sample member's decision to participate in a passive data collection using three different online panels. The first study explores nonparticipation bias in a financial tracking study and finds evidence of bias in the self-reported measures of financial behaviors, and that prior experience with the research organization positively impacts a sample member's decision to participate. Studies two and three employ deception studies (i.e., the passive data collections were presented as real rather than hypothetical, but no data was passively collected) in which respondents received experimentally varied invitations to participate in a smartphone-based passive data collection. The second study varies the type of data requested and the study topic to understand better how these study components interact. The findings suggest that the type of data requested impacts participation while the study topic does not. The second study utilized video messages presented to all sample members who chose not to participate. These videos asked the sample member to reconsider, varying whether or not they reiterated the data and security measures of the study from the initial invitation. The results suggest that offering a follow-up video increased participation. Finally, the third study experimentally varied the level of control the sample member would have over what

data is shared with researchers during a passive data collection. The findings suggest that an offer of control may not increase participation in app-based passive data collection. The three studies suggest that sample members are more likely to participate in a survey when they have prior experience with such a request and may be converted to participate with a video message, but that the type of data requested greatly impacts the decision to participate. Future work should include replicating these studies with different requested data types and shifting to samples not drawn from online panels.

KEYWORDS: passive data collection, web surveys, smartphones, nonparticipation

NONPARTICIPATION ISSUES RELATED TO PASSIVE DATA COLLECTION

by

Alexandra Marie Brown Breslin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Doctoral Committee:
Assistant Professor Chris Antoun, Co-Chair
Professor Stanley Presser, Co-Chair
Associate Professor Vanessa Frias-Martinez
Professor Florian Keusch
Professor Frauke Kreuter

Preface

Recently, passive data collection techniques on smartphones and PCs have been developed that enable the replacement of self-reports with direct observation of behavior. In contrast to active measurement, once participants have taken the steps to enable the production of the requested data, passive measurement allows for increased granularity in time and space through real-time continuous measurement (Couper 2019). These measurements include the use of sensors in a smartphone to track usage, accessing an accelerometer or GPS tool that the sample member carries, or providing access to financial accounts – such as credit cards (Agarwal et al. 2007) or bank accounts (Angrisani et al. 2023). This dissertation focuses on such passive tools, though its findings may also apply to active measurement.

Passively collecting data offers many benefits and opportunities. Such data collections have the potential to bypass the challenges of self-report surveys by directly capturing the behaviors of study participants when they interact with the device or sensor. Compared to self-reports, collecting this information directly may have much less measurement error (e.g., Revilla et al. 2017; Scherpenzeel 2017) and be less burdensome to respondents (e.g., Geurs, Veenstra, & Thomas 2013). In addition, passively collected data can link to self-report data (e.g., Stier et al. 2020). However, participation rates for passive data collection are lower than for traditional self-report surveys.

Passive data collection participants are recruited in two main ways: either by drawing a fresh sample of participants or using existing panel survey. A review of the literature suggests that recruiting participants through existing panels is the most common method. Passive data collection may use various technologies, for example:

1. Apps that track information from a device,
2. Plug-ins that collect web browsing history and other device information (either continuously or for a point in time) or
3. Proxies that configure to the networks used to connect to the Internet through a given device (e.g., WIFI router).

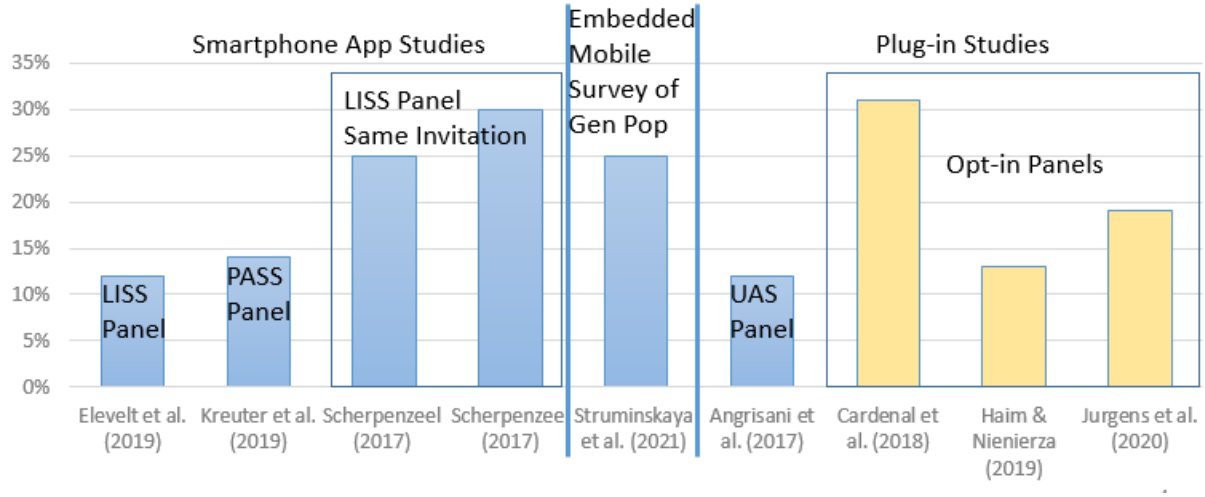
Data collection may occur independently or as part of a larger data collection effort, such as in combination with a self-report survey. The majority of recent passive data collections conducted through apps or plug-ins have been part of larger data collection efforts as ways to reduce burden and/or improve data quality.

Unfortunately, the participation rates of these studies tend to be lower than those of traditional self-report surveys. Participation in a passive data study requires a sample member to accept the invitation and take the step(s) required to share the requested data with researchers. Not all sample members who indicate a willingness to participate in the study will take all the required steps to share the requested data. For example, in two mobile app-based passive data collections, Scherpenzeel (2017) found that 37% of LISS Panel members invited to participate in a Mobility and Time Use Survey (TUS) indicated they would participate. However, only 68% of these participated in the TUS by sharing their geolocation data. Only 81% of those who said they were willing to participate actually participated in the Mobility study by sharing their accelerometer data.

Figure 1 shows participation rates for recent mobile app and plug-in studies. The mobile app studies invite sample members to download an app and include other components such as completing self-report surveys or other tasks such as uploading pictures. The plug-in studies track specific data types – financial transactions or interactions with social media. Besides

Angrisani et al. (2023), the plug-in studies included self-report survey components during the data collection period. Struminskaya et al. (2021) used a different approach, embedding a passive data collection in a mobile web survey. Rather than use an existing panel, the sample members had all previously responded to a general population survey from Statistics Netherlands. Such an embedded passive data collection design may be easier for participants to complete but does not allow for data collection once the web survey is complete.

Figure 1: Final Participation Rates in Mobile App and Plug-In Studies



Authors have speculated that participation rates vary by the burden required in the steps to share the data (Kreuter et al. 2019) and with the sensitivity of the requested data (Struminskaya et al. 2021). Work by Nissenbaum (2010) may help to explain why participation rates are lower for passive data collections than to the associated self-report surveys. Nissenbaum (2010) proposes a framework of contextual integrity to explain when and how individuals may wish to share or conceal their personal information. Nissenbaum posits that when the expected norms associated with the (perceived) appropriate flow of personal information are violated, this becomes a violation of privacy (referred to as a violation of contextual integrity). Key to her formulation is *who is collecting what information under what conditions (e.g., features of the study such as how*

the data is stored/transmitted or for what purpose). The following three chapters explore how the type of information collected, and the conditions of the passive data collection impact participation. The impact of the actors (i.e., who is collecting the data) on participation in a passive data collection is an important subject for future research.

Groves, Cialdini, and Couper's (1992) theoretical framework for understanding survey participation posits that decision-making about survey participation is guided by heuristics irrelevant to the central features of the task of being a respondent. Research suggests the decision to participate is quick and shallow, based on very little information, and subject to rapid change (Groves & Couper 1998; de Heer & de Leeuw 2002) – with little evidence that any one factor positively impacts participation more than another factor. Leverage-salience theory (Groves, Singer, and Corning 2000) posits that the decision process should vary systematically across sample members as the result of a sample member's inclination to accept a fixed component of the survey (e.g., expected survey norms around interaction) and the salience of that component in the survey request.

Bradburn (2015) described the expected interaction between an interviewer and respondent as a "script," defined as a general set of stylized behaviors that are recognized as forming a goal-oriented unit. Nissenbaum (2010) also explores the notion of a script to establish expected norms in specific interactions, such as ordering in a restaurant. A divergence from expectations may result in an unwillingness to participate in the interaction. As surveys shift from the expected norms of questions and responses to passive data collection methods, sample members may question whether their expected norms of surveys still hold.

Acknowledging the changes that passive data collection brings to the interaction between survey organizations and respondents, my work contributes in several ways to understanding nonparticipation within passive data collections that use web-based panels to recruit respondents. My first study explores the nature of nonparticipation bias in a passive data collection of financial transactions, which departs dramatically from the traditional survey request that asks respondents to self-report their spending and saving and gives respondents control over what is reported. My second study explores whether the passive data collection context (defined as the combination of the data requested and the study topic) impacts the decision to participate in a mobile web-based passive data collection and whether addressing data privacy concerns by reiterating the study's data security measures (through a recorded video) reduces nonparticipation. Reiterating the data security information may tie the passive data collection closer to the traditional survey context with which most people are more familiar. Finally, my third study explores whether offering participants control over what data are shared with researchers reduces nonparticipation in a mobile app-based passive data collection.

Findings from this work may also have implications for collecting paradata, data collected 'in the background' to understand the survey process. As web survey platforms expand what can be automatically collected (e.g., latitude/longitude), the ethics of collecting such data without consent ignores the right respondents have in deciding whether or not to participate in the research[1]. Studies that have explored the impact of consent to collect paradata suggest that asking for consent lowers participation, but the results are mixed. In an earlier study, Couper and Singer (2013) found that when collecting traditional paradata such as timing data, keystroke data, mouse click data, and information about the type of interface such as the web browser and screen

[1] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979

resolution, any mention of the paradata collection in the consent request resulted in lower rates of participation. Asking consent to collect other forms of potential paradata - i.e., location – has not yet been tested. Understanding the impact of a request for consent is important because as technology and user-written scripts develop, it is easier or more common to collect different forms of paradata to which sample members might not wish to consent. Thus, my results on ways to increase participation in passive data collections may inform the future of paradata collection as well.

Dedication

This work is dedicated to my family.

Acknowledgement

I am profoundly grateful for all the support and feedback I received while preparing this work.

First, to my advisors. I cannot thank you enough for all the time, guidance, and editorial support. This research would not have come to be without you. Stanley, you have been a consistent presence and support throughout my PhD, beginning with my first PhD seminar class. Your incredible expertise and teaching provided a path to follow, for which I am incredibly grateful. Chris, you supported my research from the earliest stages, brought me in to observe the research process firsthand, and provided a space to discuss my research plans. Your support and guidance have kept me going in this program, and I am extremely grateful for it.

Next, to my full committee. I am honored to be surrounded by such amazing researchers. Frauke, thank you for taking the time to meet with me before I even applied to the PhD program. Your perspective and constant presence at many of the conferences and seminars I attended truly led me down the path toward this degree. Florian, I am so thankful that you taught me the first survey methodology course that I took. Your openness to chatting over the years has been so incredibly supportive. Finally, thank you, Vanessa, for joining my committee and offering such great feedback with fresh eyes.

I am also incredibly grateful to the larger survey methodology community. The years I have spent at JPSM will be cherished. The feedback I received while presenting early iterations of this work at workshops, AAPOR, and CIPHER improved my research plan immensely. As did the

feedback I received while applying for grants and funding. I especially appreciate the feedback I received from the TESS committee on multiple occasions. I also want to thank Ipsos for their incredible financial support in administering the survey used in chapter two.

Thank you to my parents, whose constant support of my desire to learn sent me down this long, winding path of knowledge.

Lastly, I want to thank my husband. So much about our life has changed since I began my PhD journey. But after many years, two kids, a pandemic, and two moves, you have always been there to offer support and listen to the many topics I pursued over the last few years.

Thank you all.

Table of Contents

Abstract .....
Title Page.....
Copyright Statement.....
Preface ii
Dedication and Acknowledgement..... viii
Table of Contents.....x
List of Tables xi
List of Figures xiii
Chapter One: 1
Introduction 1
Literature Review 2
Methods and Data.....7
Results.....12
Discussion.....21
Limitations and Next Steps.....22
Chapter Two:.....25
Introduction 25
Literature Review 26
Methods and Data.....31
Results.....41
Discussion.....45
Limitations and Next Steps.....47
Chapter Three: 50
Introduction 50
Literature Review 52
Methods and Data.....60
Results.....68
Discussion.....73
Limitations and Next Steps.....75
Conclusion: 77
Appendix A: Introduction for Financial Tracking Study 80

Appendix B: Refusal Conversion Questionnaire.....81
Appendix C: Offer of Control Questionnaire 86
References.....90

List of Tables

Chapter One:

- Table 1: Descriptions of Estimates, Sources, and Sample Size……10
- Table 2: Chi-square Comparisons of Measures of Self-Reported Financial Behaviors of Nonparticipants and Participants to the Passive Data Collection……..…13
- Table 3: Pearson Correlation Matrix – Self-Reported Financial Behaviors and Sample Member Attitudes……..16
- Table 4: Logistic Regression Model 1 - Relationship between Self-reported Financial Behaviors, Demographics and Participation…...……...……17
- Table 5: Logistic Regression Model 2 - Relationship between Self-reported Financial Behaviors, Demographics, Attitudes and Participation…..……20

Chapter Two:

- Table 1: Experimental Assignment by Study Topic and Requested Data……33
- Table 2: Requests to be withdrawn by Study Topic and Requested Data……35
- Table 3a: Demographics – Social Network Size Topic Invitations……..…..……37
- Table 3b: Demographics – Smart Phone Topic Invitations……..……..37
- Table 3c: Demographics – Location Topic Invitations……..37
- Table 4: Percent of Respondents Who Report Using Smartphones for Activities other than Phone Calls and Text Messaging, by Invitation Topic……..38
- Table 5: Video Timing……...40
- Table 6: Percent of Respondents Who Agreed to Participate in to Passive Data Collection……...41
- Table 7: Regressions – Initial Participation……...42
- Table 8: Initial and Final Participation Rates, Overall and by Requested Data Type…...43
- Table 9: Initial and Final Participation Rates, by Assigned Recording……44
- Table 10: Participant Demographics, Initial and Final Samples……45

Chapter Three:

- Table 1: Requests to be withdrawn……62
- Table 2: Demographics of sample, by experiment……...…63
- Table 3: Respondent Relationship with Smartphone……64
- Table 4: Open response categorization for why a respondent choose not to participate……...65
- Table 5: Open response categorization for why a respondent choose not to participate, by experimental group……66
- Table 6: Comfort with Installing Apps (for all respondents)……...……69
- Table 7: Concern about sharing personal information over the Internet……...71
- Table 8: Regression Output……..72

List of Figures

Preface:

- Figure 1: Final Participation Rates in Mobile App and Plug-In Studies……...iv

Chapter One:

- Figure 1: Trust in Organization……..……11

Chapter Two:

- Figure 1 – 3: Experiment Invitations……33

Chapter Three:

- Figure 1: Invitations to Studies……..67
- Figure 2: Participation Overall, by Experimental Treatment, and Privacy Concerns…...69

Chapter One: A Nonparticipation Bias Analysis of a Passive Data Collection of Financial Transactions

Abstract

Passively collected data allows researchers to work with highly detailed objective data. However, the passive collection of data, such as financial transactions, may be perceived as a privacy violation, triggering privacy concerns for invited sample members that may result in higher rates of nonparticipation depending on sample member's attitudes and demographics. Using data from a study conducted on the Understanding America Study (UAS) online probability panel, comparisons are made between participants in a passive data collection of financial transactions and nonparticipants in terms of their self-reported financial behaviors, demographics, and attitudes. The results suggest an association between self-reported financial behaviors and participation, which does not fully diminish after controlling for sample member demographics and attitudes. Overall, nonparticipation in a passive data collection of financial transactions results in biased key estimates of interest for such a study.

Introduction

The nonparticipation bias present in a sample mean is the property of an individual statistical estimate and driven by the systematic differences between participants and nonparticipants. This work explores the impact of nonparticipation on key estimates of interest (i.e., self-reported financial behaviors) within a study that passively collected financial transaction data from Understanding America Study (UAS) panelists. As this data allows for access to self-report data for both participants and nonparticipants in the UAS study, it is possible to explore how key estimates, such as self-reported financial behavior, differ between participants and nonparticipants with varying attitudes and demographics. If the results suggest nonparticipation bias may exist for the estimates of financial behavior, this raises concerns for any conclusions drawn using such unique and innovative data.

Angrisani et al. (2023) began to explore representation issues for a financial tracking study conducted on the UAS Panel. They found that younger and more educated panelists participated when compared to their counterparts. The current work extends Angrisani et al. (2023) by utilizing a larger and more extensive dataset from the UAS Panel to explore whether differences

in self-reported financial behaviors exist between UAS Panel members who participated in the study and those who did not participate. The work by Angrisani et al. (2023) invited panelists to participate in a passive data collection conducted on a different platform (Yodlee) than the current work (Plaid) and utilized a much smaller sample of panelists (just over 1,000 panelists compared to over 7,000 panelists in the current work). The current work begins by replicating the analysis of Angrisani et al. (2023) by exploring the demographic differences between participants and nonparticipants, then goes on to extend the work by (1) first exploring the different self-reported financial behaviors of participants and nonparticipants and (2) then determining whether these differences continue once respondent demographics and attitudes are incorporated into the analysis.

The current work takes advantage of available data from both participants and all the invited panelists to explore whether a relationship between participation in a passive data collection of financial transactions and sample member demographics (such as age and education) persists after controlling for other demographics and attitudes (such trust and privacy concerns). While it might be possible to create weights to include certain demographics, such as age and education, attitudes cannot easily be accounted for in traditional weighting schemes, raising concerns for the validity of key estimates for such work if differences persist.

Literature Review

Nonparticipation Bias within Passive Data Collections

Two studies explore nonparticipation bias within a passive data collection included as a component of a larger study (e.g., these studies also included traditional self-report surveys as well). The first by Elevelt et al. (2019) invited members from the Dutch LISS panel to a smartphone-only version of the Dutch Time Use Survey (TUS). Participants were asked to

perform five sets of tasks that were embedded within a traditional self-report survey, including accepting an invitation and installing an app, answering pop-up questions, and giving permission to record GPS location and call data. The authors found nonparticipation bias in estimates of time use by linking data collected in previous surveys. Specifically, participants who agreed to all tasks worked significantly more hours and watched significantly less television than participants who agreed to only some or none of the tasks.

Overall, the authors found that busy participants (i.e., those who work a lot) were willing to participate at higher rates in all parts of the TUS, similar to findings from offline TUSs (Van Ingen et al. 2008; Abraham et al. 2006). This finding may be closely related to the participants' attitudes, as Stoop (2005) found that busy or working people are more involved in society. Elevelt et al. (2019) speculate that societal involvement is related to a more positive attitude toward smartphone-based activities.

In a cross-sectional general population randomized experiment in which sample members were invited to participate in several tasks during a web survey (including sharing their GPS location), Struminskaya et al. (2021) did not find nonparticipation bias in the estimates used to proxy participant location. However, the authors found bias in the estimates used to proxy tasks, which included taking videos and pictures of the sample member's homes/surroundings. This study asked participants to provide their location during a web survey (rather than through an app). It defined nonparticipation bias as the difference between the full sample of survey respondents and the subset of respondents who shared additional information. The authors found that privacy concerns decreased willingness to share and actual sharing of geolocation, but not of the other tasks that required taking videos or pictures.

Together, these two studies suggest that sample member attitudes relate to the decision to participate in a passive data collection. In comparison to these studies, the current work explores a passive data collection that was not part of a larger study (i.e., there was no traditional self-report survey, just the passive data collection). Without a traditional self-report survey, sample members invited to participate may have heightened privacy concerns as the parameters of a passive data collection differ from a traditional survey.

Nissenbaum states there are parameters around information sharing that are expected and used to assess comfort with the request: the actors (e.g., interviewers/research organizations), the information requested, and features of the request (e.g., how the data is shared between the study participants and the research organization). In a traditional self-report survey, Bradburn (2015) described the expected norms of the interaction between the interviewer (actor) and respondent, which includes the data request (in the form of a question) and other features of the study (e.g., whether the request was made by an interviewer or on a web-based platform). In the passive data collection environment, the actor is different (a research organization rather than an interviewer) and the data request may include the need for a third party or application on a smartphone to collect the requested data. Given these differences, sample members with existing privacy concerns or who have lower levels of trust may choose not to participate in such a passive data collection. Beyond attitudes, the passive data collection literature suggests there are other correlates to participation in a passive data collection.

Correlates to Participation

Sample member demographics – such as age and education - appear to impact a sample member's decision to participate in a passive data collection (Mulder & de Bruijne 2019; Revilla et al. 2017; Revilla et al. 2019; Angrisani et al. 2023; Jäckle et al. 2023). Angrisani et al. (2023)

explored representation issues for a financial tracking study conducted on the UAS Panel. They found that younger adults were more likely to consent to participate when compared to older adults, as were more educated panelists when compared to less educated panelists. The authors found that those who consented to participate were more likely to check their account balances online compared to nonconsenters. They speculated that nonconsenters may have security concerns that make them less likely to check their account balances and to participate in such a passive data collection of financial transactions, which would result in inaccurate estimates of online banking activities in the resulting data. However, due to the low sample size, the authors were unable to explore these differences in the actual participants as only 135 panelists linked their financial institutions in the study, compared to the 500 that consented (or agreed) to download the app (but then did not link any financial institutions).

Till now, the relationship between attitudes and participation in a passive data collection has only been explored within an invitation to a hypothetical passive data collection (Bach et al. 2024; Wenz et al. 2019; Keusch et al. 2019) or to active app-based data collections (Jäckle et al. 2019; 2023). Work by Jäckle et al. (2019; 2023) revealed that participation in mobile app-based studies correlates with other indicators of cooperativeness, such as respondents who consent to link their credit rating data to the survey being over-represented among app participants. The additional studies focused on invitations to hypothetical passive data collections found sample members who had previously shared similar data (Bach et al., 2024), had higher levels of trust, and those who were more civically engaged (e.g., regular volunteers) (Wenz et al. 2019; Keusch et al. 2019) were more likely to participate while privacy concerns had a negative impact on the sample members willingness to participate. While Wenz et al. (2019) and Keusch et al. (2019)

asked directly about sample member attitudes, unfortunately, this is not possible in the current work.

However, it may be feasible to capture respondent's attitudes toward privacy with measures of personal information protection behaviors. Such protection behaviors include whether the sample member knows how to adjust the privacy settings on social media networks and/or smartphone apps. The literature suggests that privacy protection behaviors serve as a proxy for privacy concerns, as the two are correlated (Cho 2010; Madden & Smith 2010; Litt 2013; Kamp 2016). Assuming the invited sample member possesses general Internet skills (Buchi et al. 2016), the literature suggests that those with privacy concerns invest time in learning how to protect their personal information and will engage with more technological privacy tools than those with lower levels of privacy concern. For the current work, it is assumed that the sample members possess general Internet skills as they are active members of an online probability panel, and privacy concerns are measured by the respondent's familiarity with privacy protection measures.

The current work first extends what Angrisani et al. (2023) found, testing whether differences in financial behaviors exist in a financial tracking study through comparisons of self-reported financial behaviors from participants and nonparticipants on the UAS panelists. The current work will then explore the relationship between participation and sample member characteristics to determine if a relationship between participation and sample member demographics (such as age and education) persists after controlling for other demographics and attitudes (such as trust and privacy concerns).

Hypothesis & Research Questions

RQ1: Do differences exist between the UAS Panelists who refused to participate in the passive data collection of financial transactions and those who actually participated in the passive data collection in terms of self-reported financial behaviors?

RQ2: If there are differences, do they persist after accounting for sample member attitudes and demographics?

> H1: Following Angrisani et al. (2023), it is expected that participants in the financial tracking study conducted on the UAS Panel will be younger and more educated than those UAS Panel members who choose not to participate.

Methods

Data

Data from the Understanding America Study (UAS) panel will be used for this work. The UAS is a US population-representative household Internet panel managed by the Center for Economic and Social Research (CESR) at the University of Southern California (USC) of approximately 9,000 respondents at the time of this writing. Panel members are recruited by address-based sampling, and respondents without prior Internet access are provided Tablets and broadband Internet access if needed.

The data for this work comes from the UASFin study, which has an approximately 12% participation rate. For sample sizes used in the current work, please see Table 1 below. The UASFin study is an ongoing study in partnership with the Financial Health Network (https://finhealthnetwork.org/) with the goal of understanding better the financial health and financial flow (income and expenditures) of Americans. In total, 7,747 UAS Panelists (as of

January 2022) have been invited to participate in the UASFin study through the course of four rounds of invitations, with additional invitations planned to include all panelists who report having a bank account that can be accessed online. Of those 7,747 invited, 952 (or 12.31%) agreed to participate and linked at least one bank account. Sample members are asked to participate in a study in which all the financial transactions for the accounts they link are tracked through a third-party financial aggregator (Plaid). Upon agreeing to participate, the sample member is immediately directed to the UASFin study website that the UAS created. Through this site, participants create an account, link their financial institutions (which includes sharing passwords), and view their corresponding account balances.

Following work by Olson (2006) and Peytcheva and Groves (2009), this study estimates that the nonparticipation analysis will be drawn from surveys that all sample members completed, resulting in an analytic sample of approximately n = 900. Olson (2006) found that nonresponse bias analyses based on survey reports yielded conclusions similar to those based on records. Peytcheva and Groves (2009) further justify the use of this approach by including in their meta-analytic dataset nonresponse bias studies that reported nonresponse bias based on comparisons between respondents and nonrespondents to screener interview data as well as follow-up studies of samples of nonrespondents, essentially comparing responses to survey questions across groups rather than comparing responses to records.

Research Question Variables of Interest

Nonparticipation Bias Analysis (Research Question 1)

To address research question 1 as to whether differences exist between the participants who linked at least one account and the invited sample on variables related to financial activity in the

self-report data, a review of the differences between the participants and nonparticipants in terms of self-reported financial behaviors is conducted.

To begin the analysis, the UASFin data are linked with survey responses collected in previous UAS surveys, providing background information on both participants and nonparticipants in the UASFin study. The survey responses collected in previous UAS surveys are all drawn from surveys of either the full panel or a random sample selected from the full panel. The self-reported measures of financial behavior include whether the sample member uses online banking, has ever checked an account balance over the Internet, ever transferred money from one account to another over the Internet, banked online (at a computer), used a bank "app" on a mobile phone, and/or whether the respondent pays [bills] with automatic bill payment.

Multivariate Logistic Regressions: Dependent and Independent Variables (Research Question 2)

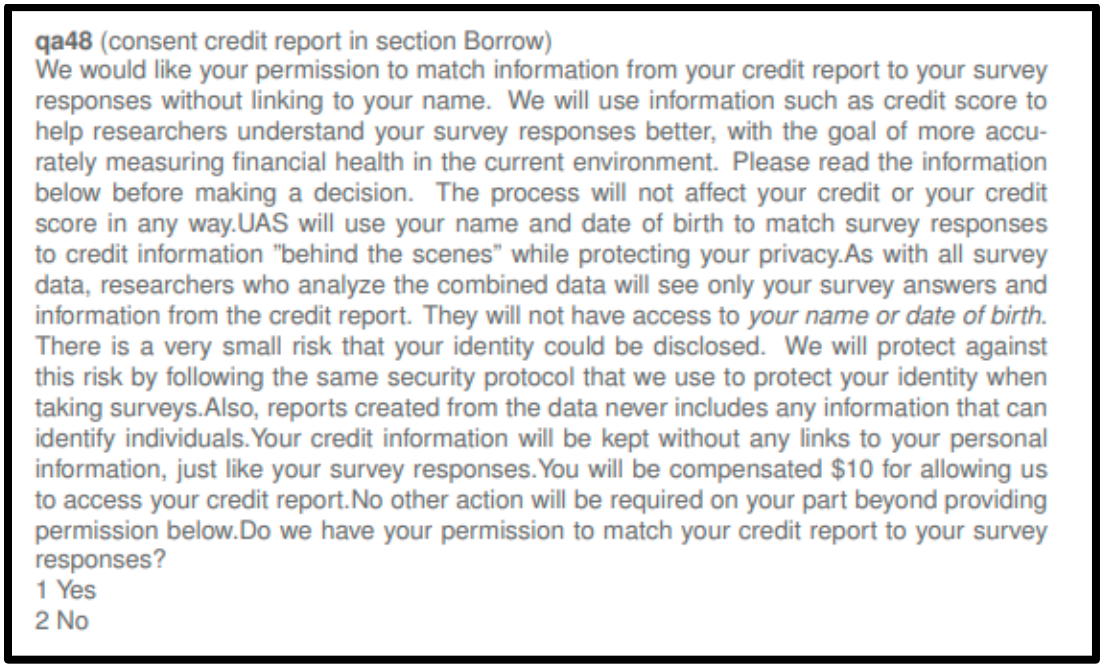
Following the comparison of participants to nonparticipants, a multivariate logistic regression is used to understand the relationship between participating in the passive data collection and sample member demographics, financial behaviors, and attitudes. The dependent variable to address research question 2 is whether the sample member participated in the passive data collection. Participation is defined as the sample member having linked at least one financial account to the financial aggregator.

The independent variables include demographics, self-reported measures of financial behavior, and measures of sample member privacy and trust, as described below in Table 1.

Table 1: Descriptions of Estimates, Sources, and Sample Size

	Measure	Description	Source	Sample Size
	Participate	Plaid Consent Surveys (From April 2019 to July 2021)	UAS 181, UAS 192, UAS 297	7,747
	Demographics: Age, Education, Gender, Income	Regular Household Surveys	My Household	7,747
Self-Reported Financial Variables	I have checked my account balance over the Internet	Yodlee Administrative Record Internet Banking Project (July 2016 to August 2018)	UAS 35	904
	I have transferred money from one account to another over the Internet		UAS 35	904
	I pay bills using online banking	Financial Services & Decision Making Wave 3 (April 2020 to June 2022)	UAS 239	5,580
	I use a computer to conduct banking activities online		UAS 239	5,581
	I use a banking "app" on a mobile phone to conduct bank activities		UAS 239	5,581
	I set up and use automatic bill payment through my bank		UAS 239	5,580
Measure of Privacy Concern	Awareness of Privacy Practices: "I know how to adjust privacy settings"	What do People Know about Social Security Wave 3 (April 2020 to June 2022)	UAS 231	7,413
Measures of Trust	Overall trust, Big Five – "I am someone who is generally trusting"	Financial Literacy; Personality; Understanding Probabilities; Numeracy; Well-Being (January 2018 to October 2022)	UAS 121, UAS 237	7,738
	Trust in Organizations - Provided consent to link credit report (Question provided in Figure 1 below)	Household Finances Annual Pulse Survey (April 2020 to May 2020)	UAS 233	6,453

Figure 1: Trust in Organization



Participants Demographics (Hypothesis 1)

To address hypothesis 1 as to whether participation increases with the respondent's education and decreases with the respondent's age, the various logistic regressions described below will include several demographics from all invited UAS panelists. These demographics include the invited panelist's age, education, gender, and income. These demographics are collected and updated quarterly for all UAS panel members.

Results

All analyses presented here are unweighted, as this work aims is to describe the UAS sample in terms of participation rather than to infer to the broader population.

Nonparticipation Analysis (Research Question 1)

Table two below suggests differences in the self-reported financial variables of interest and the decision to participate. A series of chi-square tests for independence indicate significant differences between those UAS panelists who chose not to participate and those who did participate in terms of each reported financial behavior. This suggests that, for research question one, differences exist between the full sample of UAS Panelists invited to participate in the passive data collection of financial transactions and those who actually participated in terms of self-reported financial behaviors. Further, for each self-reported financial behavior, participants report participating at higher rates than those who choose not to participate. The results from multivariate logistic regressions in the next section will explore whether the self-reported financial behaviors remain different between participants and nonparticipants when accounting for sample member demographics and attitudes.

Table 2: Chi-square Comparisons of Measures of Self-Reported Financial Behaviors of Nonparticipants and Participants to the Passive Data Collection

	n	% Participants	% Nonparticipants	Percentage Point Differences	*X*2
I pay bills using online banking.	816	80.9% (n=115)	59.8% (n=701)	21.1%	20.50***
I have checked my account balance over the Internet (at least once).	863	96.7% (n=119)	86.4% (n=744)	10.3%	13.02***
I have transferred money from one account to another over the Internet.	863	86.9% (n=119)	68.6% (n=744)	18.3%	18.35***
I use a computer to conduct banking activities online.	815	79.1% (n=115)	63.6%(n=700)	15.5%	11.40***
I use a bank "app" on a mobile phone to conduct bank activities.	815	67.0% (n=115)	44.3% (n=700)	22.7%	20.59***
I set up and use automatic bill payment through my bank.	816	70.4% (n=115)	58.1% (n=701)	12.3%	6.50**
Notes: ** p<0.05, ** p<0.01, *** p<0.001*					

For this work, the standard errors are computed using the SRS formula as this work focuses on nonparticipation to the passive data collection, not the cumulative nonparticipation for the entire panel (Olson 2006; Struminskaya et al. 2021; Yeager et al. 2011). The full survey sample is drawn from the frame representing the target population and is treated as a simple random sample (SRS) of the target population. Overall, the variances are homogenous, and the standard

errors around each estimate of the self-reported financial behaviors are similar for those who choose not to participate (ranging from 0.7 to 1.2) and the participants (ranging from 1.5 to 1.7).

In additional demographic comparisons, participants tend to be younger and more likely to have a higher education, which aligns with the earlier work by Angrisani et al. (2023) and with Hypothesis 1. This finding suggests that controlling for age and education in a traditional weighting scheme should narrow the gap between the estimates for the nonparticipants and the participants, which is explored in the next section.

Multivariate Logistic Regressions (Hypothesis 1 & Research Question 2)

Given that differences exist between the participants and those who choose not to participate in self-reported financial behaviors, the question remains whether the observed association persists after accounting for sample member attitudes and demographics. To respond to the second research question, a series of multivariate logistic regressions are utilized. Initially, the relationship between participation and each of the self-reported financial behaviors is adjusted by incorporating age and education (following Angrisani et al. 2023) as well as income and gender (in equation/model 1). Then the measures of trust and privacy enter the model (in equation/model 2). Measures of privacy and trust are pertinent to understanding the relationship between the sample member and the research organization, as discussed above (Bradburn, 2015; Nissenbaum, 2010). All the independent variables are either drawn from the full sample of UAS panelists or from surveys that drew a random sample from the full list of UAS panelists. This maintains the random sampling of the respondents included in the equations below.

$$Participation = \beta_o + \beta_1 Fin.Behave + \beta_2 Gender + \beta_3 Age + \beta_4 Edu + \beta_5 Inc \quad \text{[Eq. 1]}$$

$$\ldots + \beta_6 Privacy + \beta_7 Trust \quad \text{[Eq. 2]}$$

For these models, each financial behavior will enter the model independently of the others for a few reasons. First, this will directly test whether the estimate of this behavior may contain bias once the demographics are included in the model. Second, this avoids any issues with multicollinearity between the measures of self-reported financial variables. For these measures, two are drawn from one survey, and the remaining four are drawn from a separate survey – both of which selected a random sample from the UASPanel. While there does not appear to be a strong relationship between any of these binary measures, the two from the same survey ["checked an account balance over the Internet" and "transferred money from one account to another over the Internet"] have the highest correlation of 0.51, as shown in Table 3.

Table 3: Pearson Correlation Matrix – Self-Reported Financial Behaviors and Sample Member Attitudes

	…use online banking	… balance over the Internet	… transfer money over the Internet	… have banked online	…use a bank "app"	…auto. bill pay	Privacy
I pay bills using online banking.	1.0						
I have checked my account balance over the Internet (at least once).	0.30	1.0					
I have transferred money from one account to another over the Internet.	0.37	0.51	1.0				
I use a computer to conduct banking activities	0.42	0.31	0.38	1.0			
I use a banking "app" on a mobile phone to conduct bank activities	0.20	0.21	0.23	0.02	1.0		
I set up and use automatic bill payment through my bank	0.20	0.23	0.18	0.26	0.23	1.0	
Privacy	0.25	0.26	0.24	0.20	0.25	0.06	1.0
Trust	0.05	0.09	0.09	0.10	0.11	0.16	0.05

Table 4 below displays the results that with the inclusion of demographics, the relationship between participation and having "set up and use automatic bill payment through my bank" is no longer significant (p=0.1). At the same time, the other five self-reported financial variables continue to have a significant relationship with participation. In terms of the demographics, there does not appear to be any relationship between gender and participation or between income and

participation. However, age has a relationship with participation, with the odds of participating decreasing with age; and education has a relationship with participation, with the odds of participating increasing with education. This result suggests hypothesis 1 should not be rejected.

Table 4: Logistic Regression Model 1 –
Relationship between Self-reported Financial Behaviors, Demographics and Participation

Model	1a	1b	1c	1d	1e	1f
	Coeff (p)	Coeff (p)	Coeff (p)	Coeff (p)	Coeff (p)	Coeff (p)
Financial Variables						
I pay bills using online banking.	0.88***					
I have checked my account balance over the Internet (at least once).		1.18*				
I have transferred money from one account to another over the Internet.			0.78**			
I use a computer to conduct banking activities				0.62*		
I use a banking "app" on a mobile phone to conduct bank activities					0.71**	
I set up and use automatic bill payment through my bank						0.37
Gender (Male)	0.04	0.05	0.03	0.00	0.05	-0.00
Age (18-40)						
41-64	-0.59*	-0.46*	-0.45	-0.63**	-0.41	-0.56*
65+	-0.88**	-0.83**	-0.81**	-0.99***	-0.63*	-0.89**
Income (<$35,000)						
$35,001-$100,000	-0.04	0.04	0.00	-0.01	0.07	0.08
$100,001+	-0.09	-0.01	-0.05	-0.08	-0.00	0.05
Education (HS/GED or less)						
Associates degree, Vocational program, or Some college	0.24	0.06	0.01	0.25	0.29	0.28
Bachelor's or more	1.02**	0.92**	0.83**	1.00**	1.09***	1.08***
n=	816	863	863	815	815	816
Notes: * *p<0.05, ** p<0.01, *** p<0.001*						

Model 2 further explores how adding sample member characteristics to the model might adjust the relationship between participation and self-reported financial behaviors by including the sample member attitudes towards privacy and trust, as shown in Table 5 below. The measures of privacy and trust were derived from surveys that drew a random sample from the full UAS panel.

The measure of privacy concern is derived from a question of awareness of privacy practices. Respondents were asked to rate how true the following statement was for themselves: "I know how to adjust the privacy settings while using the Internet." For this work, the responses were recoded as a 0/1 binary response, with "0" defined as the respondent not knowing how to adjust privacy settings and "1" otherwise. With this coding, 1 indicates the respondent has a privacy concern and is expected to have a negative relationship with participation. Overall, 73% of the analytic sample of respondents (n=592) to this question indicated they knew how to adjust their privacy settings.

Trust is measured in two ways in this work. First, the measure of how trusting the respondent considers themselves is derived from a question asking for the respondent's level of agreement with the statement: "I am someone who is generally trusting." For this work, the responses were recoded as a 0/1 binary response, with "0" defined as "disagreeing" or "neither agreeing nor disagreeing," indicating the respondent is not trusting, and "1" defined as agreeing they are trusting, which is expected to have a positive relationship with participation. Overall, 78% of the analytic sample of respondents (n=675) to this question indicated they were trusting.

The second measure of trust captures the respondent's trust in the research organization requesting the data. This binary measure was derived from whether the respondent had previously consented to link a credit report in a previous study, with "0" defined as not having

previously consented and "1" defined as having consented and expected to have a positive relationship with participation. In total, 56% of the analytic sample of respondents (n=436) requested to share their credit report previously consented to the request.

Before the results of model 2 are presented, the relationship between participation and privacy concerns, and the relationship between participation and trust is explored (not shown). Using a logistic regression to model the relationship between participation and the respondent demographics and attitudes indicates that the attitudes of privacy and trust in the research organization are significantly related to the decision to participate - regardless of whether they enter the model independently or together. When each of the attitudes entered the model separately, the final odds ratio of trust in the organization measure was 6.42 (=exp(1.86); $p<0.000$), suggesting that the odds of a sample member participating in the passive data collection is nearly six times higher if the sample member trusted the research organization enough to participate in an earlier study that linked credit report information. Further, those with privacy concerns have 1.30 (=exp (0.26); $p<0.000$) higher odds of participating in the passive data collection, which was not the direction expected. The trust measure of whether the respondent considers themselves a trusting person does not have a relationship with the decision to participate.

As shown in Table 5 below, when the financial behavior variables are independently brought into the model with respondent demographics and attitudes, two of the six self-reported financial behaviors continue to have a significant relationship with participation. Specifically, the respondent reporting they "pay bills using online banking" and stating they have ever used "a banking "app" on a mobile phone to conduct bank activities" continue to have a relationship with participation, while the four other financial behaviors do not. Further, trust in the organization

and the oldest age group (65+) and most educated group (bachelor education or more) have a statistically significant relationship with participation. This result continues to suggest that hypothesis 1 should not be rejected. This Table indicates that while controlling for respondent demographics reduces the nonparticipation bias, two of these variables continue to include some level of bias, which is not expected to dissipate by including traditional weights or respondent attitudes in the model.

Table 5: Logistic Regression Model 2 – Relationship between Self-reported Financial Behaviors, Demographics, Attitudes and Participation

Model	2a	2b	2c	2d	2e	2f
	Coeff (p)	Coeff (p)	Coeff (p)	Coeff (p)	Coeff (p)	Coeff (p)
Financial Variables						
I pay bills using online banking.	0.86**					
I have checked my account balance over the Internet (at least once).		0.85				
I have transferred money from one account to another over the Internet.			0.51			
I use a computer to conduct banking activities				0.54		
I use a banking "app" on a mobile phone to conduct bank activities					0.54*	
I set up and use automatic bill payment through my bank						0.15
Privacy	0.02	0.17	0.20	0.12	0.08	0.18
Trust in Orgs	1.73***	1.73***	1.73***	1.71***	1.67***	1.72***
Gender (Male)	0.04	0.08	0.07	0.00	0.06	0.01
Age (18-40)						
41-64	-0.57*	-0.51	-0.49	-0.66*	-0.48	-0.60*
65+	-1.00**	-0.99**	-0.96**	-1.14***	-0.85*	-1.04**
Income (<$35,000)						
$35,001-$100,000	-0.15	0.04	-0.08	-0.12	-0.08	-0.05
$100,001+	-0.14	-0.05	-0.09	-0.12	-0.08	-0.03

Education (HS/GED or less)						
Associates degree, Vocational program, or Some college	0.38	0.17	0.11	0.34	0.39	0.38
Bachelor's or more	1.27***	1.12**	1.04**	1.20**	1.29***	1.29***
n=	699	728	728	698	698	699
Notes:* *p<0.05, ** p<0.01, *** p<0.001*						

Discussion

There is evidence in the literature that demographics, as well as the attitudes of sample members, impact their decision to participate in a survey. The current work explored whether sample member demographics and attitudes impact participation in a passive data collection of financial transactions and whether nonparticipation bias might exist within the resulting estimates of financial behaviors. Similar to the earlier work by Angrisani et al. (2023), these results suggest that the first hypothesis, that participation in a passive data collection is positively related to a panelist's education and negatively related to a participant's age, should not be rejected. Younger adults and those with higher education levels tend to be early adopters of mobile banking (Laukkanen 2008; Aktunan 2012). It may be younger and more educated sample members are more likely to participate in the passive data collection of financial transactions because they are more familiar with mobile banking than their counterparts, as familiarity with a technology is a predictor of participation in a data collection that utilizes the same or similar technology (Wenz 2019).

The first research question explored whether there was a difference between those who participated in the study and the sample of invited UAS panelists who chose not to participate in terms of their self-reported financial behaviors. For this, the answer is yes; there appears to be a difference between those who were invited to participate in the study and those who actually did participate in the study in terms of self-reported financial variables. For the second research

question as to whether this observed association might persist after accounting for sample member attitudes and demographics, the results suggest that for two of the six self-reported financial behaviors of interest, there continued to be a significant relationship between participation and that financial behavior after accounting for the respondent's demographics and attitudes. This suggests that differences between participants and nonparticipants continue and that nonparticipation bias in these measures of financial behaviors will persist and be a concern for researchers working with these estimates.

Overall, differences exist between nonparticipants and those who ultimately participated in the study, and nonparticipation bias is a concern in the resulting estimates of self-reported financial behaviors.

Limitations and Implications

While these results do not appear encouraging for researchers interested in using a similar approach to estimate financial behaviors, there are limitations to this work and several areas for future research to understand better what was found here. First, the self-reported financial variables likely contain some level of measurement error that cannot be identified. Future work should consider ways to measure actual financial behaviors rather than those that are self-reported to avoid this error.

Further, participation is defined as linking at least one financial account. This is a limitation in that, again, it is self-reported, and the actual number of accounts the sample members has is unknown. However, it is a large concern and limitation for those working with the underlying financial transaction data. The current work may be a starting point in helping to understand nonparticipation bias in the context of financial transaction research, but the larger question remains as to what participants choose to link once agreeing to participate in the study. Do they

link all the financial accounts asked for or just a portion of those they have? Suppose those who only link a portion of their accounts also have similar financial behaviors. In that case, the bias in the resulting estimates might be much larger than initially thought.

Third, several sample members dropped, and cases were missing across the models as the independent variables were drawn from different surveys with samples that did not completely overlap. Future research might include independent variables all pulled from the same sample. It is expected that future surveys administered on the UAS panel might provide such data with larger sample sizes.

An implication of the current work for the field is the finding that traditional weighting schemes should not be expected to remove nonparticipation bias found in estimates of financial transactions and behaviors. While it may begin to alleviate the issue, the bias remains. Further, many substantive researchers don't use weights. This may be particularly true of economists, the primary users of such financial tracking data. Traditionally weights do not account for the attitudes of the sample members.

Attitudes and the impact of earlier studies cannot easily be accounted for in traditional weighting schemes. Participation in a passive data collection of financial transactions is significantly associated with the respondent having trusted the research organization during an earlier data collection. The result that trust is positively related to participation in a passive data collection of financial transactions is aligned with the previous findings on the importance of attitudes found by Struminskaya et al. (2021) and Elevelt et al. (2019). However, as trust appears to play such a large role in the decision to participate, anything correlated with trust might also be plagued with nonparticipation bias. Capturing information about the sample member's exposure to previous

studies by the research organization, previous invitations to link their data, or invitations to participate in a passive data collection may help to alleviate this bias if included in the analysis.

Future work should also strive to include a measure of the sample member's level of civic engagement and a measure of privacy concerns, which both Wenz et al. (2019) and Keusch et al. (2019) found to be related to a sample member's willingness to participate in a passive data collection. Unfortunately, for civic engagement, such a measure could not be identified from the UAS Panel that allowed for a large enough sample size to include. A proxy was included for privacy concerns as a self-reported measure of privacy concerns was not available. Future work should also test the assumption of the current research that a third-party intermediary in the data collection process might change the sample member's attitudes toward the study.

Finally, the current work looked to identify whether nonparticipation bias may exist within the passive data collection of financial transactions but makes no effort to correct it. Future research into such adjustments is planned.

Chapter Two: Addressing Sample Member Data Security Concerns to Reduce Nonparticipation in a Passive Data Collection

Abstract

The current literature amongst Westernized countries suggests that approximately half of sample members invited to participate in a passive data collection site data security concern as the reason they do not want to participate. The current study explores whether offering a message focused on alleviating data security concerns to sample members who choose not to participate is an effective tool to decrease the number of refusals. This study further explores whether the participation rate in a passive data collection is impacted by varying the topic of the study (e.g., a time use study, a study of smartphone activities, or a study of social network size) and the requested data (e.g., location data, a capture of apps on the phone, or a captured of store phone numbers). The results suggest that the type of data requested has a larger impact on participation than the study topic and that offering a refusal conversion message may be an effective tool to convert invited sample members who initially chose not to participate in the passive data collection.

Introduction

Nissenbaum's (2010) framework of contextual integrity suggests there are instances when individuals may wish to conceal personal information they are comfortable sharing in other instances. Sample members view the instances in which they are asked to share their data through a number of parameters, specifically who is asking for the data, what data is requested and for what reason, and features of the study such as how the data will be collected. What data is requested (i.e., GPS data) and for what reason (i.e., to study travel patterns) is jointed defined as the context around the data request and allows the sample member to assess and understand the study. If the context around the data request is aligned with the sample member's expectations for such a study, they may be more comfortable participating. However, suppose the requested data does not align with the intended use or topic of the study. In that case, the flow of information may be perceived as inappropriate, resulting in a perceived violation of privacy. For example, for a study focused on travel or time use, the sample member may feel that collecting GPS data is in alignment with the study; however, for the same study, requesting a

capture of apps on their phone may not make sense and be viewed as out of alignment. If the sample member perceives the information flow is inappropriate, they may decide not to participate in the passive data collection.

It is most common to collect passive data alongside a more extensive study that includes a self-reporting component, such as a web survey. Such combination studies are only expected to increase as social science research shifts towards digital research methods (Struminskaya et al. 2020a). However, many sample members are unwilling to participate in passive data collections due to data privacy concerns, as Keusch et al. (2019) and Struminskaya et al. (2021) both found that approximately 44% of sample members were not willing to participate in a hypothetical study due to privacy and data security concerns.

This research tests the hypotheses that (1) sample members care about the context of the data collection, which includes a description of the study's topical focus and the type of data that is collected, and (2) that providing a follow-up message reiterating the privacy guarantees of the study to sample members not willing to participate will reduce refusals within a smartphone, web-based passive data collection.

Literature Review

Alignment of Study and Requested Data

Leverage-salience theory states that a sample member's decision process when deciding whether to participate in a data collection varies systematically across sample members (Groves, Singer, and Corning 2000). This variation results from a sample member's inclination to accept a fixed component of the survey (e.g., what data is collected or the topic of the study) and the salience of that component in the survey request (e.g., alignment or misalignment of fixed components).

To date, within the passive data collection literature, there is evidence that the type of data collected may impact participation (Revilla et al. 2017; Revilla et al. 2019; Wenz et al. 2019). However, Keusch et al. (2019) did not find that topic (consumer behavior, mobility, or social interaction) impacted respondents' willingness to participate in a smartphone app-based study. In this work, Keusch et al. (2019) posed vignettes to respondents, with each vignette collecting the same five types of data, making it difficult to disentangle which type of requested data may have been viewed as in or out of alignment with the topic. The current work aims to disentangle the separate impact of the requested data type and the topic. Keusch et al. (2019) requested (1) the technical characteristics of the phone, (2) the telephone network currently used, (3) the current location (every 5 minutes), (4) what apps are used and what websites are visited on the phone, and (5) number of incoming and outgoing phone calls and text messages on the phone. Overall, the authors found that 35% of the invited sample members were willing to participate in at least one of the studies, with 11% willing to participate in all of the proposed vignettes.

In a smartphone app-based study focused on the impact of long-term unemployment conducted by Kreuter et al. (2019), the authors found a 16% participation rate. Participants were able to choose which of the requested data they were willing to share, and the authors reported little variation across the actual sharing of different types of data, with the majority of the participants willing to share all the requested data. The data requested in this study included: (1) mobile phone network quality and location information collected every half hour, (2) mobile phone interactions (collected participants' incoming and outgoing call and text message logs), (3) characteristics of social networks,[2] (4) activity data and periods of activity (collected from the

[2] The information collected was gender and nationality of the phonebook entries by matching the first name of each contact with information from the website Genderize (https://api.genderize.io) and first and last names with

built-in accelerometer), and (5) information on apps installed on the participant's smartphone and the frequency of their usage. From the results, the authors suggested that sample members were comfortable sharing detailed information to help researchers understand the impact of un/employment, though the topic was not varied.

Without experimentally varying the topic and the requested data, it is unclear whether the topic may impact participation differently, given what data is requested. As Nissenbaum's (2010) framework of contextual integrity suggests, it might be that sample members have higher privacy concerns (and rates of refusal) if the requested data does not appear to align with the topic.

A Refusal Conversion Message

Sample members decide whether to participate in a study quickly, giving researchers only a few moments to share the study context, promise privacy, and provide persuasive reasons to participate (Groves et al. 2004). A follow-up message to sample members may help persuade them to participate, alleviate concerns, and continue the conversation (Dutwin et al. 2014). As noted by Dillman (2009), follow-up information may help to build trust by legitimizing the survey request, and sample members construe it as a "reward" of sorts. However, Olson et al. (2006) found that tailoring the content of follow-up information was less effective than sending a generic letter that appealed to the respondent's sense of community, so it is unclear what type of follow-up information is most effective in avoiding refusals.

While it is unclear what type of follow-up information may be the most effective in the passive data collection context, there is evidence in the literature that approximately half of the time, sample members refuse to participate due to privacy or data security concerns. Keusch et al.

information from the website NamePrism (www.name-prism.com). In neither case were data transmitted to the outside websites.

(2019) found that 44% of sample members were not willing to participate in a hypothetical study due to privacy or data security concerns, while an additional 12% were not willing because they did not have enough information or control over what happens to the data. Struminskaya et al. (2021) found the most frequently cited reason for an individual to be unwilling to participate in a hypothetical smartphone-based app study was privacy and anonymity concerns (44.3%), with an additional 10% concerned with the misuse of data and data safety. Struminskaya et al. (2021) also found that of those individuals that were unwilling to participate, 7.1% stated they might be willing if they could be guaranteed privacy.

If the sample member has privacy or data security concerns, a way to address this may be to reiterate when and how the collected data will be used and who has access to the data, as well as to provide contact information for the researcher again in case any questions arise (Nissenbaum 2010; Lavrakas 2008). To date, no work has not provided any follow-up message to sample members who initially refuse to participate in a passive data collection to determine whether refusal conversion is possible within this context. For the current work, respondents are provided with the privacy and data security practices in text in the introduction, and this information is reiterated in a video format. The goal of this design is twofold. First, to increase the visibility of the privacy and security components of the study (Sala and Lynn 2009) and, second, to allow respondents who played the video to be identified - which leads to the assumption that the respondent received the information.

Overall, this work attempts to fill several gaps in the passive data collection literature. First, this work experimentally varies the topic and the type of requested data in the invitation to the passive data collection. Second, this work offers a follow-up video message to respondents who refuse to participate in the passive data collection embedded within the web survey. Respondents

who refuse to participate will randomly receive either a generic message asking them to reconsider participating (the "control" message) or reiterating the privacy and data security promised in the study.

Research Question & Hypotheses:

RQ1: How does the alignment between the study topic and the requested data affect participation rates?

H1: The interaction between the study topic and the data collected will significantly affect participation in a passive data collection, with greater alignment between the topic and data collected leading to greater participation rates. Specifically, it is expected that:

- For the daily travel patterns study, participation will increase when the requested data is a capture of the sample member's location than when the requested data is for a capture of their detailed screen time or contact phone numbers;
- For the smartphone activities study, participation will increase when the requested data is a capture of the sample member's detailed screen time than when the requested data is for their location or contact phone numbers; and
- For the social network size study, participation will increase when the requested data is a capture of the sample member's contact phone numbers rather than when the requested data is for their location or detailed screen time.

H2: Sample members who refuse to participate due to data security concerns can be converted to participate when provided with a follow-up message that reiterates when, how, and by whom the collected data will be used.

RQ2: Does the demographic makeup (in terms of employment and student status, ethnicity, gender, and age) of the final sample differ from the demographic makeup of the sample that initially agreed to participate?

Methods

Study Design

The study includes two experiments embedded in a single web-based survey to test the hypotheses and explore the research question.

Experiment 1

The first experiment is a 9-cell (=3 study topics * 3 types of data collected) factorial design vignette that explores the interaction between the study topic and the data collected. The study topics are (1) daily travel pattern, (2) how individuals interact with their smartphones, and (3) social network size. The types of data requested in the experiments are: (1) GPS/locations, (2) apps installed on the participant's smartphone and the frequency of their usage, and (3) phone numbers stored in the participant's recent contact list. The study descriptions are detailed below.

The types of data included in this experiment (GPS/location information, information on apps installed on the participant's smartphone and the frequency of their usage, and phone numbers stored in the participant's recent contact list) are chosen for several reasons. First, all types of collections are feasible and becoming easier to implement as technology advances. Second, all types are valuable to supplement social science research. All can be used to validate (and potentially replace) self-reported measures regarding smartphone interactions and location. In addition, all the measures are useful for supplementing data collections related to such topics as social network integration, travel behavior, and modes of transportation. Specifically for measures of GPS, such data is valuable for time-use surveys (Scherpenzeel 2017; Elevelt et al. 2021).

Experiment 2

The second experiment is the random assignment of sample members who refuse to participate in the passive data collection to receive a video message that either reiterates the data security of the study or simply asks the sample member to reconsider participating (control). It is expected that approximately half of the respondents who refuse to participate in the passive data collection will refuse due to data security concerns (Keusch et al. 2019; Struminskaya et al. 2021). Because of this relatively even split, the follow-up messages were randomly assigned to respondents. After receiving the message, respondents were asked one more time whether they would like to participate. The full survey is provided in Appendix B.

For this work, the described passive data collection was embedded in a web survey, reflecting Struminskaya et al. (2021); however, this is a deception study as no data was actually collected. Allowing the sample member to think this is a data collection that will actually occur has the advantage of avoiding any potential gaps in terms of sample members who might say they want to participate in a hypothetical study but then later choose not to participate in the actual study. For example, Angrisani et al. (2023) found that 60% of invited Understanding America Study (UAS) panel members were willing to participate in a hypothetical study tracking their financial transactions. However, only 12.2% fully participated in the actual study.

Data

This experiment was conducted through Prolific's nonprobability panel, and the survey was coded in Qualtrics. As participation in the described studies required the sample members to respond on their mobile devices, the sample selected through Prolific was restricted to participants on their mobile devices. This restriction allowed for a scenario that appeared real to

the respondent. In total, 6,989 sample members participated in the survey (resulting in 2,964 refusals), as shown in Table 1, with a median time of 2 minutes 32 seconds.

Table 1: Experimental Assignment by Study Topic and Requested Data

Data Type		Topic		
		Travel	Use of Smartphones	Social Network Size
	GPS/Location Data	780	772	778
	Recent Contacts	773	778	780
	Capture of Smartphone Activities	782	771	775

Dependent variable:

For hypothesis 1, the dependent variable is whether the sample member chose to participate in the passive data collection, with invitations shown in Figures 1-3 below.

For hypothesis 2 and the research question, the focus is the conversion of respondents who initially refused to participate in the embedded passive data collection.

Independent variables:

This section includes the questions from the study that are relevant to the experimental design.

Experimental Design – Refusal Conversion:

Figures 1 – 3: Experiment Invitations

GPS Study

Will you provide your location so we can better understand {*daily travel patterns / how people interact with their smartphones / social networks*}?

Your location will be recorded as an image, like the one displayed here.

(a) Yes, I am willing to share my location (b) No, I am not willing to share my location

If (b): Experiment
[CONTROL or DATA SECURITY message].

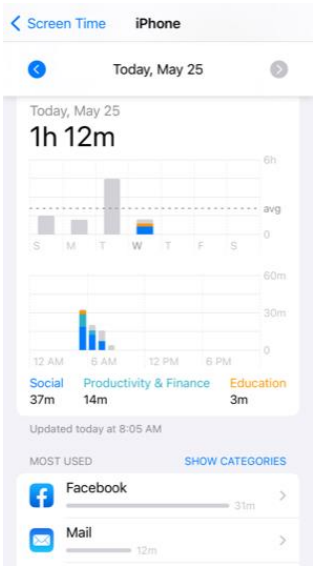
Knowing this, would you now be interested in sharing your location?
(a) Yes (b) No

Smartphone Activities

Will you provide screen time details from today so we can better understand {*daily travel patterns / how people interact with their smartphones / social networks*}?

Your information will be recorded as an image, like the one displayed here.

(a) Yes, I am willing to share information about the apps on my smartphone (b) No, I am not willing to share information about the apps on my smartphone



If (b): Experiment
[Display CONTROL or DATA SECURITY message].

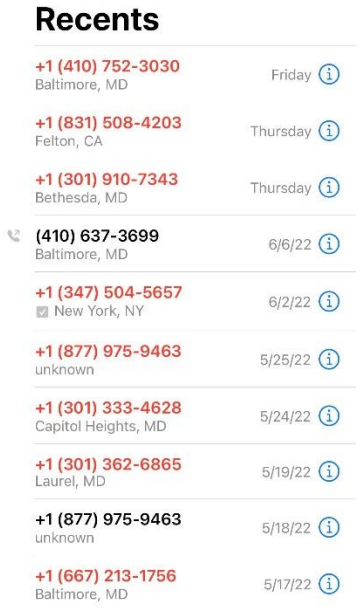
Would you now be interested in sharing your detailed screen time information?
(a) Yes (b) No

Phone Numbers in Recent Contacts

Will you provide a list of the phone numbers saved in your recent contacts from the last 20 days so we can better understand {*daily travel patterns / how people interact with their smartphones / social networks*}?

Your information will be recorded as an image, like the one displayed here.

(a) Yes, I am willing to share my contacts and interactions (b) No, I am not willing to share my contacts and interactions



If (b): Experiment
[Display CONTROL or DATA SECURITY message].

Would you now be interested in sharing your location? (a) Yes (b) No

If the sample member indicated they were interested in participating in the data collection, they were immediately informed of the deception on a debriefing page. Respondents then had the

opportunity to withdraw their responses from the study. If the sample member indicated they were not interested in participating, they were randomly assigned to receive one of the video-recorded follow-up messages. After the video and follow on question asking whether they would now like to participate, they were immediately informed of the deception on a debriefing page and given the opportunity to withdraw from the study, regardless of their response to participating. In total, 111 respondents did ask to be withdrawn from the study. As shown in Table 2 below, the number of respondents who requested to be withdrawn from the study was approximately equal across all experiments.

Table 2: Requests to be withdrawn by Study Topic and Requested Data

		Topic					
Data Type		Travel		Use of Smartphones		Social Network Size	
		Refuse – Test	Refuse – Control	Refuse – Test	Refuse – Control	Refuse – Test	Refuse – Control
	GPS/Location Data	3	8	7	4	6	4
	Recent Contacts	10	6	5	6	6	5
	Capture of Smartphone Activities	3	8	8	9	7	6

Follow-up messages

DATA SECURITY Script - Recording

Hello! My name is Alex Breslin, and I am the researcher organizing this study at the University of Maryland, College Park. Thank you for your participation so far. I understand you don't want to share {*your location / your detailed screen time information / the phone numbers of your contacts*}. I know it can be unnerving to share such information with a stranger. I want to assure you that your data will be kept anonymous. I will be the only one that will ever see it, and I promise to never present my findings in any way that would allow for you to be identified. I would greatly appreciate if you would reconsider participating. My email address was shared at the start of the survey and will be shared again at the end. Please reach out to me if you have any feedback or concerns.

CONTROL - Recording:

Hello! My name is Alex Breslin, and I am the researcher organizing this study at the University of Maryland, College Park. Thank you for your participation so far. I understand you don't want to share {*your location / your detailed screen time information / the phone numbers of your contacts*}. I would greatly appreciate if you would reconsider participating.

As discussed earlier, video recordings were used here to reiterate the study's privacy and data security components. This information was initially shared with the sample members during the invitation to the study, which appeared as text on the screen, but the follow-up message appeared as an embedded video in the web survey.

Respondent Information

As respondents were randomly assigned to one of nine invitations, demographic differences across the nine experimental groups were first evaluated. Logistic regressions exploring the demographic composition of each experimental group show no significant relationships between the experimental group and any demographic except for ethnicity within the invitations requesting location-based data (ranging from p=0.019-0.022). The confidence intervals for the specific group demographics compared to the survey sample suggest that the nine groups are fairly comparable, as shown in Tables 3(a)-(c) below. The final logistic regressions control for ethnicity to account for the slight variation in ethnicity across the groups.

Table 3a: Demographics – Social Network Size Topic Invitations

		Total	GPS/Location Data Requested	Capture of Smartphone Activities Data Requested	Phone Number and Interaction Data Requested
Employment	Full time	54.2 [52.8, 55.6]	53.6 [47.0, 60.1]	53.1 [46.6, 59.5]	51.5 [44.5, 58.5]
	Part-Time	17.2 [16.2, 18.3]	16.8 [12.4, 22.4]	14.9 [10.9, 20.1]	18.4 [13.5, 24.4]
	Not in paid work	10.9 [10.1, 11.8]	11.8 [8.2, 16.8]	11.8 [8.2, 16.7]	12.8 [8.8, 18.2]
	Unemployed	11.4 [10.5, 12.3]	13.2 [9.3, 18.3]	14.9 [10.9, 20.1]	13.3 [9.2, 18.8]
	Other	6.3 [5.7, 7.0]	4.5 [2.5, 8.2]	5.3 [3.0, 9.0]	4.1 [2.1, 7.9]
Student	Yes	20.4 [19.3, 21.5]	15.4 [11.3, 20.7]	26.3 [21.2, 32.2]	22.9 [17.6, 29.2]
Ethnicity	Asian	8.3 [7.7, 9.0]	6.0 [3.8, 9.4]	5.6 [3.5, 8.9]	8.6 [5.7, 12.5]
	Black	8.9 [8.2, 9.6]	7.4 [4.9, 11.0]	7.3 [4.9, 10.9]	7.1 [4.5, 10.8]
	Mixed	6.8 [6.3, 7.4]	7.0 [4.6, 10.6]	6.6 [4.3, 10.1]	4.5 [2.6, 7.7]
	Other	3.5 [3.1, 4.0]	4.7 [2.8, 7.8]	3.3 [1.8, 6.1]	4.1 [2.3, 7.2]
	White	72.4 [71.3, 73.4]	74.8 [69.6, 79.4]	77.1 [72.0, 81.5]	75.8 [70.4, 80.6]
Gender	Female	49.8 [48.7, 51.0]	52.5 [46.8, 58.1]	53.5 [47.8, 59.0]	52.2 [46.3, 58.1]
Age	18-25	22.2 [21.2, 23.2]	22.7 [18.3, 27.8]	24.4 [19.9, 29.6]	23.7 [19.0, 29.1]
	26-35	36.1 [35.0, 37.2]	35.8 [30.6, 41.4]	35.0 [29.8, 40.5]	31.9 [26.6, 37.6]
	36-45	22.3 [21.3, 23.3]	20.1 [15.9, 25.0]	22.8 [18.4, 27.8]	24.4 [19.7, 29.9]
	46-55	10.5 [9.8, 11.3]	12.7 [9.4, 17.0]	10.6 [7.6, 14.6]	10.0 [6.9, 14.2]
	56-65	6.1 [5.5, 6.6]	6.0 [3.8, 9.4]	4.6 [2.8, 7.7]	6.3 [2.9, 9.9]
	66-75	2.1 [1.8, 2.4]	2.0 [0.9, 4.4]	2.3 [1.1, 4.8]	3.3 [1.7, 6.3]
	76+	0.8 [0.6, 1.0]	0.7 [0.2, 2.6]	0.3 [0.0, 2.3]	0.4 [0.1, 2.6]

Table 3b: Demographics – Smart Phone Topic Invitations

		Total	GPS/Location Data Requested	Capture of Smartphone Activities Data Requested	Phone Number and Interaction Data Requested
Employment	Full time	54.2 [52.8, 55.6]	54.7 [49.9, 59.4]	55.2 [50.4, 59.9]	53.4 [48.7, 58.1]
	Part-Time	17.2 [16.2, 18.3]	17.3 [14.0, 21.3]	18.2 [14.7, 22.2]	18.2 [14.8, 22.2]
	Not in paid work	10.9 [10.1, 11.8]	10.8 [8.2, 14.2]	11.4 [8.7, 14.8]	10.4 [7.8, 13.7]
	Unemployed	11.4 [10.5, 12.3]	11.3 [8.6, 14.8]	8.7 [6.4, 11.8]	13.7 [10.7, 17.3]
	Other	6.3 [5.7, 7.0]	5.8 [3.9, 8.5]	6.5 [4.5, 9.4]	4.3 [2.7, 6.7]
Student	Yes	20.4 [19.3, 21.5]	20.3 [16.8, 24.4]	23.8 [20.0, 28.0]	21.4 [17.9, 25.4]
Ethnicity	Asian	8.3 [7.7, 9.0]	7.7 [5.7, 10.2]	7.9 [5.9, 10.4]	7.5 [5.6, 9.9]
	Black	8.9 [8.2, 9.6]	8.4 [6.4, 11.0]	9.8 [7.6, 12.5]	11.2 [8.9, 14.0]
	Mixed	6.8 [6.3, 7.4]	5.9 [4.2, 8.2]	7.2 [5.3, 9.6]	8.3 [6.3, 10.8]
	Other	3.5 [3.1, 4.0]	4.1 [2.7, 6.1]	3.8 [2.5, 5.8]	3.1 [1.9, 4.8]
	White	72.4 [71.3, 73.4]	74.0 [70.2, 77.4]	71.3 [67.5, 74.9]	70.0 [66.2, 73.6]
Gender	Female	49.8 [48.7, 51.0]	53.5 [49.4, 57.6]	53.7 [49.6, 57.7]	50.8 [46.8, 54.9]
Age	18-25	22.2 [21.2, 23.2]	18.9 [15.9, 22.3]	23.9 [20.6, 27.5]	24.9 [21.6, 28.6]
	26-35	36.1 [35.0, 37.2]	37.8 [33.9, 41.9]	36.1 [32.3, 40.1]	32.9 [29.2, 36.8]
	36-45	22.3 [21.3, 23.3]	20.0 [16.9, 23.5]	23.0 [19.7, 26.6]	21.5 [18.4, 25.0]
	46-55	10.5 [9.8, 11.3]	13.1 [10.5, 16.1]	7.5 [5.6, 10.0]	10.8 [8.6, 13.6]
	56-65	6.1 [5.5, 6.6]	8.0 [6.0, 10.5]	6.1 [4.4, 8.4]	7.1 [5.3, 9.5]
	66-75	2.1 [1.8, 2.4]	1.4 [0.7, 2.8]	2.6 [1.6, 4.3]	2.2 [1.3, 3.8]
	76+	0.8 [0.6, 1.0]	0.9 [0.4, 2.1]	0.9 [0.4, 2.1]	0.5 [0.2, 1.6]

Table 3c: Demographics – Travel Topic Invitations

		Total	GPS/Location Data Requested	Capture of Smartphone Activities Data Requested	Phone Number and Interaction Data Requested
Employment	Full time	54.2 [52.8, 55.6]	53.9 [48.5, 59.2]	56.7 [51.5, 61.8]	56.8 [51.8, 61.8]
	Part-Time	17.2 [16.2, 18.3]	17.7 [13.9, 22.1]	16.0 [12.5, 20.2]	16.4 [12.9, 20.5]
	Not in paid work	10.9 [10.1, 11.8]	10.8 [7.9, 14.6]	12.3 [9.2, 16.1]	11.8 [8.9, 15.5]
	Unemployed	11.4 [10.5, 12.3]	10.8 [7.9, 14.6]	7.7 [5.3, 11.0]	10.5 [7.7, 14.0]
	Other	6.3 [5.7, 7.0]	6.9 [4.6, 10.1]	7.4 [5.1, 10.7]	4.6 [2.9, 7.2]
Student	Yes	20.4 [19.3, 21.5]	18.5 [14.8, 22.9]	18.8 [15.2, 23.0]	18.9 [15.3, 23.1]
Ethnicity	Asian	8.3 [7.7, 9.0]	9.6 [7.2, 12.7]	9.3 [7.0, 12.2]	7.4 [5.4, 10.0]
	Black	8.9 [8.2, 9.6]	7.6 [5.5, 10.4]	9.7 [7.3, 12.7]	9.4 [7.1, 12.3]
	Mixed	6.8 [6.3, 7.4]	6.7 [4.7, 9.4]	6.3 [4.5, 8.9]	5.9 [4.2, 8.4]
	Other	3.5 [3.1, 4.0]	2.9 [1.7, 4.9]	4.2 [2.7, 6.4]	3.5 [2.2, 5.5]
	White	72.4 [71.3, 73.4]	73.2 [68.9, 77.1]	70.5 [66.3, 74.5]	73.8 [69.7, 77.5]
Gender	Female	49.8 [48.7, 51.0]	47.0 [42.5, 51.6]	50.2 [45.7, 54.7]	50.6 [46.2, 55.0]
Age	18-25	22.2 [21.2, 23.2]	20.3 [16.9, 24.3]	19.5 [16.1, 23.3]	21.3 [17.9, 25.2]
	26-35	36.1 [35.0, 37.2]	36.6 [32.3, 41.2]	37.4 [33.2, 41.9]	36.4 [32.2, 40.7]
	36-45	22.3 [21.3, 23.3]	20.3 [16.9, 24.3]	20.5 [17.1, 24.4]	20.3 [17.0, 24.1]
	46-55	10.5 [9.8, 11.3]	11.7 [9.0, 15.0]	11.3 [8.8, 14.5]	13.0 [10.3, 16.3]
	56-65	6.1 [5.5, 6.6]	7.5 [5.4, 10.3]	6.9 [4.9, 9.6]	6.3 [4.5, 8.8]
	66-75	2.1 [1.8, 2.4]	2.9 [1.7, 4.9]	4.0 [2.5, 6.1]	2.2 [1.2, 4.0]
	76+	0.8 [0.6, 1.0]	0.7 [0.2, 2.0]	0.4 [0.1, 1.7]	0.4 [0.1, 1.6]

To gauge the sample's comfort with smartphone use, sample members were asked about ten activities they might do on a smartphone other than phone calls or text messages. The top two activities were "browsing websites" (99%) and "checking the weather near you" (98.8%), as presented in Table 4 below. Nearly 40% of respondents indicated they participate in 9 or 10 of the activities in question. Together, this suggests that the sample members are comfortable using a smartphone. Comparisons across the types of data requested indicate the random assignment worked, with the largest differences across groups being 2.8 percentage points.

Table 4: Percent of Respondents Who Report Using Smartphones for Activities other than Phone Calls and Text Messaging, by Invitation Topic

		Total	Travel	Use of Smartphones	Social Network Size
Browsing websites		99.0	99.2	98.7	99.0
Checking the weather near you		98.8	98.9	98.9	98.8
Following directions from location-aware apps (for example, Google Maps)		98.0	97.7	98.0	98.2
Interacting with content on social media websites/apps		96.0	95.5	96.5	96.1
Finding restaurants, events, and/or shops near you (for example, Foursquare)		93.3	93.3	92.7	93.9
Connecting to other electronic devices via Bluetooth		92.9	92.3	93.6	92.8
Finding transportation near you (for example, Uber)		67.1	66.3	68.9	66.2
Receiving targeted offers or coupons based on your location		54.9	55.4	56.1	53.3
Tracking the number of steps you take or your workouts		57.3	57.3	57.3	57.4
Playing games that require your location (for example, Pokémon Go)		38.5	39.5	39.3	36.8
Smartphone Use	<=5 interactions	5.5	6.0	5.2	5.5
	6	9.5	9.9	9.3	9.3
	7	19.1	18.6	18.0	20.8
	8	27.2	26.5	27.8	27.3
	9	22.9	22.1	23.8	22.9
	10	15.6	16.7	15.8	14.1

Analysis

To explore the first research question and test the two hypotheses, two experiments are included in the study.

Experiment 1: Variation of Topic and Data Collected - Impact on Participation

To understand the impact of the two experimentally varied factors (topic and type of data requested) on participation, descriptive statistics will be presented, followed by a logistic regression. Equation one below explores the impact of the main effects, and equation two includes the interactions. The dependent variable for this logistic regression measures whether

the sample member agreed to participate in the passive data collection, modeled as 1 for agreeing to participate and 0 for not agreeing to participate.

$$Participation = \beta_o + \beta_1 Topic + \beta_2 Data \quad \text{[Equation 1]}$$

$$\ldots + \beta_3 (Topic * Data) \quad \text{[Equation 2]}$$

Experiment 2: Refusal Conversion

The regressions are followed by a comparison of the participation rates between those who received a data security message and those who received a generic message to reconsider participating (control). This comparison will be further disaggregated by the topic and data request that the sample member received. Finally, to explore the research question, the initial and ending participant sample demographics will be compared to better understand how follow-up messages may impact non-response bias.

The videos that re-emphasized the privacy and data security measures ranged from 34 seconds to 37 seconds. The control videos ranged in duration from 12 seconds to 13 seconds. As a check to ensure sample members who received the video message watched the video message, the time spent on the page and the number of clicks the sample member made while on the page were recorded during the web survey. On average, the respondents stayed on the page with the video for longer than the duration of the video, as shown in Table 5 below. Further, the analysis only included individuals who clicked more than two times while on the page, suggesting the respondents clicked on the video. It is not possible to capture whether the video played.

Table 5: Video Timing

	Expected Video Time (seconds)	Mean Video Time (seconds)	CI	
Contacts Control (n=491)	13	30	28	31
Contacts Test (n=490)	36	48	46	51
Screen Time Control (n=482)	13	30	29	32

Screen Time Test (n=484)	37	50	48	52
Location Control (n=506)	12	28	26	30
Location Test (n=511)	34	49	47	51

Results

Experiment 1: Variation of Topic and Data Collected - Impact on Participation

The participation rates by experiment are shown in Table 6 below. These rates indicate that refusals are not the lowest when the requested data aligns with the topic. As such, there is limited support for the first hypothesis, with only the topic of social network size following expectations.

The cells in Table 6 that were hypothesized to be in alignment are presented in **bold** and *italics*. These cells were expected to contain the largest percentage of respondents who agreed to participate in the respective column. As the response rates for each type of requested data are similar (within four percentage points), this suggests that data type had a large effect on respondent's decision to participate, whereas the study topic had essentially no effect. This finding is reiterated in Table 7 and is in line with previous research that found the type of data requested impacts the willingness of sample members to participate, while the topic of the study did not have a similar impact.

Table 6: Response Rate by Invitation

		Topic		
Data Type		Travel	Social Network Size	Use of Smartphones
	GPS/Location Data	***58%***	63%	62%
	Recent Contacts	38%	***35%***	39%
	Capture of Smartphone Activities	73%	76%	***74%***

Using logistic regressions, the analysis explores the main effects and then the interaction effects (of the data requested and the topic of the study) on participation in the study for the sample of respondents that initially chose to participate. As shown in Table 7 below – exploring the impact of data and topic on (initial) participation - the main effects in equation one suggests that the

study topics do not have a large impact on the decision to participate in the study. However, the type of data requested does have an impact – with both a request for a capture of location data and of smartphone activities having a positive impact on participation compared to a request of recent contacts. This finding is aligned with the descriptive statistics shown in Table 5 above, as a request for recent contacts had the highest percentage of refusals.

When interactions are included in equation 2, the type of data collected continues to have the most significant impact on participation. However, the interaction for a travel study collecting GPS/location data also gained significance (p=0.013), while the main effect for the travel study itself is not significant (p=0.154). This suggests that the first hypothesis – that participation will be positively impacted when the study topic and requested data are aligned – should be rejected.

Table 7: Regressions – Initial Participation

	Main Effects (Equation 1)	Interactions (Equation 2)
Studies (ref: Social Network Size)	Coefficient	Coefficient
Travel	-0.062	0.150
Use of Smartphones	0.021	0.192
Types of Data (ref: Recent Contacts)		
GPS/Location Data	0.965***	1.170***
Capture of Smartphone Activities	1.572***	1.763***
Interactions		
Travel*GPS/Location Data		-0.367*
Travel*Capture of Smartphone Activities		-0.279
Use of Smartphones*GPS/Location Data		-0.243
Use of Smartphones*Capture of Smartphone Activities		-0.290
Notes: ** p<0.05, ** p<0.01, *** p<0.001*		

Tables 6 and 7 together suggest the first hypothesis should be rejected. Overall, the results suggest that the type of data collected may have a bigger impact on the respondent's decision to

participate than the topic of the study and that alignment of the topic and type of data collection does not improve participation. This is aligned with the earlier work of Revilla et al. (2017, 2019), Wenz et al. (2017), and Keusch et al. (2019).

Experiment 2: Refusal Conversion

Digging into the possibility of converting respondents to participate in a passive data collection, Table 8 below displays the initial and final participation rates. Across all experiments, chi-square tests reveal that when the video follow-up message was offered, there was a statistically significant shift in participation with an eight-percentage point increase. Further, for each data type requested, the difference in participation before and after the participants received the video follow-up is statistically significant.

Table 8: Initial and Final Participation Rates, Overall and by Requested Data Type

Data Type	Initial Participation	Final Participation	Percentage Change
Overall	57.6%	65.9%	8.3%***
GPS/Location	61.1%	65.9%	4.8%*
Smartphone Activities	74.2%	77.9%	3.7%*
Recent Contacts	37.4%	41.9%	4.5%*
Notes: * *p<0.05, ** p<0.01, *** p<0.001*			

Table 9 below further disaggregates the findings by whether the respondents received the recording reiterating the data security measures in place or the generic recording. Chi-square tests indicate no significant differences between the initial and final participation rates when disaggregated by the data type requested and the study topic. This finding, along with the suggestion that the first hypothesis be rejected, suggests alignment between the study topic and the requested data does not affect participation rates (first research question).

However, significant differences emerge when comparing the final participation rate of those who received the control or test videos. Specifically, there are significant differences for the

study that (1) requested recent contacts (with the topic of a social network study) and (2) the study that requested GPS data (with the topic of a travel study), as shown in Table 9 below.

Overall, comparing the final participation rates by message suggests that hypothesis two should not be rejected. Sample members who refuse to participate due to data security concerns can be converted to participate with a follow-up message reiterating when, how, and by whom the collected data will be used.

Table 9: Initial and Final Participation Rates, by Assigned Recording

		Initial Participation	Final Participation	Percentage change	Difference by Messages
Treatment Groups – Data Security Message					
Data Type	**Topic**				
GPS/Location	Travel	58.0%	Control: 62.5%	4.5%	5.3%*
			Test: 67.8%	9.8%	
	Smartphone Interaction	62.0%	Control: 65.0%	3.0%	3.0%
			Test: 68.0%	6.0%	
	Social Network Size	63.2%	Control: 66.2%	3.0%	2.8%
			Test: 69.0%	5.8%	
Smartphone Activities	Travel	73.2%	Control: 75.4%	2.2%	2.7%
			Test: 78.1%	4.9%	
	Smartphone Interaction	73.8%	Control: 77.3%	3.5%	3.5%
			Test: 80.8%	7.0%	
	Social Network Size	75.6%	Control: 77.6%	2.0%	2.5%
			Test: 80.1%	4.5%	
Recent Contacts	Travel	38.2%	Control: 40.7%	2.5%	4.2%
			Test: 44.9%	6.7%	
	Smartphone Interaction	39.3%	Control: 41.6%	2.3%	4.2%
			Test: 45.8%	6.5%	
	Social Network Size	34.8%	Control: 38.1%	3.3%	6.8%**
			Test: 44.9%	10.1%	
Notes: * *p<0.05,* ** *p<0.01,* *** *p<0.001*					

Finally, to explore the second research question as to whether the converted participants might differ demographically from the initial sample of participants, demographic comparisons are made between the initial sample of participants and the final sample after the refusal conversion

video message. The demographics include employment status, student status, ethnicity, gender, and age, as shown in Table 10 below. The comparison of confidence intervals across the samples indicates no significant shifts in participant demographics between the sample that initially agreed to participate and the end sample. This finding suggests that the converted sample members have characteristics like those who initially agreed, so the additional participants would not decrease potential nonparticipation bias. The patterns remain the same when the samples are disaggregated by the requested data types (contact list, smartphone interactions, and location).

Table 10: Participant Demographics, Initial and Final Samples

		Overall Initial Participation	**Overall Final Participation**
Employment	Full time	54.6 [52.8, 56.4]	53.9 [52.2, 55.6]
	Part-Time	17.2 [15.9, 18.6]	17.3 [16.1, 18.7]
	Not in paid work	11.4 [10.3, 12.6]	11.4 [10.4, 12.5]
	Unemployed	11.2 [10.2, 12.4]	11.4 [10.3, 12.5]
	Other	5.6 [4.8, 6.5]	5.9 [5.2, 6.8]
Student	Yes	20.7 [19.3, 22.1]	21.2 [19.8, 22.5]
Ethnicity	Asian	7.8 [7.0, 8.7]	7.7 [7.0, 8.6]
	Black	8.9 [8.1, 9.9]	8.9 [8.1, 9.8]
	Mixed	6.6 [5.9, 7.4]	6.6 [5.9, 7.4]
	Other	3.7 [3.2, 4.3]	3.7 [3.2, 4.3]
	White	72.9 [71.5, 74.3]	73.0 [71.7, 74.3]
Gender	Female	51.5 [50.0, 53.0]	51.7 [50.2, 53.1]
Age	18-25	22.0 [20.8, 23.3]	22.9 [21.7, 24.1]
	26-35	35.7 [34.3, 37.2]	35.6 [34.2, 37.0]
	36-45	21.3 [20.1, 22.6]	21.1 [19.9, 22.3]
	46-55	11.2 [10.2, 12.2]	10.8 [10.0, 11.8]
	56-65	6.7 [6.0, 7.5]	6.5 [5.8, 7.3]
	66-75	2.5 [2.1, 3.0]	2.4 [2.0, 2.9]
	76+	0.6 [0.4, 0.9]	0.6 [0.4, 0.9]

Discussion

Overall, this work explored two research questions as to (1) whether alignment between the study topic and the requested data affect participation rates and (2) whether a refusal conversion

strategy might result in a final sample that differs demographically from the initial sample of participants. Specifically related to the research question on alignment of the study topic and requested data, this research tested the hypotheses that (1) sample members care about the context of the data collection, which includes the combination of the stated purpose of the study and the type of data that is collected, (2) that providing a follow-up message reiterating the privacy guarantees of the study to sample members not willing to participate is an effective refusal conversion strategy within a smartphone web survey based passive data collection.

While the passive data collection literature suggests there is evidence that the task or type of data collected may impact participation (Revilla et al. 2017; Revilla et al. 2019; Wenz et al. 2017), there has been limited exploration into whether the topic of the study might also impact the participation rate (Keusch et al. 2019). The current work is the first attempt to experimentally vary the topic of the study, and the type of data collected in a study invitation that did not appear to be hypothetical. Without such variation, it has been unclear whether the topic impacts participation differently, given what data is requested. However, the current work indicates little support for this hypothesis. Rather, the requested data type impacts participation more than the larger context of the data collection, which includes the combination of the stated purpose of the study and the type of data that is collected.

In terms of the second hypothesis, the analysis suggests that a follow-up message may help recruit sample members who initially refuse to participate in a passive data collection, and so hypothesis two should not be rejected. The literature indicates that approximately half of the time sample members refuse to participate it is due to privacy or data security concerns (Keusch et al., 2019; Struminskaya et al., 2021), and the current work suggests that providing a message that reiterates the privacy and data security promised in the study is an effective way to decrease

refusals. It is possible that reiterating this information in another format increased the saliency of this information; however, a better understanding of the impact of the message format is left for future research.

Finally, the second research question asked whether the converted participants differed from the initial sample of participants who agreed to participate in the study in such a way as to minimize nonparticipation bias that may be present. However, there is no evidence that the converted participants differed in demographics from the initial sample. This finding suggests that while there is a shift in the number of participants, any nonparticipation bias that may be present would likely not be alleviated by increasing the number of participants in the study through a video message reiterating the safety and security of the study.

Overall, the current work is a first attempt to understand the interaction between the requested data type and the study topic and how refusal conversion techniques perform within the passive data collection environment. There are several limitations of the current work that open the door for future research.

Limitations and Next Steps

There are several limitations with the current work related to the sample drawn and the simplistic experiments embedded with the web-based survey. First, relying on Prolific's non-probability panel resulted in a sample of respondents who appear more willing to participate in passive data collection than the general population of smartphone owners in the United States. This may be because panelists on Prolific's panel opted to participate in web-based surveys, which means they are more willing to participate in studies in general and may be more comfortable with web-based (and smartphone-based) studies. Overall, this resulted in a much lower rate of refusals than anticipated, which many have masked the impact of experimentally varying the topic and the

data requested in the study. Following work by Keusch et al. (2019) and Struminskaya et al. (2021), it was assumed that approximately 35% of respondents would be willing to participate in such a passive data collection as both studies found very similar rates of participation for a similar study. However, in the actual survey, 58% of the respondents were willing to participate in the studies when initially invited. This resulted in 2,964 refusals instead of the expected 4,230 refusals, which was desired given the power analysis conducted prior to the experiment. Future work should again test hypothesis one using a more general sample on a probability-based platform to better understand the interaction between the study topic and requested data.

Second, the current work invited the sample member to a passive data collection that was extremely short in duration – described as a single data collection within the web survey. While this reflected the work of Struminskaya et al. (2021), gaining a better understanding of the impact of time on the sample member's decision to participate (initially and after a follow-up message is provided) will be useful to determine when it might be appropriate to budget for a follow-up message.

Third, while the current work utilized a video for the follow-up message, future work should test the impact of the format of the follow-up message. Namely, would a message written in the text have a similar impact on refusal conversion as a video message?

Next, future work should further explore whether such a refusal conversion tool is useful in reducing nonparticipation bias within a probability-based panel. The finding in the current work that those converted to participate did not differ demographically from those who initially agreed to participate might be a characteristic of the non-probability-based panel and high levels of willingness to participate rather than an artifact of the tool itself.

Finally, future research efforts should be dedicated to better understanding the difference between data privacy concerns and data security concerns, as viewed by an invited sample member. This work took a broad definition in the invitation and follow-up message, essentially combining the two. However, Keusch et al. (2019) found that 44% of sample members were not willing to participate in a hypothetical study due to privacy or data security concerns, while an additional 12% were not willing because they did not have enough information or control over what happens to the data. Struminskaya et al. (2021) found the most frequently cited reason for an individual to be unwilling to participate in a hypothetical smartphone-based app study was privacy and anonymity concerns (44.3%), with an additional 10% concerned with the misuse of data and data safety. Struminskaya et al. also found that of those individuals that were unwilling to participate, 7.1% stated they might be willing if they could be guaranteed privacy. These responses suggest that different follow-up messages may be more effective depending on the actual concerns of the sample members. In terms of the concern related to a lack of control, chapter three will explore this in further detail.

These suggestions for future work will help translate the current work's results into actionable results that may also be applicable to research related to paradata collection, record linkage consent, and follow-up questions to avoid item missing data.

Chapter Three: An Offer of Control over Data Shared with Researchers to Reduce Nonparticipation in a Passive Data Collection

Abstract

In self-report surveys, respondents may choose to censor responses to particular questions to preserve their privacy. However, such censoring is not possible within the passive data collection context, and sample members may choose not to participate in the study rather than share their data. This work explores whether offering sample members the opportunity to control what data is shared with researchers during a passive data collection impacts participation. Sample members recruited from an online probability-based panel were randomly assigned to receive an invitation to download a smartphone app to track their location at regular intervals. The invitation varied in terms of the control sample members were given over the collected data. Overall, this work finds no evidence that an offer of control improves participation.

Introduction

Concerns over the privacy of data collected during a passive data collection are a large barrier to participation (Keusch et al. 2019; Revilla 2019; Chin et al. 2012). Compared to a traditional self-report survey, sample members invited to participate in a passive data collection do not have much control over what data is shared once they agree to participate. For example, upon agreeing to participate in a traditional self-report survey, respondents can still choose to respond to a question fully, censor their response, or avoid responding altogether (i.e., item nonresponse). This has become the expected scenario (or norm) for survey participants. However, it is assumed that participants in a smartphone app-based study share all the requested activities on their phone unless they delete the app or turn off required data (e.g., location services), potentially ending their participation in the study. While both are positioned as surveys, the participant experiences are very different. In an effort to align the norms associated with how data is transmitted or shared between the actors (research organization and study participant) in the passive data collection to the self-report survey experience, this work explores whether offering participants control over what data is shared with researchers' impacts participation in an actual study invitation.

The current literature exploring offers of control within a passive data collection generally finds positive but somewhat mixed results in terms of the effectiveness of such an offer. Keusch et al. (2019) found offering sample members the ability to temporarily turn off data collection increased willingness to participate in a study that sample members knew to be hypothetical. In other work utilizing an invitation to a hypothetical study, Struminskaya et al. (2020b) found an offer of control to edit the data after the data collection did not increase willingness to participate in an app-based study. However, later work by Struminskaya et al. (2021) found that an offer of control to edit the data after the data collection increased geolocation sharing. Finally, work by Wenz and Keusch (2023) directly compared offering sample members the ability to temporarily turn off data collection and offering control to edit the data after the data collection. Wenz and Keusch (2023) found that either offer of control increased willingness to participate in a study that sample members knew to be hypothetical, with an offer of editing data increasing willingness slightly more than an offer to turn off the data collection temporarily.

These studies offered control at different time points in the data collection: an offer before the data is collected (i.e., pre-emptive control) compared to an offer after the data is collected (i.e., reactive control in which collected data can be edited). A particular offer might be preferable to respondents for different reasons. For example, invited sample members may be aware that they already possess such pre-emptive control by their ability to turn their phone off or leave it at home when they go out. Such awareness makes an offer of pre-emptive control unnecessary and might not be an attractive feature of the study. The findings by Keusch et al. (2019) and Wenz and Keusch (2023) suggest this might not be the case, but in the study invitation, the sample members were aware of the hypothetical nature of the study. The current work aims to test this offer within the setting of a study that is perceived to be real.

Pre-emptive control also differs from reactive control in the number of interactions the participant has with the data collection tool. With pre-emptive control, the participant is asked to consider their actions in real-time and then interact with the tool to turn off the data collection. Depending on how comfortable the participant is with the data collection tool, this may be burdensome and difficult. This also raises the question of the participant's options if they forget to turn off the data collection for a particular activity. On the other hand, reactive control allows the participant to go about their normal activities and only asks for a single interaction point- to review and edit the collected data. However, participants may be concerned that edited information is still visible to the researcher, preferring instead to have certain activities not recorded at all. A third option might also exist, in which participants prefer both pre-emptive and reactive control together to protect their data best.

The current work moves away from presenting the study as hypothetical to determine whether control impacts participation at the moment of an actual invitation, as the relationship between hypothetical willingness to participate and actual participation is unclear. This work further explores whether offering pre-emptive control (similar to Keusch et al. 2019; Wenz & Keusch 2023) or reactive control (similar to Struminskaya et al. 2020b; Struminskaya et al. 2021; Wenz & Keusch 2023) alone have a different impact than when both forms of control are offered together as these offers of control have never been tested together to determine which may have a larger impact on participation.

Literature Review

The Expected Norms of a Passive Data Collection

Bradburn (2015) refers to an expected "script" between the survey respondent and the interviewer. This is aligned with Nissenbaum's (2010) definition that a script guides the expected

interactions between people in different scenarios – for example when interacting with a server in a restaurant, both customer and server know the "role" that they play in the interaction. This script extends to the expected interactions between an individual and an institution. As Nissenbaum states: "By calling its online offering a university, a shoe store, a church, a medical center, a friendship network, or a bank, a company gives users a way to understand the services or activities that take place there, and it invites evaluation against respective norms, whether these are embodied in law or simply arise from reasonable expectations."

Similarly, by calling a passive data collection a survey or connecting it to a survey (e.g., embedded within a web-based survey), the research institution gives respondents a way to understand what participation may be like. However, as the experience of a passive data collection is quite different from a self-report survey, sample members may question if the expected norms related to the privacy of the respondents and truthfulness from the research organization still hold (Bradburn 2015). If a sample member suspects these norms do not hold or if the expected script does not occur, this might trigger anxiety or resistance, resulting in a refusal to participate.

Specifically within the context of passively collecting data within a public location (such as monitoring with cameras in a public park setting), Nissenbaum states, "If people expect to be monitored, if they anticipate that their recorded views will be shared with particular third parties for money or favors, they are likely to be more watchful, circumspect, or uncooperative." The discomfort with being monitored relates directly to passive data collections in which the participant knows they are being tracked and have limited to no control over what is tracked once they agree to share their data. The current work will provide insight into how giving participants control over how they are monitored or how the data is tracked impacts participation in a passive

data collection, with the expectation that sample members will be more likely to participate when they have more control over how they are monitored.

An Offer of Control

Instilling the expected survey norms for the respondents may be possible by offering control over what is shared. This control may be offered by allowing respondents to turn off the passive data collection at a certain point or allowing respondents to review and edit the material before it is sent to the researchers, which is similar to a respondent choosing to skip or not respond to a question on a traditional self-report survey. The ability to control what is shared can be considered from two different points in time: a pre-emptive offer of control in which the data is never collected and a reactive offer of control in which the data is collected but not shared.

Pre-Emptive Control

Recruiting respondents from a German nonprobability panel, Keusch et al. (2019) used vignettes to investigate the effects of different study characteristics on expressed willingness to participate in hypothetical studies involving passive mobile data collection. The vignettes described hypothetical research involving an app that participants would be asked to download to their smartphone to allow for passive mobile data collection by researchers. The survey sample was restricted to those known to have a smartphone, as they previously responded to an early wave of the survey on such a phone. In addition to other study characteristics, the vignettes varied in terms of whether or not there was an option to switch off the app during the field period. After seeing the vignette, the respondents were asked to rate their willingness to participate in the described study on a scale from 0 (“Definitely would not”) to 10 (“Definitely would participate”), which was dichotomized into those that would not participate (0-5) and those who would participate (6-10) and used as the dependent variable in the study. The authors found sample members were significantly more likely to express a willingness to participate if the app

had the option to temporarily switch off data collection, with an average marginal effect of 4 percentage points ($p < 0.001$) (or 3.47 higher odds). While this has not yet been tested, the ability to program an app to turn tracking off is feasible for survey research as many existing apps allow turning on and off tracking, such as location tracking. Further, it is important to note that sample members invited to participate in a mobile app-based study already possess the ability to pre-emptively avoid data collection by either turning their phone off or leaving it at home, for example, if they do not want an activity tracked. Such a positive impact might be interpreted as the value the sample member places on being able to keep their phone on or with them during the course of the data collection through the offer of an additional mechanism to pause the data collection.

Reactive Control

Struminskaya et al. (2021) experimentally varied participant's control over an actual data collection and found sample members were willing to share certain types of data at higher rates if given the option to edit what is shared after data collection. This study utilized a cross-sectional general population randomized experiment embedded within a web survey administered by Statistics Netherland. Willingness to share sensor data was measured using a binary choice (yes/no), and the individual data requests were asked in one fixed order (geolocation, a video, and photos of the house, receipt, and self). Respondents were randomized in the same conditions for all five data-sharing requests such that if they were offered the ability to edit their data, it was offered for all five types of data requested.

Struminskaya et al. (2021) found offering participants control over their data by allowing for editing after the data collection increased the sharing of geolocation (as well as sharing a photo of a receipt) when respondents were asked about their willingness to share five different types of

data with geolocation asked first. Emphasizing control over the geolocation data increased willingness to share by 8.3 percentage points (p=.027) compared to the control in which no editing was offered. Within a logistic regression, control over the data collection had an average marginal effect of 0.11 for geolocation data. This was 1.62 times higher odds for willingness to share geolocation.

However, in other work, Struminskaya et al. (2020b) found the offer of control (which stated that the respondents would be able to view and change their data later in the questionnaire) did not increase willingness to participate in a hypothetical app-based study. Struminskaya et al. (2020b) focused on survey respondent's willingness to share data collected using smartphone sensors (GPS, camera, and wearables) in a probability-based online panel of the general population of the Netherlands (the LISS Panel). The experiment varied several study factors, including the emphasis placed on control over the data collection process. In the survey, when respondents were asked about their willingness, it was made clear that no actual measurements would be taken. The random assignment to the experimental conditions was performed at the respondent level, which meant that the conditions into which a respondent was randomized were repeated for all five questions. The authors note that emphasizing control did not influence willingness, which is inconsistent with previous research (Keusch et al. 2019). However, the authors also found that respondents who indicated they would want to receive feedback in summary reports had a significantly higher willingness. They hypothesize that the new forms of measurements using smartphones and wearables may create expectations about feedback that influence willingness. Hence, the responsible factor here might be interest rather than the ability to control data collection.

The conflicting results between the two Struminskaya et al. studies may result from of a number of differences between the studies, first, one is known to be hypothetical (in which control has no effect) and the other is a true data collection, which may change the value the sample member places on the offer of control. This offer might be more important and impactful in a real study as respondents decide to participate than in a study in which no data will be collected. Second, a question order effect may be present. In the true data collection, respondents were asked to share five types of data, shown in the same order for each respondent (geolocation, a video, and photos of the house, receipt, and self), while for the earlier hypothetical study, the order of the requests was randomized. As such, the positive effect may partly be a question-order effect. The authors note that in studies that randomized the order of multiple data-sharing requests, the first request achieved higher willingness than later requests (Silber et al. 2018; Walzenbach et al. 2019).

Pre-Emptive vs. Reactive Control

Recent work by Wenz and Keusch (2023) directly compares these offers of control within a vignette-style study that invited respondents to participate in an app-based passive data collection that offered ,among other things, the option to either (1) switch off the app during the field period, (2) to edit the data collected before it is shared with researchers or (3) provided no additional control over the data which was stated as: "*There would be no option to switch the research app on and off during the course of the study*." The sample was drawn from the NORC AmeriSpeak Panel (a probability-based online panel of the general population aged 18+ in the United States). The authors randomly assigned respondents to eight vignettes (without replacement) that described the hypothetical studies to which they would invite respondents to participate. The studies included downloading a research app on their smartphone that would administer survey questions about the individuals' time use and collect data about the technical characteristics of their phone, whether their phone is currently in motion, their current location,

what apps are used, and what websites are visited, as well as the number of incoming and outgoing phone calls and text messages on their phone. After seeing the vignette, the respondents were asked to rate their willingness to participate in the described study on a scale from 0 ("Definitely would not") to 10 ("Definitely would participate"), which was dichotomized into those that would not participate (0-5) and those who would participate (6-10) and used as the dependent variable in the study (like in Keusch et al. 2019). In total, 44 percent of respondents being willing to download the research app to participate in the described study. Through a series of multilevel logistic regression models, Wenz and Keusch (2023) concluded that giving respondents more control over their data significantly increases their willingness to participate. In total, respondents had a 6 percentage points higher predicted probability of being willing to participate in the mobile app-based data collection if they can temporarily turn off the data collection compared to not having this option. There was an 8 percentage points higher predicted probability of willingness if they are offered the option to review and edit the data that has been collected.

When adding interactions to understand better how respondent's privacy concerns interact with the offers of control, Wenz and Keusch (2023) found a positive or neutral relationship between privacy concerns (in one of their four measures for privacy) and willingness to participate given the offer of control – which was not expected. Giving more control over the data collection did not significantly affect willingness (compared to no offer of control). Lastly, when the authors add demographics to their models, the results show that giving respondents control over their data collection seems to be more effective in increasing participation rates for those with higher levels of education, with those who have obtained a college degree being more willing to participate in a mobile app-based study if they can temporarily switch off the data collection or

review the data before submission as opposed to not having either of these options. Giving these options to individuals with lower levels of educational attainment did not lead to significant increases in willingness to participate.

Overall, while work by Keusch et al. (2019), Struminskaya et al. (2021), and Wenz and Keusch (2023) suggests that both pre-emptive and reactive offers of control should lead to higher willingness to participate in a passive data collection, some questions remain. In terms of pre-emptive control, Keusch et al. (2019) and Wenz and Keusch (2023) tested this hypothesis within a hypothetical study format. As such, the effect they found may differ from a study in which respondents expect to have such data passively tracked. Respondents might also recognize that they already possess such pre-emptive control by either turning their phone off or leaving it elsewhere, and such an offer of control might not appear as attractive. Further, the null finding by Struminskaya et al. (2020b) was contradictory to the positive impact of an offer of reactive control found by the later Struminskaya et al. (2021) and the Wenz and Keusch (2023) study. While there may be two factors at play here - the difference between the hypothetical and real data collections, as well as the ordering of the requested data in the Struminskaya et al. studies - the current work hopes to provide some clarity as to the impact of an offer of reactive control. Finally, the two offers of control (pre-emptive and reactive) have never been tested collectively. It might also be expected that providing both offers together should capture respondents who prefer one over the other, leading to a larger increase in participation. It might also be that such an offer is equivalent to the offer of reactive control if sample members recognize they already possess the ability to take pre-emptive control of their data. This work builds on the previous studies by experimentally offering the three forms of control (pre-emptive, reactive, and a joint

offer) in a data collection request that the respondents will perceive to be real rather than hypothetical.

Hypothesis and Research Questions:

RQ1: Does offering sample members both pre-emptive and reactive control over their collected data increase participation in a mobile app-based passive data collection compared to offering sample members only pre-emptive control, reactive control, and no control?

> H1: It is expected that an offer of control (of any form) over the data collected during a passive data collection will increase participation in a mobile app-based passive data collection.

RQ2: Do privacy and burden concerns impact which offer of control is more effective at increasing participation rates, such that reactive control may be more effective for a respondent with concerns over burden and pre-emptive control may be more effective for a respondent with greater privacy concerns?

Methods

Study Design

An experiment with 4 different offers of control was conducted – no control, pre-emptive control, reactive control, and both pre-emptive and reactive control. Sample members were recruited through the GfK KnowledgePanel to participate in a geotracking passive data collection, with the survey coded in Qualtrics and administered in February 2023. The KnowledgePanel, recently purchased by Ipsos, is the oldest and largest probability-based online panel in the United States, with approximately 60,000 members. The panel was built through address-based sampling (ABS).

As a smartphone is required to complete the app-based geotracking passive data collection, respondents were restricted to completing the survey on a smartphone. Upon opening the emailed invitation, respondents received a short survey to gauge their relationship with their smartphone. Questions included the frequency of turning their phone off, their tendency to keep their phone with them, the intensity of smartphone use, and their comfort in installing apps.

After the short survey, respondents were invited to participate in the passive data collection. An invitation to a geotracking study was designed to match the passive data collection in the study by Struminskaya et al. (2021). Sample members were randomly assigned to receive introductions to the passive data collection that varied in terms of the offered control. However, it is important to note that this was a deception study and that no data was collected if the respondent was interested in continuing with the follow-up study. If sample members chose not to participate in the passive data collection, they were asked to describe in their own words why they did not want to participate.

Regardless of whether the respondent indicated they wanted to participate in the study or not, all respondents were informed of the deception at the end of the survey and offered the opportunity to be removed from the study. In total, 12% of all respondents withdrew their data, resulting in a loss of 79 observations, as described below in Table 1. Overall, the requests to be withdrawn were fairly even across all experiments, with slightly more requests from respondents who initially refused to participate in the passive data collection.

Table 1: Requests to be withdrawn

	Originally Agreed to Participate	Originally Refused to Participate
No Information Provided (n=18)	9	9
Pre-emptive Information Provided (n=21)	9	12
Reactive Information Provided (n=17)	6	11
Both Pre-emptive and Reactive Information Provided (n=23)	8	15

Those respondents who indicated they did not want to participate in the geotracking passive data collection received an open-ended question about why they did not want to participate. See Appendix C for full question wording.

Data

Respondent Information

In total, 570 respondents were included in this study, with approximately 140 respondents assigned to each invitation. As respondents were randomly assigned to one of four invitations, demographic differences across the four experimental groups are first evaluated. A series of logistic regressions with the outcome variable (1) being assigned to the experiment (and 0 otherwise) confirms that the demographics do not differ more than by chance. None of the respondent demographics are significantly related to the assigned experiment, except for those with a mid-level income being assigned to the preemptive experiment at higher rates ($p=0.018$) compared to the lowest income category. As shown in Table 2 below, the demographics do not vary widely, and the final analysis controls for income to account for this difference.

Table 2: Demographics of the sample, by experiment

		Full Sample	Reactive	Preemptive	Both	Control
Education	No high school diploma or GED	6.7 [4.9,9.0]	5.7 [2.4, 13.0]	13.3 [7.5, 22.4]	4.1 [1.3, 11.9]	3.8 [1.2, 11.2]
	High diploma or equivalent	28.2 [24.7,32.1]	25 [17.0, 35.1]	22.9 [15.1, 33.2]	28.4 [19.3, 39.7]	32.9 [23.5, 44.0]
	Some college or Associate degree	25.6 [22.2,29.4]	30.7 [21.9, 41.1]	30.1 [21.2, 40.8]	23 [14.8, 33.9]	21.5 [13.8, 32.0]
	Bachelor's degree	21.9 [18.7,25.5]	22.7 [15.1, 32.7]	20.5 [13.1, 30.5]	24.3 [15.9, 35.4]	29.1 [20.2, 40.1]
	Master's degree or above	17.5 [14.6,20.9]	15.9 [9.6, 25.1]	13.3 [7.5, 22.4]	20.3 [12.6, 31.0]	12.7 [6.9, 22.0]
Race/Ethnicity	2+ Races, Non-Hispanic	2.8 [1.7,4.5]	3.4 [1.1, 10.1]	3.6 [1.2, 10.6]	8.1 [3.7, 16.9]	1.3 [0.2, 8.5]
	Black, Non-Hispanic	10.2 [7.9,12.9]	15.9 [9.6, 25.1]	9.6 [4.9, 18.1]	10.8 [5.5, 20.2]	8.9 [4.3, 17.5]
	Hispanic	14.7 [12.1,17.9]	14.8 [8.8, 23.8]	18.1 [11.2, 27.9]	16.2 [9.4, 26.5]	15.2 [8.8, 24.9]
	Other, Non-Hispanic	6.1 [4.4,8.4]	6.8 [3.1, 14.4]	7.2 [3.3, 15.2]	6.8 [2.8, 15.3]	5.1 [1.9, 12.8]
	White, Non-Hispanic	66.1 [62.1,69.9]	59.1 [48.5, 68.9]	61.4 [50.6, 71.3]	58.1 [46.6, 68.8]	69.6 [58.6, 78.8]
Gender	Female	50.2 [46.1,54.3]	52.3 [41.9, 62.5]	39.8 [29.8, 50.6]	51.4 [40.1, 62.5]	55.7 [44.6, 66.2]
	Male	49.8 [45.7,53.9]	47.7 [37.5, 58.1]	60.2 [49.4, 70.2]	48.6 [37.5, 59.9]	44.3 [33.8, 55.4]
Income	Under $10,000	3.2 [2.0,5.0]	4.5 [1.7, 11.5]	2.4 [0.6, 9.2]	4.1 [1.3, 11.9]	2.5 [0.6, 9.6]
	$10,000 to $24,999	7.0 [5.2,9.4]	13.6 [7.9, 22.5]	10.8 [5.7, 19.6]	0	6.3 [2.7, 14.4]
	$25,000 to $49,999	15.3 [12.5,18.5]	14.8 [8.8, 23.8]	15.7 [9.3, 25.2]	17.6 [10.5, 28.0]	19 [11.8, 29.2]
	$50,000 to $74,999	14.6 [11.9,17.7]	13.6 [7.9, 22.5]	13.3 [7.5, 22.4]	13.5 [7.4, 23.4]	16.5 [9.8, 26.3]
	$75,000 to $99,999	13.0 [10.5,16.0]	15.9 [9.6, 25.1]	8.4 [4.1, 16.7]	8.1 [3.7, 16.9]	10.1 [5.1, 19.0]
	$100,000 to $149,999	18.9 [15.9,22.4]	13.6 [7.9, 22.5]	20.5 [13.1, 30.5]	23 [14.8, 33.9]	20.3 [12.8, 30.6]
	$150,000 or more	28.1 [24.5,31.9]	23.9 [16.1, 33.9]	28.9 [20.2, 39.6]	33.8 [23.9, 45.3]	25.3 [16.9, 36.1]
MSA Status	Metro	86.1 [83.0,88.7]	87.5 [78.8, 93.0]	89.2 [80.4, 94.3]	78.4 [67.5, 86.3]	83.5 [73.7, 90.2]
	Non-Metro	13.9 [11.3,17.0]	21.5 [7.0, 21.2]	10.8 [5.7, 19.6]	21.6 [13.7, 32.5]	16.5 [9.8, 26.3]
Region of USA	Midwest	21.2 [18.1,24.8]	20.5 [13.3, 30.2]	21.7 [14.1, 31.9]	24.3 [15.9, 35.4]	19 [11.8, 29.2]
	Northeast	16.0 [13.2,19.2]	19.3 [12.3, 28.9]	10.8 [5.7, 19.6]	13.5 [7.4, 23.4]	11.4 [6.0, 20.5]
	South	38.4 [34.5,42.5]	39.8 [30.1, 50.3]	34.9 [25.5, 45.8]	33.8 [23.9, 45.3]	46.8 [36.1, 57.8]
	West	24.4 [21.0,28.1]	20.5 [13.3, 30.2]	32.5 [23.3, 43.3]	28.4 [19.3, 39.7]	22.8 [14.8, 33.3]
Employment	Not working	34.7 [30.9,38.8]	36.4 [27.0, 46.9]	34.9 [25.5, 45.8]	33.8 [23.9, 45.3]	40.5 [30.3, 51.7]
	Working full-time	53.2 [49.0,57.2]	47.7 [37.5, 58.1]	56.6 [45.8, 66.9]	52.7 [41.4, 63.8]	51.9 [40.9, 62.7]
	Working part-time	12.1 [9.7,15.1]	15.9 [9.6, 25.1]	8.4 [4.1, 16.7]	13.5 [7.4, 23.4]	7.6 [3.4, 15.9]
Age	18-25	11.9 [9.5,14.9]	15.9 [9.6, 25.1]	8.4 [4.1, 16.7]	8.1 [3.7, 16.9]	13.9 [7.9, 23.5]
	26-35	16.1 [13.3,19.4]	14.8 [8.8, 23.8]	14.5 [8.4, 23.8]	16.2 [9.4, 26.5]	17.7 [10.8, 27.8]
	36-45	20.9 [17.7,24.4]	15.9 [9.6, 25.1]	26.5 [18.1, 37.0]	21.6 [13.7, 32.5]	16.5 [9.8, 26.3]
	46-55	16.5 [13.7,19.8]	14.8 [8.8, 23.8]	15.7 [9.3, 25.2]	18.9 [11.5, 29.5]	19 [11.8, 29.2]
	56-65	17.5 [14.6,20.9]	20.5 [13.3, 30.2]	16.9 [10.2, 26.5]	21.6 [13.7, 32.5]	16.5 [9.8, 26.3]
	66-75	13.9 [11.3,17.0]	14.8 [8.8, 23.8]	13.3 [7.5, 22.4]	12.2 [6.4, 21.8]	12.7 [6.9, 22.0]
	76+	3.2 [2.0,5.0]	3.4 [1.1, 10.1]	4.8 [1.8, 12.2]	1.4 [0.2, 9.0]	3.8 [1.2, 11.2]

Sample members report regularly keeping their phone on and with them, as shown in Table 3 below. However, as described above, respondents have control outside of the data collection tool to pre-emptively stop their data from being collected. Overall, nearly two-thirds of the respondents to this survey report at least occasionally turning off their smartphones, with just over a quarter (26%) reporting that they turn their phone off most days or a few times a week. Further, just about half of the respondents (49%) do not have their phone on them during some portion of the day. In total, 15% of the respondents report they turn their phone off most days or

a few times a week and only keep their phone on them during only a portion of the data, which is approximately equal across all experiments. This reiterates that respondents can act pre-emptively to avoid a data collection but do not typically have the ability to act reactively – or to review and edit data that may have been collected while the phone is on or with them. The current work aims to understand the value respondents may place on being offered further pre-emptive control compared to being offered reactive control within the data collection tool.

Table 3: Respondent's Relationship with Smartphone

		% (SE)
Frequency of Turning Phone Off	Most days	16.5 (1.6)
	A few times a week	9.1 (1.2)
	A few times a month	15.8 (1.5)
	Less than once a month	31.4 (1.9)
	Never	27.2 (1.9)
Frequency of Keeping Phone with them	It is with me some of the day	6.1 (1.0)
	It is with me most of the day	42.6 (2.1)
	It is always with me	51.2 (2.1)
Activities with Smartphones (*Activity with geolocation*)	Playing Games	25.4 (1.8)
	Tracking Steps	40.5 (2.1)
	Receiving Targeted Coupons	41.4 (2.1)
	Finding Transportation	45.3 (2.1)
	Connecting to Bluetooth	80.7 (1.7)
	Finding Restaurants, Events, and Shops	82.4 (1.6)
	Interacting with Social Media	82.5 (1.6)
	Browsing Websites	93.7 (1.0)
	Following Direction	94.6 (1.0)
	Checking Weather	96 (0.8)
Intensity of Smartphone Use	Less than once a month	1.4 (0.5)
	A few times a month	1.6 (0.5)
	A few times a week	4.4 (0.9)
	Every day or most days	13.7 (1.4)
	Multiple times a day	78.6 (1.7)

When asked how frequently respondents use their phone for activities other than phone calls or text messages, 92% report they do such activities every day or most days, with the vast majority (79%) of the respondents reporting they do such activities multiple times a day. When asked

specifically about activities they might be doing on a smartphone, the top two activities reported require their phone to know a location – checking the weather near the respondent (96%) and following directions from a location-aware app (95%). Further, over 80% of respondents connect to Bluetooth-enabled devices (which require close proximity) and find restaurants, events, and shops near them - suggesting that the vast majority of respondents are willing to provide their location to interact with their smartphones and immediate physical environment.

However, when asked why a respondent chose not to participate in the geotracking passive data collection, 43% indicated that something about the request for a location is just "too much" for them to agree to, as shown in Tables 4 and 5 below. An additional 19% stated the level of burden they associated with the study was too high – e.g., their phone did not have enough storage space, or they did not have enough time. Privacy was cited as the next most common reason (8%) for not participating, at the same rate that trust was cited (8%). Concerns related to the level of incentives and the impact of a previous experience were also cited at a lower rate (2%).

Table 4: Open response categorization for why a respondent choose not to participate

n=195	**n**	**%**	**Example**
Study request is "too much"	83	43%	"I am nervous about being tracked in this way"
Burden	37	19%	"I don't have room on my phone." "I don't have time for that."
Miscellaneous	35	18%	"Just no." "I am not interested."
Privacy	16	8%	"Privacy concerns."
Trust	15	8%	"Don't trust this."
Too low of incentive	4	2%	"Not enough monitory incentive. I would not accept anything below $100."
Negative previous experience	3	2%	"I downloaded an app once for vpn and it kept me from logging in to my job and caused issues"

When the open response categories were split by the experimental groups, the top issue cited continued to be the request for a location is just "too much" for them to agree to, ranging from 36% (in the pre-emptive study) to 50% (in the combined offers of control study).

Table 5: Open response categorization for why a respondent choose not to participate, by experimental group

	Open Response Category	**n**	**%**
No Info (n=55)	Study request is "too much"	25	45%
	Burden	10	18%
	Privacy	9	16%
	Trust	3	5%
Pre-emptive (n=50)	Study request is "too much"	18	36%
	Burden	11	22%
	Privacy	6	12%
	Trust	5	10%
Reactive (n=47)	Study request is "too much"	19	40%
	Burden	5	11%
	Privacy	7	15%
	Trust	8	17%
Pre-emptive & Reactive (n=42)	Study request is "too much"	21	50%
	Burden	2	5%
	Privacy	4	10%
	Trust	2	5%

Experimental Conditions

Within the invitation to participate in the geotracking study, four offers of control were offered – no control, pre-emptive control, reactive control, and both pre-emptive and reactive control – as shown in the invitations in Figure 1 below. Each was offered to approximately 140 respondents.

Figure 1: Invitations to Studies

Invitation to Smartphone App Study

[PAGE ONE]

We would like to invite you to participate in an additional app-based study that focuses on **where Americans use their smartphones**.

To participate in the study, please download the app provided on the next page and keep it on your phone for one week. You will receive $20 for your participation, with $10 provided for downloading the app and $10 provided at the end of the week long data collection.

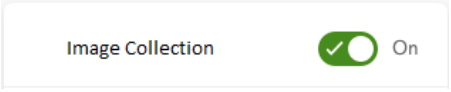
The app will record your location on a map once every hour, as displayed here. (Click "next page" to read more.)

___

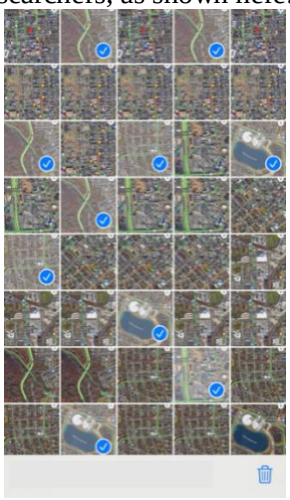
[PAGE TWO]

<u>Important additional study information:</u>

- Your data will remain anonymous and will never be used in a way that would personally identify you.
- *[Pre-emptive Control Offer]* The image collection can be turned off if you feel uncomfortable with sharing your location, as shown here.



- *[Reactive Control Offer]* At the end of the week, you will be provided with the images to review. You may edit the series of images before they are sent to our researchers, as shown here.



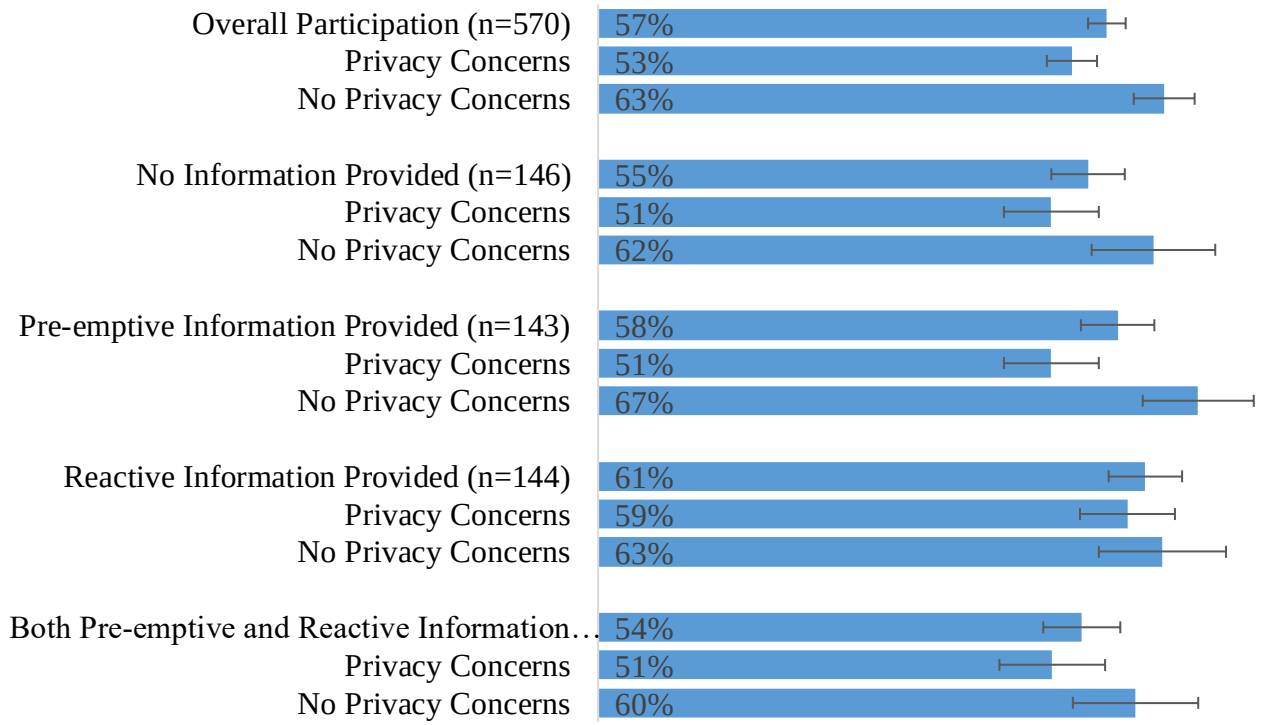
Results

Willingness to Participate

Overall, 57% of the respondents indicated they wanted to participate in the geotracking passive data collection. This participation rate is much higher than expected as the range of participation rates found from a literature review is between 12% and 30% (as shown in Figure 1 in the Preface). The high willingness to participate may be related to the sample being restricted to smartphone users who are part of an online panel, as panelists are familiar with participating in surveys and smartphones. As shown in Figure 2 below, there is very little difference in the participation rates across the invitations, though there is some variation when the privacy concerns of the respondents are taken into account. Because none of the experimental groups differ significantly from each other, the answer to RQ1 is No and the H1 is not supported.

In terms of privacy concerns, those who indicated they are concerned with sharing their information over the Internet participate in the passive data collection at lower rates than those who indicate they do not have any concerns. This is true for the overall participation rate and those respondents assigned to the experimental group, for which the invitation to the passive data collection offered them pre-emptive control of the data collected in the app, as shown by the non-overlapping confidence intervals. The offer of pre-emptive control may have reminded the respondent they already have such control, which increased their willingness to participate.

Figure 2: Participation Overall, by Experimental Treatment, and Privacy Concerns



Burden Measures

To better understand the impact of burden on participation in the passive data collection, respondents received the following question in the survey:

Measure of Burden: Ability to complete tasks, following Jäckle et al. (2019) and Read (2019)

> How comfortable are you with installing new apps (e.g. from iTunes, Google Play Store) on a smartphone? (a) very comfortable (b) comfortable (c) not comfortable

Table 6 summarizes the responses to the measure of burden question above.

Table 6: Comfort with Installing Apps (for all respondents)

	% (SE)
Not comfortable	8.1 (1.1)
Comfortable	27.9 (1.9)
Very comfortable	64.0 (2.0)

Respondents who indicated they were comfortable or very comfortable with installing an app agreed to participate in the geotracking study at very similar rates (56% and 59%, respectively).

Those who reported they were uncomfortable agreed to participate 39% of the time. This makes sense as participants in the passive data collection are required to download an app to participate. For the sake of analysis, this measure will be dichotomized, combining "comfortable" and "very comfortable" together is a low-burden group and leaving "not comfortable" in the high-burden group.

Those that indicated they weren't comfortable with installing apps and received a pre-emptive offer of control were willing to participate at a much higher rate (56%) than the other invitations. It may be that those who received the offer of pre-emptive control were reminded this is a control they possess outside of the data collection app as well, which inflated their willingness to participate. However, it might also be that burden here is not capturing the same burden being expressed by the respondents, which was reported more along the lines of space constraints on a smartphone or no available time in the open-ended responses.

Privacy

To better understand the impact of privacy on participation in the passive data collection, respondent's answers to the following question were used:

Measure of Privacy: Concern about sharing personal information on the Internet (IPSOS Household Survey)

> How concerned are you about providing personal information over the Internet? (a) not at all concerned (b) slightly concerned (c) somewhat concerned (d) very concerned

As shown in Table 7 below, very few of the respondents feel "not at all concerned" while over half (62%) have higher levels of concern, indicating that they are "somewhat" or "very" concerned with sharing their personal information over the Internet. For the sake of analysis, this measure will be dichotomized, combining "not at all concerned" and "slightly concerned" together in a no- or low-concern group and combining "somewhat concerned" and "very concerned" into a higher-concern group.

Table 7: Concern about sharing personal information over the Internet

	% (SE)
Not at all concerned	4.5 (0.9)
Slightly concerned	33.9 (2.1)
Somewhat concerned	36.9 (2.1)
Very concerned	24.7 (1.9)

Logistic Regressions

Lastly, logistic regressions are used to understand further how the offers of control are associated with participation, including the respondent characteristics of privacy and burden concerns. While none of the experimental groups differ significantly from each other in Table 2 above, there is variation by level of education (especially by the lowest education group), employment status, race/ethnicity, age, gender, income, metro area, and region of the country with the estimates falling outside the comparable confidence intervals by a percentage point or two. To account for this variation, the logistic regressions below control for these demographics.

As shown in Table 8 below, the dependent variable for these logistic regressions measures whether the sample member decided to participate in the passive data collection or not, modeled as 1 for participating and 0 for not participating. Model one simply included the four offers of control from the invitations, as shown in equation one. As expected, there were no significant relationships between the offers of control and the decision to participate, with or without the additional controls for respondent demographics.

$$\text{logit (p)} = \log(\frac{p}{1-p}) = \beta_o + \beta_1 Control_l \qquad \text{[Equation 1]}$$

Model two builds from model one by including the specific respondent characteristics of perceived burden and privacy concerns, as described above and shown below in equation two. Equation two includes the binary variable of privacy, modeled as 1 for respondents who indicated they were “not at all concerned” or “slightly concerned” in providing their personal

information over the Internet and 0 for respondents who indicated they were "somewhat concerned" or "very concerned" indicating these respondents had high levels of privacy concerns. Equation two also includes the binary variable of burden, modeled as 1 if the respondent indicated they were "very comfortable" or "comfortable" installing apps on their smartphone and 0 for respondents who indicated they were "not comfortable" indicating they would experience more burden by installing the passive data collection app on their smartphone. Together, the interactions show the respondents who choose to participate in the passive data collection with lower privacy concerns and those who would experience less burden when installing the passive data collection app.

$$\text{logit (p)} = \log(\frac{p}{1-p}) = \cdots + \beta_2 ResChar_2 + \beta_3 (Control * ResChar)_3 \quad \text{[Equation 2]}$$

Table 8: Regression Output

Dependent variable: Participation	Equation 1		Equation 2	
	Coeff	p	Coeff	p
Level of Control (No Control is Comparison)				
Pre-emptive Control	0.160	0.501	1.009	0.295
Reactive Control	0.287	0.228	-0.318	0.750
Both Pre-emptive and Reactive Control	-0.004	0.987	-0.288	0.747
Respondent Characteristics				
Privacy			-0.304	0.383
Burden			0.951	0.138
Interactions (Both Pre-emptive and Reactive Control)				
No Information * Privacy			0.050	0.920
Pre-emptive Control * Privacy			-0.367	0.456
Reactive Control * Privacy			0.170	0.734
No Information * Burden			-0.302	0.724
Pre-emptive Control * Burden			-1.011	0.289
Reactive Control * Burden			0.235	0.810

Overall, the regressions indicate no differences by invitation or respondent characteristics with or without the additional controls for respondent demographics.

Discussion

Concerns over the privacy of data collected during a passive data collection are a large barrier to participation (Keusch et al. 2019; Revilla 2019; Chin et al. 2012). Previous research exploring offers of control given to respondents either pre-emptively (Keusch et al. 2019; Wenz & Keusch 2023) or reactively (Struminskaya et al. 2021; Wenz & Keusch 2023) suggested that either offer of control should lead to higher willingness to participate in a passive data collection. However, the current analysis resulted in a null finding, similar to Struminskaya et al. (2020b), whose research included an offer of reactive control. Overall, the current work found no relationship between participation in a geotracking study and offering respondents control over the collected data.

Given the study design of the current work, there are a few suspected reasons why the offers of control presented in the introduction did not increase participation in the passive data collection. First, as there was an unexpectedly high willingness to participate in the passive data collection (57%), it might be that the survey sample was drawn from a population that is generally more willing to participate in such studies than the general population of smartphone owners targeted by Keusch et al. (2019) or Struminskaya et al. (2021). However, the sample drawn in the current work is theoretically similar to that drawn by Wenz and Keusch (2023), as both are drawn from probability-based online panels of Americans. The differences are that Wenz and Keusch (2023) included population-based weights in their analysis and had a larger sample size (n=1,876) compared to the current work, which did not include population weights and had a smaller sample (n=570). Wenz and Keusch (2023) found a 44% rate of willingness to participate, while Keusch et al. (2019) and Struminskaya et al. (2021) saw rates of approximately 35% of respondents that would be willing to participate/did participate.

With such a willing sample as the one from the current work, any sample member who chooses not to participate might not be influenced to participate by an offer of control. It might be that requesting location data in real-time may be too concerning regardless of the parameters set for controlling the data – which is suggested by the responses to the open-ended questions. Or those who chose not to participate did not believe the offers of control and so could not be recruited. Future research should consider varying other elements of the invitation. As suggested by Struminskaya et al. (2020b), offering feedback to the respondents based on the data collected might be a useful tool to recruit more participants.

Second, the sample was recruited from an active online-based probability panel. As such, the sample was recruited from a population of people who use a smartphone and are familiar with taking surveys – with many likely taking surveys on a smartphone. This may result in a population that feels comfortable with controlling the data that can be collected from their phone. Alternatively, this population has already shared so much through their devices, such as computers and phones, that their concerns over the privacy of their data may be lower than the general population of smartphone owners in the United States. Future research should consider retesting this experiment by recruiting respondents from the general population of smartphone users rather than those associated with an online panel and regularly taking surveys online or on a smartphone.

Third, this study was conducted two years after the next earliest work by Wenz and Keusch (2023), which administered their data collection in July and August of 2021 compared to February 2023. In that time, the general attitudes held by smartphone owners in the United States may have shifted toward such studies. Wenz et al. (2019) suggest that having been invited to a similar study in the past is positively associated with participating in a current invitation to a

passive data collection. Future research should consider collecting data on prior study invitations and time in an online panel and bringing this into the analysis.

Fourth, in work by Keusch et al. (2021) and Wenz and Keusch (2023), the authors explicitly state in the vignette that: "*There is no option to switch the research app on and off during the course of the study*". It may be that emphasizing that there is no option to switch off the research app caused higher rates of refusal, rather than the offers of control increasing participation more than what would have been expected had the text omitted such language. Struminskaya et al.'s (2020b) work omitted this text in the no-control group and did not find a difference in willingness to participate between the invitations. However, Struminskaya et al. (2021) also omitted this text and found a positive relationship between reactive control and actual participation. In this current work, the invitation that did not offer additional control to the respondent omitted the text altogether. This may be an area for future work, comparing an omitted statement to a direct statement like the one employed by Keusch et al. (2019) and Wenz and Keusch (2023).

Finally, it might be that respondents were willing to say they would participate in the passive data collection and believe that such a data collection will occur but are also aware that they still need to take the additional step of downloading an app to their phone. This additional step allows respondents to opt out of the study even after agreeing to participate.

As detailed in the next section, these suspicions lead to several future research avenues.

Limitations and Next Steps

Given the unexpectedly high willingness to participate in the passive data collection, it is worth replicating this study with another form of recruitment, varying different elements of the study

beyond the offer of control, especially in terms of the type of requested data as the open-ended responses indicate there was something about the current study that was "too much." Another future avenue of research would be to replicate the study by removing the additional step (the requirement to download an app) that participants needed to participate in the passive data collection.

Additionally, the invitation to the collective control that offered both pre-emptive and reactive control should consider randomizing the order in which these offers are presented. The current work presented the offer of pre-emptive control first for each respondent who received the offer of joint control. It may be that the pre-emptive offer is simply not attractive enough, and the respondent did not consider the additional offer of reactive control.

A large limitation of this work is a lack of information on whether the respondents have been invited to similar prior studies, and the amount of time they spent in an online panel. These would be valuable components to include in future analysis.

Conclusion

This work aimed to expand what was known about nonparticipation within passive data collections that use web-based panels to recruit respondents. Overall, a number of similar findings came out of the three chapters in this work. As each chapter utilized a different online panel, this suggests that the findings may be generalizable to other passive data collections that use web-based panels to recruit participants. First, across the chapters, there is evidence that nonparticipation bias within the passive data collection context is a large and looming issue. Secondly, the tools and techniques explored across these chapters did not diminish the impact of nonparticipation bias. Finally, there is evidence that attitudes towards privacy and trust are important to account for within an analysis of passively collected data, especially when key estimates of interest may be related to these attitudes. Future work should further explore the correlation between privacy and participation in a passive data collection, as well as the techniques that might be employed to diminish concerns related to privacy.

Nonparticipation Bias

This work assessed the nature and size of nonparticipation bias in a passive data collection in two different ways. First, for a passive data collection of financial transactions with participants recruited from an online probability panel (Chapter One); second, for participation in a number of different passive data collections embedded within a traditional survey with participants recruited from a nonprobability panel (Chapter Two). Together, these chapters suggest that nonparticipation bias within passive data collection may be a large concern for the resulting estimates of interest, particularly if the estimates of interest are correlated with the attitudes that might drive the decision to participate – privacy concerns and perceptions of trust.

Tools and Techniques to Address Nonparticipation Bias

This work explored two different tools to combat nonparticipation and nonparticipation bias. Chapter two explored whether offering a follow-up video message focused on alleviating privacy and data security concerns might decrease nonparticipation bias in a probability-based panel. However, while it was found that the follow-up message did decrease nonparticipation, those sample members who were recruited to participate were not demographically different from those who had initially agreed to participate – suggesting that the follow-up message would not alleviate any nonparticipation bias. Chapter three then explored whether offering sample members control over the data that was shared with researchers reduced nonparticipation but found no difference in terms of participation rates across various offers of control, suggesting that at least for a probability-based panel in which respondents are familiar with the research organization administering the study, offering additional control over their data likely won't increase participation in the study.

Future Work

A theme running through all these studies was the inclusion of respondent attitudes and perceptions towards passively sharing data to understand differences between sample members who chose to participate and those who refused to participate. In chapter one, it was found that participation in a passive data collection of financial transactions was significantly associated with the respondent having consented to linking their personal information (e.g., credit scores linked through a four-digit SSN number match) to an earlier data collection by the same research organization. The result that trusting in the research organization is positively related to participation in a passive data collection of financial transactions is aligned with the previous findings of the importance of attitudes in participation in passive data collections found by Struminskaya et al. (2021) as well as Elevelt et al. (2019).

However, in chapter three – which focused on whether an offer of control increased participation in a geotracking passive data collection utilizing a probability-based panel – there was no relationship between participation and sample member's attitude of privacy or burden with using smartphones. This is counter to the expectation from the literature and the results on chapter two, which both suggest an offer of control would positively impact participation rates. In chapter two, when asked why survey respondents choose not to participate in the passive data collection – this study experimentally varied invitations to a number of passive data collections embedded within a traditional survey – respondents primarily cited security as a concern and a reason for not participating. This suggests that attitudes towards the research organization might play a larger role in the decision to participate than trusting in the technology used to participate in the passive data collection. However, further research is needed to test this hypothesis.

All three of these studies suggest that nonparticipation bias may impact the final datasets and that traditional refusal conversion techniques (such as follow-up messages) might not be effective. Future work should focus on better understanding how nonparticipants might be converted, especially those who are older and have lower levels of education, so passive data collections might be more representative of the target populations. Future work should also continue to test whether a prior relationship with the research organization results in more participation in passive data collection and the underlying mechanisms for this. It may be that having a prior relationship with the research organization triggers less concern over privacy violations (Nissembaum 2010) as the "script" between the research organization and the sample member is known – even if the interactions of the study itself differ from a traditional self-report study.

Appendix A: Introduction for Financial Tracking Study

We are looking for volunteers for a study about how people in America manage their money.

If you agree to participate, we will ask you to sign up to a secure website where you can track all your financial transactions. There is no cost to using this service. In fact, we will pay you in return for allowing UAS researchers to learn from your everyday transactions.

Your financial information will be secure, protected, and confidential. As with all Understanding America Study surveys you take, the information you share with us is never linked to you individually. We have developed the website in collaboration with Plaid, a company trusted by millions of consumers to connect their bank and other financial accounts such as investments or retirement savings, to help them manage their money. The data we access will include account type, institution holding the account, account balance, transaction dates, transaction amounts, and the type of merchant.

Your passwords and IDs will remain private. Plaid will never share your login IDs, passwords, or any other personal information with UAS researchers or anyone else. We at UAS also protect your private information. For example, when we write up results, we only ever publish tables and graphs that make it impossible to identify anyone individually.

Time required: Your participation in this study will require very little of your time. After you add your accounts, we will ask you to periodically review and update them as needed. You can remove any or all of your accounts from the study at any time without consequences. We also may ask you if you are willing to talk to us about your experience using the site. If so, you will be paid for your time and we will arrange to speak to you when it is convenient. You do not have to agree to do an interview to participate in the rest of the study.

To get started: You will simply create your own private password for the secure Plaid website and then follow prompts to log into it and add financial institutions with which you have accounts.

You will be paid for your participation, now and monthly: You will earn $[[]] for each financial institution that you add. After that you will continue to earn $[[]] every month for each active institution that you have added.

If you have any questions about this study: Please let us know. You can find out more information about Plaid here: https://plaid.com/company. For other questions, just call our help desk at 855.872.8673 (9am to 5pm Pacific; Monday through Friday) or send an email to uashelp@usc.edu.

I am interested in participating.
1 (YES) Yes
2 (NO) No

---

consentyes

Thank you for agreeing to participate in this study!

We have set up a secure website which requires that you set up a new password that only you will know.

Please click on the link below to create the new password for your account on the secure website. Once your password is set, you will be asked to login to the study site – Called UASFin - using username **[PRIMARY KEY[]]** and the new password you created. Please write it down and keep it in a secure place so you will be able to log in to the study site in the future.

SETUP PASSWORD (opens in a new window)

IMPORTANT: After creating your password and adding your accounts, Please remember to return to this page and complete the rest of this survey, to get your $2 reward in addition to the rewards you will earn from linking your accounts in the UASFin site.

Appendix B: Refusal Conversion Questionnaire

(Restriction: The participant must respond on a smartphone.)

General Introduction

[page 1]

Thank you for participating in this survey! Researchers from the University of Maryland are interested in learning about what smartphone owners use their phone to do. This should take no more than 3 minutes and you will be provided $1.00 upon completion by Prolific. Please make sure you are starting this survey on a smartphone!

Click continue to read more about this survey.

[page 2]

Your responses will be kept anonymous used solely for research purposes to improve future smartphone based surveys. Your participation is completely voluntary and you may skip any question you do not want to answer. There are no known risks or direct benefits to participating in this survey. This project has been reviewed by the University of Maryland Institutional Review Board (irb@umd.edu; Package Number 1958251-1).

If you have any questions regarding this research, please reach out to Alexandra Brown from the University of Maryland College Park at abrown53@umd.edu. By consenting below, you indicate that you are at least 18 years of age, have read this consent information or have had it read to you, your questions have been answered to your satisfaction and you voluntarily agree to participate in this research study. Please click here in order to download a copy of this information for your records. If you do not consent, the following screen will direct you back to Prolific.

Question 1-10: Smartphone App Activity

Do you use your smartphone for the following activities? (a) Yes (b) No

{RANDOMIZE}

- Browsing websites
- Interacting with content on social media websites/apps
- Following directions from location aware apps (e.g. Google maps)
- Finding restaurants, events, and/or shops near your current location (e.g. Foursquare)
- Receiving targeted offers or coupons based on your location
- Checking the weather in your current location
- Playing games on an app that requires your location (e.g. Pokémon Go)
- Finding transportation near you (e.g. Uber)
- Tracking the number of steps you take or your level of exercise
- Connecting to other electronic devices via Bluetooth

Random Assignment – Travel Patterns, Smartphone Interaction, or Social Network Size Prompt:

****

Question 11: Travel Patterns

Will you provide your location so we can better understand {*daily travel patterns/how people interact with their smartphones/social networks*}?

Your information will be recorded as an image, similar to the one displayed here.

(a) Yes

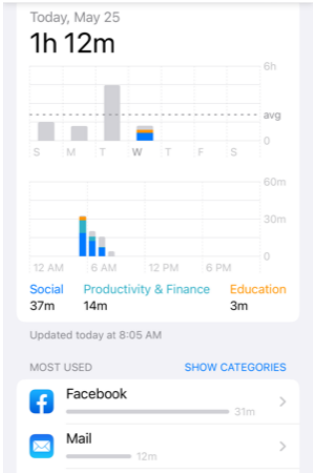
(b) No

Question 12: Travel Patterns FU if Q11=b
[DATA SECURITY or CONTROL message – shown at bottom].

Would you now be interested in sharing your location? (a) Yes (b) No

****

Question 11: Smartphone Interaction



Will you share your screen time details from today so we can better understand {*daily travel patterns/how people interact with their smartphones/social networks*}?

Your information will be recorded as an image, similar to the one displayed here.

(a) Yes

(b) No

Question 12: Smartphone Interaction FU if Q11=b
[DATA SECURITY or CONTROL message – shown at bottom].

Would you now be interested in sharing your detailed screen time information? (a) Yes (b) No

****

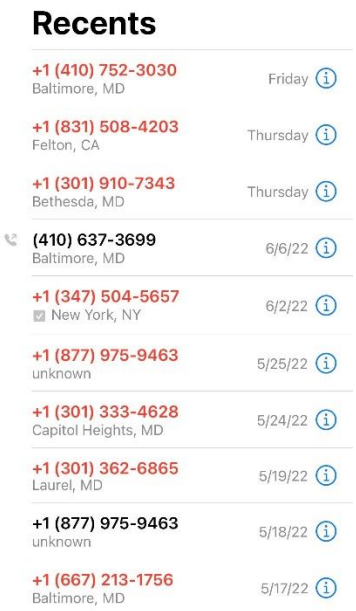
Question 11: Social Network Size Study

Will you share a list of the phone numbers saved in your recent contacts from the last 20 days so we can better understand {*daily travel patterns/how people interact with their smartphones/social networks*}?

Your information will be recorded as an image, similar to the one displayed here.

(a) Yes

(b) No



Question 12: Social Network Size Study FU if Q11=b
[DATA SECURITY or CONTROL message – shown at bottom].

Would you now be interested in sharing your detailed screen time information? (a) Yes (b) No

****

Descriptive Text if Q11=a, skip to "Debriefing on Final Page"

Question 13: Expected Data Collection
Finally, before you go, please tell me which of these data types you expect to be requested from your smartphone if you were invited to participate in a study of {*daily travel patterns / how people interact with their smartphones / social network size*}.

{SELECT ALL THAT APPLY}

(a) apps on the smartphone
(b) frequency of app usage
(c) number incoming and outgoing phone calls and text messages on the phone
(d) technical characteristics of the smartphone
(e) telephone network currently used
(f) the current location of the phone collected through the course of the study
(g) a list of websites visited
(h) activity data and periods of activity (collected from built-in accelerometer)
(i) characteristics of social networks, such as the names of contacts within your phonebook

Descriptive Text if Q11=b, "Debriefing on Final Page"
Thank you for your participation in this survey. Your participation is greatly appreciated!

Earlier you were informed that the purpose of this survey was to learn {*about daily travel patterns / how people interact with their smartphones / about social network size*}. In actuality, this study is about understanding whether a short video message can help convince individuals to share their data, as well as to understand whether the decision to share data is impacted by the topic of the study. In actuality, no additional data will be collected.

Unfortunately, in order to properly test my hypothesis, I could not provide you with that information prior to requesting the additional data. This ensured that your reaction to the invitation was spontaneous, genuine, and not influenced by prior knowledge about the purpose of the study. If I had told you the actual purpose of the study, your ability to decide whether or not to share your data could have been affected. I regret the deception of requesting that you share any additional data beyond the survey questions but hope that you understand the reason for it.

Please note that although the purpose of this study has changed from the originally stated purpose, everything else that you consented to is correct. This includes the ways in which we will keep your data secure and not personally identifiable.

Now that you know the true purpose of the study and are fully informed, you may decide that you do not want your data used in this research. If you would like your data removed from this study and permanently deleted please indicate that below. Whether you agree or do not agree to have your data used for this study, you will still receive the $1.50 from Prolific.

If you have any questions or concerns regarding this study, its purpose or procedures, or if you have a research-related problem, please feel free to contact me, Alexandra Brown from the University of Maryland College Park at abrown53@umd.edu.

*****Please keep a copy of this form for your future reference. Once again, thank you for your participation in this study!*****

a) Proceed and submit my response for analysis.
b) Withdraw my consent for this study and do not use my response.

[DATA SECURITY or CONTROL SCRIPTS]

DATA SECURITY Script – Recording:
Hello! My name is Alex Breslin and I am the researcher organizing this study at the University of Maryland. Thank you for your participation so far. I understand you don't want to share {*your location / your detailed screen time information / a list of recently contacted phone numbers*}. I know it can be unnerving to share such information with a stranger. I want to assure you that your data will be kept anonymous. I will be the only one that will ever see it and I promise to never present my findings in any way that would allow for you to be identified. I would greatly appreciate if you would reconsider participating. My email address was shared at the start of the survey and will be shared again at the end. Please reach out to me if you have any feedback or concerns.

CONTROL Script - Recording:
Hello! My name is Alex Breslin and I am the researcher at the University of Maryland organizing this study. Thank you for your participation so far. I understand you don't want to share {*your location / your detailed screen time information / a list of recently contacted phone numbers*}. I would greatly appreciate if you would reconsider participating.

Appendix C: Offer of Control Questionnaire

(Restriction: As the follow-up study will be completed on a smartphone, the participant must have access to a personal smartphone.)

General Introduction

[page 1]

Thank you for participating in this survey! Researchers from the University of Maryland are interested in learning how Americans interact with their smartphones. This should take no more than 2 minutes but you need to be on a smartphone to continue.

You will receive 2000 points through Ipsos for completing this survey and at the end, you will be invited to participate in an additional study. If you have any questions regarding this research, please reach out to Alexandra Brown at abrown53@umd.edu.Click "next page" to continue reading about this survey.

[page2]

The data collected through this survey will solely be used to research purposes and your responses will be kept anonymous. Your participation is completely voluntary and you may skip any question you do not want to answer.

This project has been reviewed by the University of Maryland Institutional Review Board (irb@umd.edu; Package Number 1955676-1). By continuing with this survey, you indicate that you are at least 18 years of age, approve of Ipsos sharing data collected through general household surveys, have read this consent information or have had it read to you, your questions have been answered to your satisfaction and you voluntarily agree to participate in this research study. Here is a link to download a copy of this information for your records.

(Section 1: Activities that suggest respondent may take pre-emptive control)

Question 1: Turning phone off

Please estimate how often you turn your personal smartphone off, even for a short period of time. (a) most days (b) a few times a week (c) a few times a month (c) less than once a month (d) never

Question 2: Keeping phone with them

Please estimate how often you keep your personal smartphone with you (i.e., on or near you) during the day. (a) it is always with me (b) it is with me most of the day (c) it is with me some of the day (d) it is never with me

(Section 2: Respondent comfort with smartphone and apps)

Question 3: Intensity of smartphone use

Do you use your smartphone for the following activities? (a) Yes (b) No

{RANDOMIZE}

- Browsing websites
- Interacting with content on social media websites/apps
- Following directions from location aware apps (e.g. Google maps)
- Finding restaurants, events, and/or shops near your current location (e.g. Foursquare)

- Receiving targeted offers or coupons based on your location
- Checking the weather in your current location
- Playing games on an app that requires your location (e.g. Pokémon Go)
- Finding transportation near you (e.g. Uber)
- Tracking the number of steps you take or your level of exercise
- Connecting to other electronic devices via Bluetooth

Question 4: Intensity of smartphone use

How often do you use a smartphone for activities other than phone calls or text messaging? (a) multiple times a day (b) every day or most days (c) a few times a week (d) a few times a month (e) less than once a month (f) never

Question 5: Comfort with installing apps

How comfortable are you with installing new apps (e.g. from iTunes, Google Play Store) on a smartphone? (a) very comfortable (b) comfortable (c) not comfortable

Question 6: Awareness of Privacy Practices within app

On a scale from 0 to 10, with 0 being "completely unsure" and 10 being "completely sure", how sure do you feel in your ability to adjust the privacy settings for the apps currently on your phone?

Such settings may be related to how the app gathers, stores, and/or uses the personally identifiable information it collects.

[0-10 scale]

(Section 3: Invitation to additional study)

Question 7: Invitation to Smartphone App Study (EXPERIMENT)

[PAGE ONE]

We would like to invite you to participate in an additional app-based study that focuses on **where Americans use their smartphones**.

To participate in the study, please download the app provided on the next page and keep it on your phone for one week. You will receive $20 for your participation, with $10 provided for downloading the app and $10 provided at the end of the week long data collection.

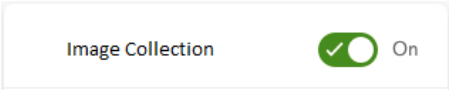
The app will record your location on a map once every hour, as displayed here. (Click "next page" to read more.)

___

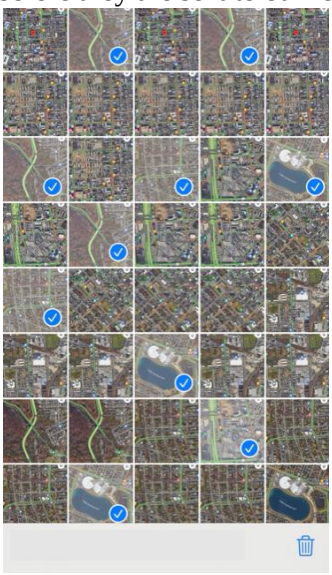
[PAGE TWO]

Important additional study information:

- Your data will remain anonymous and will never be used in a way that would personally identify you.
- *[Pre-emptive Control Offer]* The image collection can be turned off if you feel uncomfortable with sharing your location, as shown here.



- *[Reactive Control Offer]* At the end of the week, you will be provided with the images to review. You may edit the series of images before they are sent to our researchers, as shown here.



Question 8: Follow-up to Question 5

If (a) Yes, proceed to "Debriefing on Final Page"

If (b) No

We understand that you don't want to participate. Thank you for taking the time to complete the short survey.

Before you go, is there anything you would like to share with us about why you choose not to participate? [TEXT BOX]

[Proceed to "Debriefing on Final Page"]

"Debriefing on Final Page"

Thank you for your participation in this survey. Your participation is greatly appreciated!

Earlier you were informed that the purpose of this survey was to learn how Americans interact with their smartphones. In actuality, this study is about understanding how an offer of control over data that is passively collected impacts the decision to participate in a study like the one presented, in which a participant's location is tracked for one week. No such study is planned.

Unfortunately, in order to properly test my hypothesis, I could not provide you with that information prior to inviting you to participate in the location tracking study. This ensured that your reaction to the invitation was spontaneous, genuine, and not influenced by prior knowledge about the purpose of the study. If I had told you the actual purpose of the study, your ability to decide whether or not to participate could have been affected. I regret the deception of inviting you to a fictitious location tracking study but hope that you understand the reason for it.

Please note that although the purpose of this study has changed from the originally stated purpose, everything else that you consented to is correct. This includes the ways in which we will keep your data secure and not personally identifiable.

Now that you know the true purpose of the study and are fully informed, you may decide that you do not want your data used in this research. If you would like your data removed from this study and permanently deleted please indicate that below. Whether you agree or do not agree to have your data used for this study, you will still receive 2000 points for your participation in this survey.

If you have any questions or concerns regarding this study, its purpose or procedures, or if you have a research-related problem, please feel free to contact me, Alexandra Brown at abrown53@umd.edu.

*****Please keep a copy of this form for your future reference. Once again, thank you for your participation in this study!*****

a) Proceed and submit my response for analysis.
b) Withdraw my consent for this study and do not use my response.

References

Abraham, K., Maitland, A., & Bianchi, S. (2006). Nonresponse in the american time use survey: Who is missing from the data and how much does it matter? Public Opinion Quarterly, 70, 676–703. doi:10 . 1093 / poq /nfl037

Agarwal, S. Liu, C. & Souleles, N. S. (2007). The reaction of consumer spending and debt to tax rebates—evidence from consumer credit data. Journal of political Economy, 115(6), 986-1019.
Angrisani, M., Kapteyn, A., & Samek, S. (2023). Real-time Measurement of Household Financial Electronic Transactions in a Population Representative Panel: A Pilot Study. Field Methods, 35(4), 303-317. https://doi.org/10.1177/1525822X231195521

Akturan, U., & Tezcan, N. (2012). Mobile banking adoption of the youth market: Perceptions and interactions. *Marketing Intelligence & Planning*, 30(4), 444-459.

Bach, R. L., Silber, H., Gerdon, F., Keusch, F., Schonlau, M., & Schröder, J. (2024). To share or not to share–understanding individuals' willingness to share biomarkers, sensor data, and medical records. Information, Communication & Society, 1-19.

Baden, R., Bender, A., Spring, N., Bhattacharjee, B., & Starin, D. (2009, August). Persona: an online social network with user-defined privacy. *In Proceedings of the ACM SIGCOMM 2009 conference on Data communication* (pp. 135-146).

Bradburn, N. M. (2015). The affect balance scale: Subjective approaches (3). *Global handbook of quality of life: Exploration of well-being of nations and continents*, 269-279.

Büchi, M., Just, N., & Latzer, M. (2016). Caring is not enough: The importance of Internet skills for online privacy protection. *Information, Communication & Society*. Advance online publication. http://dx.doi.org/10.1080/1369118X.2016.1229001

Cceres, R. Cox, Lim, Shakimov, & Varshavsky (2009) Virtual individual servers as privacy-preserving proxies for mobile devices. In Proc. of 1st ACM SIGCOMM Workshop on Networking, Systems, & Applications on Mobile Handhelds (MobiHeld)

Chin, E., Felt, A. P., Sekar, V., & Wagner, D. (2012, July). Measuring user confidence in smartphone security and privacy. *In Proceedings of the eighth symposium on usable privacy and security* (pp. 1-16).

Cho, H. 2010. Determinants of Behavioral Responses to Online Privacy: The Effects of Concern, Risk Beliefs, Self-Efficacy, and Communication Sources on Self- Protection Strategies. *Journal of Information Privacy and Security*. 6, 1 (2010), 3±27.

Couper, M. P., & Singer, E. (2013, January). Informed consent for web paradata use. In Survey Research Methods (Vol. 7, No. 1, p. 57). NIH Public Access.

Couper, M. (2019). Mobile Data Collection: A Survey Researcher's Perspective. Keynote Presentation, MASS Workshop. Mannheim, 4th March 2019.

De Heer, W., & De Leeuw, E. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. *Survey nonresponse*, 41, 41-54.

Dillman, D. A. (2009). Some consequences of survey mode changes in longitudinal surveys. *Methodology of longitudinal surveys*, 127-140.

Dutwin, D. Loft, J. D. Darling, J. Holbrook, A. Johnson, T. Langley, R. E. ... & Triplett, T. (2014). Current Knowledge and Considerations Regarding Survey Refusals.

Elevelt, A, Lugtig, P, & Toepoel, V. (2019, April). Doing a Time Use Survey on Smartphones Only: What Factors Predict Nonresponse at Different Stages of the Survey Process?. In Survey Research Methods (Vol. 13, No. 2, pp. 195-213).

Elevelt, A., Bernasco, W., Lugtig, P., Ruiter, S., & Toepoel, V. (2021). Where you at? Using GPS locations in an electronic time use diary study to derive functional locations. Social Science Computer Review, 39(4), 509-526

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. Download PDF

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. Download PDF

Geurs, K., Veenstra, S., & Thomas, T. (2013, July). The setup of a mobile mobility panel for the Netherlands. *In Proceedings of the 13th World Conference on Transportation Research, Rio de Janeiro, Brazil.*

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). *Understanding the decision to participate in a survey*. Public opinion quarterly, 56(4), 475-495.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. John Wiley & Sons.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology*. John Wiley & Sons.

Groves, R. M. Presser, S. & Dipko, S. (2004). The role of topic interest in survey participation decisions. Public Opinion Quarterly, 68(1), 2-31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and an illustration. *The Public Opinion Quarterly*, 64(3), 299-308.

Jäckle, A., Burton, J., Couper, M. P., and Lessof, C. (2019), Participation in a mobile app survey to collect expenditure data aspart of a large-scale probability household panel: coverage and participation rates and biases. Survey Research Methods, 13,23–44

Jäckle, A., Burton, J., & Couper, M. P. (2023), Understanding Society: minimising selection biases in data collection using mobile apps. Fiscal Studies, 44, 361–376. https://doi.org/10.1111/1475-5890.12351

Kamp, M. H. (2016). *Determinants of privacy protection behavior on social network sites: the role of privacy beliefs, social norms and Internet skills* (Master's thesis, University of Twente).

Keusch, F, Struminskaya, B, Antoun, C, Couper, M. P, & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. Public opinion quarterly, 83(S1), 210-235.

Kreuter, F. Presser, S. & Tourangeau, R. (2008). Social desirability bias in cati, ivr, & web surveysthe effects of mode and question sensitivity. *Public opinion quarterly*, *72*(5), 847-865.

Kreuter, F, Haas, G. C, Keusch, F, Bähr, S, & Trappmann, M. (2019). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. Social Science Computer Review, 0894439318816389.

Laukkanen, T., Pasanen, M. Mobile banking innovators and early adopter: How they differ from other online users?. *J FinancServMark* 13,86-94 (2008). https://doi.org/10.1057/palgrave.fsm.476007

Lavrakas, P. J. (1987). *Telephone Survey Methods: Sampling, Selection, & Supervision (1st edition).* Thousand Oaks, CA: Sage Publications.

Lavrakas, P. J. (1993). *Telephone survey methods: Sampling, selection, & supervision* (2nd edition). Thousand Oaks, CA: Sage Publications.

Lavrakas, P. J. (2008). Fallback statements. In Paul J. Lavrakas (Ed.). *Encyclopedia of survey research methods.* Thousand Oaks, CA: Sage Publications.

Lewis, T, Gorsak, M, & Yount, N. (2019). An Automated Refusal Conversion Strategy for Web Surveys. Field Methods, 31(4), 309-327.

Litt, E. (2013). Understanding social network site users' privacy tool use. *Computers in Human Behavior*, 29(4), 1649-1656.

Mulder, J. & de Bruijne, M. (2019). Willingness of Online Respondents to Participate in Alternative Modes of Data Collection. Surv Pract, 12(1), 1-11.

Mun, M. Hao, S. Mishra, N. Shilton, K. Burke, J. Estrin, D. ... & Govindan, R. (2010, November). Personal data vaults: a locus of control for personal data streams. In Proceedings of the 6th International Conference (pp. 1-12).

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Olson, K. (2006). Survey Participation, Nonresponse Bias, Measurement Error Bias, & Total Bias, *Public Opinion Quarterly*, Volume 70, Issue 5, 2006, Pages 737–748, https://doi.org/10.1093/poq/nfl038

Peytcheva, E. & Groves, R. M. (2009). Using variation in response rates of demographic subgroups as evidence of nonresponse bias in survey estimates. Journal of Official Statistics, 25(2), 193.

Revilla, M, Ochoa, C, & Loewe, G. (2017). Using passive data from a meter to complement survey data in order to study online behavior. Social Science Computer Review, 35(4), 521-536.

Revilla, M, Couper, M. P, & Ochoa, C. (2019). Willingness of online panelists to perform additional tasks. methods, data, analyses, 13(2), 29.

Sala, E., & Lynn, P. (2009). The potential of a multi-mode data collection design to reduce non response bias. The case of a survey of employers. *Quality & quantity*, 43, 123-136.

Scherpenzeel, A. (2017). Mixing online panel data collection with innovative methods. In Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung (pp. 27-49). Springer VS, Wiesbaden.

Seong, S. Seo,Nasielski, Sengupta, Hangal, & et al. (2010) PrPl:A Decentralized Social Networking Infrastructure . In ACM Workshop on Mobile Cloud Computing and Services: Social Networks and Beyond

Silber, Henning, Bernd Weiss, Florian Keusch, Christoph Beuthner, and Jette Schröder. 2018. "Framing Consent Questions in Mobile Surveys: Experiments on Question Wording." Paper presented at the BigSurv18 Conference. Barcelona, Spain.

Souleles, N. S. (2007). The response of household consumption to income tax refunds. American Economic Review, 89(4), 947-958.

Stoop, I. A. L. (2005). The hunt for the last respondent: Nonresponse in sample surveys. The Hague: The Netherlands Institute for Social Research—SCP.

Stier, S. Breuer, J. Siegers, P. & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field.

Struminskaya, B. Lugtig, P. Keusch, F. & Höhne, J. K. (2020a). Augmenting surveys with data from sensors and apps: Opportunities and challenges. *Social Science Computer Review*, 0894439320979951.

Struminskaya, B., Toepoel, V., Lugtig, P., Haan, M., Luiten, A., & Schouten, B. (2020b). Understanding willingness to share smartphone-sensor data. *Public opinion quarterly*, *84*(3), 725-759.

Struminskaya, B., Lugtig, P., Toepoel, V., Schouten, B., Giesen, D., & Dolmans, R. (2021). Sharing Data Collected with Smartphone SensorsWillingness, Participation, and Nonparticipation Bias. *Public Opinion Quarterly*.

Tootoonchian, A. Saroiu, Ganjali, & Wolman (2009) Lockr: Better privacy for social networks. In CoNEXT,

Van Ingen, E., Stoop, I. A. L., & Breedveld, K. (2008). Nonresponse in the Dutch Time Use Survey: Strategies for response enhancement and bias reduction. Field Methods, 21, 69–90. doi:10.1177/1525822X08323099

Walzenbach, Sandra, Jon Burton, Mick P. Couper, Thomas F. Crossley, and Annette Jäckle. 2019. "Foot-in-the-door or door-in-the-face? A survey experiment on multiple requests for consent to data linkage." Paper presented at the European Survey Research Association (ESRA) Conference, Zagreb, Croatia, 19.07.2019.

Wenz, A, Jackle, A, & Couper, M. P. (2019, April). Willingness to use mobile technologies for data collection in a probability household panel. In Survey Research Methods (Vol. 13, No. 1, pp. 1-22). European Survey Research Association.

Douglas Williams, J Michael Brick, Trends in U.S. Face-To-Face Household Survey Nonresponse and Level of Effort, *Journal of Survey Statistics and Methodology*, Volume 6, Issue 2, June 2018, Pages 186–211, https://doi.org/10.1093/jssam/smx019

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and Internet surveys conducted with probability and non-probability samples. *Public opinion quarterly*, 75(4), 709-747.

Zhen, C. Taylor, J. L. Muth, M. K. & Leibtag, E. (2009). Understanding differences in self-reported expenditures between household scanner data and diary survey data: a comparison of Homescan and consumer expenditure survey. Review of Agricultural Economics, 31(3), 470-492.