UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Understanding Social Insider Intrusions to Personal Computing Devices

"Documento Definitivo"

Doutoramento em Informática
Especialidade de Engenharia Informática

Diogo Homem Marques

Tese orientada por:
Professor Doutor Luís Manuel Pinto da Rocha Afonso Carriço
Professor Doutor Tiago João Vieira Guerreiro

Documento especialmente elaborado para a obtenção do grau de doutor

2020

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Understanding Social Insider Intrusions to Personal Computing Devices

Doutoramento em Informática

Especialidade de Engenharia Informática

Diogo Homem Marques

Tese orientada por:

Professor Doutor Luís Manuel Pinto da Rocha Afonso Carriço

Professor Doutor Tiago João Vieira Guerreiro

Júri:

Presidente:

- Nuno Fuentecilla Maia Ferreira Neves, Professor Catedrático e Presidente do Departamento de Informática, da Faculdade de Ciências da Universidade de Lisboa

Vogais:

- Doutor Jason Hong, Full Professor, Carnegie Mellon University, USA;
- Doutor Florian Alt, Full Professor, Universität der Bundeswehr München, Alemanha;
- Doutor Duarte Nuno Jardim Nunes, Professor Catedrático, Instituto Superior Técnico da Universidade de Lisboa;
- Doutor Alysson Neves Bessani, Professor Associado, Faculdade de Ciências da Universidade de Lisboa;
- Doutor Tiago João Vieira Guerreiro, Professor Auxiliar, Faculdade de Ciências da Universidade de Lisboa, Orientador.

Documento especialmente elaborado para a obtenção do grau de doutor

Este trabalho foi apoiado pela Fundação para a Ciência e a Tecnologia (FCT) através de Bolsa Individual de Doutoramento (SFRH/BD/98527/2013), do financiamento à Unidade de Investigação LASIGE (UID/CEC/00408/2019), e do projecto mIDR (AAC 02/SAICT/2017, projecto 30347, cofinanciado pelo COMPETE/FEDER/FNR)

2020

Resumo

A adopção rápida e em massa de dispositivos computacionais móveis, e em especial de *smartphones*, acentuou as dificuldades que as pessoas têm em negociar a "privacidade num mundo em rede" (Palen and Dourish, 2003). Para muitos, o smartphone tornou-se o fulcro da interação com o mundo físico e social — e, desse modo, tornou-se também um ponto de centralização de informação pessoal. Este fenómeno alterou equilíbrios em vários domínios, de entre os quais, o da privacidade. Nas várias dimensões da privacidade, surgiram novos desafios. Por exemplo, na dimensão a que podemos chamar protectora, isto é, numa concepção de privacidade que prima pela resistência a tentativas de entidades estranhas de tirar vantagem material de dados pessoais, identificaram-se desafios como novas formas de extorsão, ou novas possibilidades de vigilância por agentes ligados a Estados, ou ainda novas formas de exploração comercial. Uma outra dimensão da privacidade, de especial importância num contexto de presença ubíqua de dispositivos computacionais na vida quotidiana, é a dimensão interpessoal. Nesta dimensão, os novos desafios são prementes e complexos. Prementes porque, para os utilizadores destes dispositivos, a possibilidade de acesso indevido por alguém próximo é um risco presente no dia-a-dia. E complexos porque, na dimensão interpessoal da privacidade, a procura de privacidade corresponde à procura, não de secretismo, mas de *agência*. Tal implica que, às partes numa relação interpessoal, não interessa apenas evitar exposição, mas também, e não de menor importância, interessa consentir exposição. A subtração da capacidade de consentir exposição, por exemplo por via do acesso indevido a um dispositivo, é assim, especialmente gravosa à privacidade interpessoal.

Neste trabalho, debruçamo-nos sobre os desafios que a massificação de dispositivos pessoais móveis trouxe à dimensão interpessoal da privacidade, sob a lente da Segurança Informática. Partimos, então, de um referencial analítico utilitário e engenheiral; isto é, pretendemos desenvolver um entendimento sobre o fenómeno que seja útil à análise dos sistemas informáticos a ele subjacentes, na medida em que possa informar decisões sobre alternativas de desenho. Procurámos dar a devida consideração a outros paradigmas, quando úteis a este objetivo. Contudo, é no paradigma da Segurança que centramos a análise, e é dele que adoptamos a unidade de observação: a *intrusão*.

Para efeitos de análise, definimos uma *intrusão por social insider* como um incidente de Segurança, em que ocorre um acesso não autorizado a um dispositivo computacional pessoal, perpetuada por um *social insider*. Por *social insider*, entendemos um actor que, mesmo sem conhecimento técnico especializado, pela mera proximidade social, consegue fisicamente aceder a dispositivos de outra pessoa. Uma intrusão ocorre quando tal actor consegue, por interfaces de utilizador comuns, ganhar acesso a conteúdos do dispositivo, sem a permissão explícita ou contra as expectativas do seu detentor legítimo.

Procurámos, através de uma sucessão de estudos de utilizador, caracterizar estes incidentes. Em Segurança Informática, procura-se comumente avaliar o risco associado à ocorrência de um incidente, em função da probabilidade de ocorrência, e da severidade dos efeitos. De modo a melhor compreender a probabilidade de ocorrência, procurámos quantificar a prevalência; isto é, a proporção de pessoas numa população que perpetuaram intrusões desde tipo num período de referência. De modo a melhor compreender a severidade, procurámos explicitar a experiência de intrusão; isto é, a sucessão de eventos, e a forma como as pessoas envolvidas foram afectadas.

Debruçámo-nos, inicialmente, sobre a estimação de prevalência de forma fiável. Seria problemático inquirir pessoas de forma directa acerca de práticas que podem ser julgadas censuráveis. Explorámos, assim, a adequação da técnica *list experiment* (e.g., McNeeley, 2012). Com esta técnica, através da conjugação do método de inquérito com o método experimental, é possível obter estimativas agregadas da prática de actos sensíveis, sem que os participantes tenham de revelar se os praticaram. Num primeiro estudo de validação desta técnica, com 90 participantes abordados pessoalmente, comparámos uma estimativa obtida através de inquérito directo, com uma estimativa obtida com a técnica de *list experiment*. Dos participantes questionados directamente, 10% indicou ter praticado uma intrusão; enquanto a estimativa obtida por *list experiment* foi de 60% o terem feito. De seguida, para validar a possibilidade de obtenção de estimativas com amostras de maior escala, conduzimos um estudo online, com 434 participantes, recrutados através do serviço Mechanical Turk. Administrámos um conjunto de variantes de uma *list experiment*, que nos permitiu comparar prevalências que conhecíamos à partida, com as prevalências estimadas pela técnica. Concluímos que, com alguns cuidados, a técnica produz estimativas fiáveis.

Tendo validado um método de quantificação, o passo seguinte foi estimar, em larga escala, prevalências de intrusão. Utilizando a plataforma Mechanical Turk, medimos dois tipos específicos de intrusões por *social insiders*: a inspecção a conteúdos de telefones móveis (a que nos referimos por "snooping"), e a utilização de dispositivo de outrem para acesso indevido à sua conta de Facebook ("facejacking"). Para ambos os casos, desenhámos instrumentos de inquérito de base empírica. Para o desenho do instrumento de medição de "snooping", recolhemos 2.226 respostas a inquéritos directos de pergunta única, sobre actos relacionados com a segurança online e em dispositivos móveis. Para o desenho do instrumento de medição de "facejacking", recolhemos 174 respostas a um questionário sobre vários comportamentos na utilização do Facebook. Para, então, aferir a prevalência de "snooping", conduzimos um estudo com a técnica de *list experiment*, com 1.381 participantes. Estimámos que 31% dos participantes teria, no período de um ano que antecedeu o estudo, inspeccionado os conteúdos do telefone móvel de outra pessoa, sem a sua permissão. Encontrámos, também, indícios de que seriam mais propensas a esta prática as pessoas mais jovens, e aquelas cujo telefone móvel era um smartphone. Levantámos, daí, a hipótese de que a propensão a esta prática poderia também estar associada a uma mais intensa utilização do dispositivo. Para testar esta hipótese, desenhámos uma variação do estudo, que administrámos a 653 novos participantes. Os resultados indicaram que tal associação existia. Por sua vez, para aferir a prevalência de "facejacking", conduzimos um estudo do mesmo tipo, com 1308 participantes. Desta feita, procurámos estimar não só prevalência de praticar este tipo de intrusão, como a prevalência de ser alvo de intrusão. Estimámos que 24% dos participantes tinham praticado este tipo de intrusão, e que 21% sabiam ter sido alvos. Quer no caso das intrusões de "snooping", quer no caso das intrusões de "facejacking", as estimativas apontam para as intrusões serem ocorrências comuns. Também em ambos os casos, os níveis de prevalência obtidos com a técnica de *list experiment* foram consideravelmente superiores aos obtidos com inquérito directo, bem como consideravelmente superiores a estimativas existentes anteriormente.

Finalmente, investigámos, também, a experiência de intrusão. Tratando-se de um fenómeno sobre o qual o conhecimento empírico existente é reduzido, optámos por uma abordagem qualitativa, utilizando um método de recolha de incidentes críticos. Através de um instrumento desenhado para o efeito, recolhemos 102 histórias pessoais, escritas por participantes anónimos, em formato de texto narrativo aberto. As narrativas descrevem situações em que os participantes

acederam a um smartphone de alguém que conhecem, ou alguém que conhecem acedeu ao seu. Da análise textual destas narrativas identificámos um conjunto de padrões comuns. Por exemplo, verificámos que, nestas narrativas, as partes envolvidas tendem a estar nos mais próximos dos círculos sociais, como sendo as relações de intimidade, familiares, ou de amizade próxima; distinguimos um conjunto de motivações para intrusão que variam na sua perniciosidade, mas com clara preponderância à intenção de controlar relações interpessoais entre o alvo de intrusão e pessoas terceiras; verificámos que é comum as intrusões acontecerem mesmo quando os dispositivos estão protegidos por chave, através de uma variedade de estratégias; ou que, na esmagadora maioria das situações, os conteúdos acedidos são registos de comunicação escrita, como emails ou mensagens de texto. De uma análise discursiva das narrativas, identificámos ainda dois aspectos transversais à experiência de intrusão. Primeiro, que as experiências de intrusão estão fortemente ligadas a uma necessidade de demonstração de vulnerabilidade entre pessoas, sendo essa demonstração uma condição necessária ao desenvolvimento da confiança mútua, vista, por sua vez, como necessária à construção de relações interpessoais. E, segundo, que as experiências de intrusão parecem ser julgadas sob um marcado viés cognitivo de atribuição; isto é, aqueles que praticam intrusões explicam o seu comportamento por circunstâncias que o rodearam, enquanto aqueles que são alvos de intrusão associam-no à existência de defeitos de carácter da outra parte. Do conjunto destes níveis de análise, evidencia-se que as experiências de intrusão por *social insiders* revestem-se de grande significância pessoal para as partes envolvidas.

Uma limitação importante da nossa análise prende-se com a nossa escolha do incidente como unidade de observação. O foco em episódios discretos no tempo é insuficiente para compreender aspectos fulcrais do fenómeno; em particular, a sua intersecção com a problemática da violência nas relações de intimidade. Em paradigmas que privilegiam a *identidade* como unidade de observação, esta forma de violência é conceptualizada como um padrão continuado de controlo coercivo. O nosso foco em incidentes não permite distinguir aquilo que são práticas distendidas no tempo, nas quais as intrusões a dispositivos são um de entre vários sinais. Existe um crescente corpo de investigação sobre a relação entre as tecnologias digitais e violência nas relações de intimidade (e.g., Burke et al., 2011; Dimond et al., 2011; Freed et al., 2018, 2017; Leitão, 2019; Matthews et al., 2017; Woodlock, 2017), em relação ao qual a nossa análise deve ser vista como subsidiária.

Em conclusão, esta caracterização empírica das intrusões por *social insiders* fornece evidências úteis à análise de risco. Na dimensão da probabilidade, as estimativas de prevalência que obtivemos indicam que o grau de probabilidade de ocorrência deste tipo de incidentes é assinalável, e subestimado até aqui. Na dimensão da severidade, identificámos padrões que permitem modelar ameaças de vários graus. As experiências de intrusão são, em geral, de significância pessoal para as partes envolvidas. Contudo, as consequências sentidas variam num espectro. Com efeito, em alguns padrões de intrusão, observámos consequências que se podem dizer positivas. Por outro lado, encontrámos também padrões que levaram a consequências catastróficas.

Abstemo-nos de julgar um nível de risco genérico associado a estes incidentes. Avaliando a probabilidade pela prevalência nos grupos etários mais baixos, e a severidade pelos padrões de intrusão geradores das piores consequências, estas intrusões estariam entre os maiores riscos de Segurança que os utilizadores encaram. Avaliando pelos extremos opostos das dimensões de risco, estes incidentes seriam trivialidades. Pretende-se, ao invés, que esta caracterização de intrusões seja um instrumento para uma modelação de ameaças mais sustentada em evidência; que assim

permita melhor análise de alternativas de desenho no domínio dos dispositivos computacionais pessoais, em complementaridade com outras dimensões de análise relevantes aos objectivos de desenho particulares ao sistema em causa.

Consideramos, contudo, fundamental que a dimensão do risco de incidente não seja descurada. O fenómeno de centralização do acesso a informações pessoais em dispositivos como smartphones, colocou-os num domínio precário, onde medeiam a construção das relações interpessoais dos seus utilizadores, mas também, potencialmente, a sua erosão. As intrusões por *social insiders* frustram o anseio pelo consentimento na exposição de informações pessoais. O consentimento, e a sua subversão, devem ocupar um papel central no desenho de tecnologias que se pretendam respeitadoras da privacidade.

Palavras-chave: Segurança, Interação Pessoa-Máquina (IPM), Computação Ubíqua, Aspectos Humanos e Societais da Segurança e Privacidade, Privacidade Interpessoal

Abstract

We examined the characteristics of social insider intrusions to personal computing devices. Social insider intrusions are situations in which one person physically accesses the device of someone they know, without permission. With devices like smartphones becoming hubs for social interaction, social insider intrusions also became a central challenge to interpersonal privacy. Through a series of quantitative and qualitative empirical studies, we sought to better understand intrusions. Our analysis indicates that the frequency of intrusions is substantially higher than previously thought, and even prevalent among younger segments of the populations we analyzed. We found recurring patterns in how intrusions unfold, including a variety of motivations and access strategies, often successful despite the presence of security technologies, like device locks. Our analysis offers both a snapshot in time, and insight onto foundational challenges that arise from technologies mediating interpersonal relationships.

Keywords: Security, Human-Computer Interaction (HCI), Ubiquitous Computing, Human and Societal Aspects of Security and Privacy, Interpersonal Privacy

Contents

List of Figures

List of Tables

1 Introduction

Personal computing devices have become intertwined with many aspects of our lives. In doing so, these devices became gateways to information which can be intimate, sensitive, or confidential. As long as others are interested in such information, there is a risk they may try to access it against our wishes. And try they do: malware, surveillance by state-sponsored actors, and personal data tracking for commercial purposes, are issues that have entered public discourse, and became, reasonably so, a point of concern (e.g., Pew Research Center, 2014).

However, in their daily lives, many end-users of computing devices have more pressing concerns than strangers accessing their data. How can users protect their privacy when people with whom they have close social ties can physically pick up their devices and browse through them? If we conceive of privacy as the ability to control the ways in which others know us, having data accessed by people whose opinions we care about is a violation of privacy in its most fundamental sense.

User concerns with unauthorized access to their devices have been previously documented. Muslukhov et al. (2013), for instance, in a quantitative comparison, found that smartphone users were equally concerned about strangers and insiders accessing their devices. Current security technologies, such as authentication locks, appear to be unable to alleviate user concerns with unauthorized access. Egelman et al. (2014), for instance, found that users perceived they could manage their concerns about strangers with authentication locks; however, they had more difficulty in managing access by people around them.

A key source of difficulty is that preventing unauthorized access is not the only important dimension to users. Users also value the ability to access their devices easily (e.g., Egelman et al., 2014; Harbach et al., 2016, 2014), and the ability to allow limited access, or signal allowance, to some people, some of the time (e.g., Hang et al., 2012; Karlson et al., 2009; Matthews et al., 2016; Mazurek et al., 2010). In fact, signaling non-allowance of access, much less enforcing it, is not a viable option for the many users who are coerced to relinquish control over their circumstances, such as those subjected to intimate partner abuse (e.g., Dimond et al., 2011; Matthews et al., 2017).

As a Computer Security issue, we can thus describe a class of adversaries who, without special skills or abilities, can obtain unauthorized physical access to personal computing devices belonging to people they know, by virtue of their social proximity. A security incident can occur when one of these adversaries obtains access to a device through its usual user interfaces, without explicit permission or against the expectations of its legitimate user. We will refer to this class of adversaries as *social insiders*, and to these security incidents as *intrusions*.

Social insider intrusions are poorly understood as a computer security issue, and thus

often overlooked. Our goal was, thus, to contribute to a better understanding of the characteristics of social insider intrusions to personal computing devices. In particular, we wanted to address the following three questions:

1. *How prevalent* are social insider intrusions to personal devices?
2. *What happens* in social insider intrusions to personal devices?
3. *How do people experience* social insider intrusions to personal devices?

In this document, we report on a series of empirical research studies aimed at addressing these questions. The document is organized as follows:

In **Chapter 2**, before reporting on our empirical work, we contextualize our research within existing knowledge. We critically examine existing empirical research on related topics, with a focus on the limitations of security technologies aimed at preventing unauthorized access, such as lock authentication.

In **Chapter 3**, we address the methodological challenge of quantifying the prevalence of social insider intrusions. In the course of our research, we intended to estimate how commonly people engaged in social insider intrusions. However, if we were to ask people directly whether they had engaged in intrusions or not, we could not reasonably expect honest responses, because such behavior is commonly seen as censurable. Here, we report on our exploration of whether *list experiments* (e.g., McNeeley, 2012), a survey technique developed to quantify sensitive behaviors, could provide less biased estimates of intrusion prevalence. In list experiments, participants are asked to indicate how many statements in a list (but not which ones) they identify with. One group of participants receives a list of control items, and another group a list of the same control items plus an item of interest. Without knowing the true answer for each respondent, an aggregate estimate of positive response to the item of interest can be calculated by the difference in mean number of items selected between groups. In Chapter 3, we provide a more detailed description of the technique, and the rationale for its selection over other techniques. We also report on two empirical studies we conducted to validate the applicability of list experiments. We first report on an in-person pilot study with 90 participants in which we compared administering a list experiment, to administering a direct question about social insider intrusions. Of participants that responded to a direct question, only 10% self-identified with having engaged in intrusions; while the list experiment estimate was that 60% of participants had engaged in intrusions. We concluded that, when quantifying intrusions, list experiments were an adequate means to reduce measurement bias. Then, we report on a study in which we measured how accurate are estimates when list experiments are administered to large online pools of participants. We ran a series of list experiments on Amazon Mechanical Turk, totaling 434 participants, in which we compared known quantities to list experiment estimates of those quantities. Furthermore, we tested adding attention check items to list questions, and segmented participants in groups with varying degrees of reputation. We concluded that online list experiments adequately estimate known quantities, and that independent of participant reputation, adding attention check items to list questions can reduce measurement bias.

Having concluded that list experiments were an adequate means to quantify intrusions, we then conducted two large-scale studies. We measured two specific instances of intrusions: a) "snooping" intrusions, that is, accessing someone else's smartphone without permission to inspect

data, and; b) "facejacking" intrusions, that is, accessing someone else's Facebook account on the victim's device.

In **Chapter 4**, we report on three empirical studies aimed at quantifying snooping intrusions. First, to select items to include in the list experiment, we conducted a series of online direct-question surveys on Google Consumer Surveys. From 2,226 responses, we selected 5 items to include in subsequent list experiments, including 4 control items, and the treatment item "I have looked through someone else's phone without their permission". Second, we ran a list experiment on Amazon Mechanical Turk with 1,381 participants, aimed at quantifying 1-year prevalence of snooping intrusions. We found that snooping intrusions were common: an estimated 31% of participants had "looked through someone else's phone without permission," in the 1-year period before the survey was conducted. We also found that being young and owning a smartphone predicted higher likelihood of having engaged in intrusions. In a third study, we ran a similar list experiment, with 653 participants, aimed at examining the relationship between engaging in snooping intrusions and using smartphones for privacy-sensitive activities. We found indication that such relationship existed: the more people used smartphones in ways which generated privacy-sensitive data, the more likely they were to snoop on others. A possible explanation for these findings is that, as people learn by their own usage what kinds of sensitive information is kept on smartphones, they gain a better sense of what they could have access to if they were to engage in an intrusion.

In **Chapter 5**, we report on two empirical studies aimed at quantifying facejacking intrusions. As in the previous chapter, in a first study, we selected items to include in the list questions by administering a direct-question survey to 174 participants. Then, we ran a list experiment on Amazon Mechanical Turk with 1,308 participants, aimed at quantifying the prevalence of facejacking intrusions. We again found that intrusions were common: we estimated that 24% of participants facejacked someone they knew, and that 21% were knowing victims of facejacking.

From these two quantification exercises, we concluded that social insider intrusions are common occurrences. Given the prevalence of at least two types of social insider intrusions, we reasoned that it would be appropriate to widen our understanding of these kinds of incidents. However, attempting to quantify every possible variation of social insider intrusions would not be practical. For instance, we quantified the prevalence of "looking through" someone else's cell phone, but an intruder might not inspect data at the moment, but instead install a surveillance implant (see e.g., Parsons et al., 2019). The quantitative approach we had employed was clearly not appropriate for a wide exploration of possible explanations, processes, and outcomes of intrusions. Finding those factors called for a more qualitative lens.

In **Chapter 6**, we report on a qualitative examination of social insider intrusions. To do so, in an online study, we collected 102 accounts of incidents of unauthorized access to smartphones. We asked participants to write about past situations in which either they accessed the smartphone of someone they know, or someone they know accessed theirs. We analyzed this data in two steps, first exploring *what happens* in such incidents, and secondly making sense of *how participants describe them.* From our first step of analysis, we unpack common aspects of social insider intrusions, including the context leading up to incidents, the course of events, and the consequences. For instance, we found that, in the accounts we collected, those who accessed devices were most commonly part of an "inner circle" of people close to the device owner; we found that there was an array of motivations for unauthorized access, ranging

from benign to malicious, but most commonly unauthorized access was motivated by a desire to learn about relationships of the device owner with third parties; we found that incidents often occurred when devices were just briefly unattended; or that, overwhelmingly, the most accessed data were records of written conversations, such as instant messages or email. In the second step of analysis, we identified two orthogonal themes in how participants experienced intrusions. First, participants understood trust as performative vulnerability: interpersonal trust was necessary to sustain relationships, but building trust required displaying vulnerability to intrusions. Second, participants were self-serving in their sensemaking: they blamed intrusions on a set of circumstances, or the other person's shortcomings, but rarely themselves.

In **Chapter 7**, we conclude by summarizing findings, reflecting on the limitations of this research, and drawing implications.

Prior Publication

This document reproduces the substance of reports of me and my collaborators' research which have been peer-reviewed and published. The following publications are wholly or partially reproduced here:

- Diogo Marques, Tiago Guerreiro, Luís Carriço, Ivan Beschastnikh, and Konstantin Beznosov. 2019. Vulnerability & Blame: Making Sense of Unauthorized Access to Smartphones. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Paper 589, 13 pages.
- Wali Ahmed Usmani, Diogo Marques, Ivan Beschastnikh, Konstantin Beznosov, Tiago Guerreiro, and Luís Carriço. 2017. Characterizing Social Insider Attacks on Facebook. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3810-3820.
- Diogo Marques, Ildar Muslukhov, Tiago Guerreiro, Konstantin Beznosov, and Luís Carriço. 2016. Snooping on Mobile Phones: Prevalence and Trends. In *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security (SOUPS '16)*. USENIX Association, Berkeley, CA, USA, 159-174.
- Diogo Marques, Tiago Guerreiro, and Luís Carriço. 2014. Measuring Snooping Behavior with Surveys: It's *How* You Ask It. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, New York, NY, USA, 2479-2484.

In this document, "we" denotes the co-authors of these publications.

2
Background

In this chapter, we contextualize our research. In Section 2.1, we frame the rising of social insider risks in connection to rapid changes in personal computing. In Section 2.2, we unpack how this phenomenon relates to conceptions of privacy. Finally, in Section 2.3 we analyze existing security defenses, and how they might collide with the rise of social insider risks.

2.1 The Rise of Social Insiders

The recognition of what we call *social insiders* as a privacy and security issue has tracked the profound changes that have occurred in personal computing. Until very recently, personal computing devices were non-existing; then they were rare; then they were common, but for work; then they were either mobile and single-purpose, or fixed and multi-purpose. Only with the advent of multi-purpose, networked, powerful, mobile devices, and their rapid massification, mainly in the form of smartphones, did it become clear that, like in many other kinds of computing systems, there was an "insider threat".

Users have been expressing privacy and security concerns ever since smartphones became widely available. For instance, a survey of 465 Deutsche Telekom smartphone customers conducted in 2009 (two years after the iPhone was introduced) found that 70% of respondents avoided certain phone functions due to security concerns (Ben-Asher et al., 2011). Smartphones combined features of other portable electronic devices, storing sensitive data that was previously distributed. Data stored on smartphones and deemed sensitive included email and browsing histories, which had previously been kept on computers; text messages and call logs, which had been kept on mobile phones; pictures and video, which had been kept on digital cameras; and location information, which had been kept on navigation devices (e.g., Ben-Asher et al., 2011; Karlson et al., 2009; Muslukhov et al., 2012).

Personal computing devices started to act as gateways to a host of online services, thus providing easy access to Internet-kept data, including archives of communication, media and online social interactions. By providing a unified point of access to such a large array of sensitive data, personal computing devices also became a unified point of failure. More so than home computers, the former locus of personal computing, mobile devices became, unsurprisingly, a source of security and privacy concerns (e.g., Chin et al., 2012).

Perceptions of the sensitivity of data kept in personal computing devices has since been documented to be nuanced. Studies have shown, for instance, that different kinds of personal data are attributed varying degrees of sensitivity, and that there is a high degree of variability

between users (e.g., Ben-Asher et al., 2011; Chin et al., 2012; Felt et al., 2012; Hang et al., 2012; Hayashi et al., 2012; Mazurek et al., 2010; Muslukhov et al., 2012, 2013). A crucial factor in the sensitivity of data kept in personal computing devices is, necessarily, *who* could access it.

2.1.1 Strangers Vs. Insiders

Two classes of adversaries who may be interested in sensitive data accessible through personal computing devices can be distinguished by having, or not having, interpersonal relationships with potential victims: *insiders* or *strangers*. These two classes are associated with different kinds of concerns.

Strangers. Strangers are agents not known to potential victims, and thus include most kinds of adversaries often considered in security literature, like malicious hackers, advertisers, or state-sponsored entities. People's concerns towards such agents revolve around avoiding *practical* losses. For instance, when people consider the consequences of having malware on their smartphones, they mostly worry with financial loss and data loss (Felt et al., 2012). Most of the data users keep and think of as private, such as pictures or private communication, is not easily exploitable by strangers who build malware. When participants in a study were prompted to imagine a kind of malware that would publish some of their private data, they were more concerned if it were disclosed to friends than to advertisers (Felt et al., 2012).

Insiders. In personal computing, insiders can be understood as people within an individual's social circle (e.g., Cherapau et al., 2015; Egelman et al., 2014; Johnson et al., 2012; Muslukhov et al., 2013). The amount of private data commonly accessible through these devices, and factors like their portability, make them more amenable to physical unauthorized access, and thus more susceptible to insider intrusions. Insiders are concerning adversaries because they can inflict *privacy* losses. (People's conceptions of privacy are reviewed in Section 2.2). When it comes to other people gaining physical access to their devices, participants in a 2013 study were as worried about insiders as they were of strangers, with nuances in what kinds of data were sensitive for each class of adversary (Muslukhov et al., 2013). Hang et al. (2012) found that when smartphone users willingly shared their devices with strangers, they feared theft, but when they shared them with insiders, they feared unwanted data exposure. Unwanted exposure of data to insiders can be particularly harmful, given the potential to damage social relationships (Johnson et al., 2012).

Muslukhov et al. (2013) proposed a descriptive model which abstracts differences between the threats posed by strangers or by insiders. According to the model, the objective of strangers is to obtain financial or other practical advantage (such as having a free device). Insiders can be distinguished in two types, per their objectives. "Extreme insiders" are malicious in the traditional sense, and want to obtain financial profit (e.g., selling the device), perpetrate some kind of fraud (e.g., identity theft), or conduct acts of sabotage (e.g., damage the device). "Conservative insiders", one the other hand, want to get access to sensitive data, use sensitive data or functionality for reasons not related to financial profit, and hide traces of unauthorized activity (Muslukhov et al., 2013).

While the kind of adversaries our research is concerned with substantially overlaps with the category of conservative insiders, we prefer to call them *social* insiders. We use the term "social" to underline our focus on individuals who have a social relationship with their victims, while still providing a clear demarcation from the most common usage of "insider" in

security literature, as individuals within organizations who develop malicious intent. For brevity, throughout this document, unless otherwise specified, mentions of *insiders* refer to social insiders.

2.1.2 Prevalence

When dealing with security threats, an important aspect to consider is the likelihood that risks materialize into actual intrusions. There is value in considering security problems that do not yet exist, but could exist one day, or problems that are small, but could get larger, or problems that are not important to most people, but are deeply important to some; but, surely, the priority should be in dealing with problems that deeply affect many people today.

Recent surveys suggest that people commonly experience unauthorized physical access to their personal computing devices. A 2012 Pew survey estimated that 12% of US mobile phone owners had at least once experienced another person accessing the contents of their phones in a way that made them feel their privacy was invaded (Pew Research Center, 2012b). In another survey, with an online sample, 14% of participants reported being victims of unauthorized access ("Someone used my mobile phone without my permission with intention to look at some of my data"), and 9% reported being perpetrators ("I used someone's mobile phone without the owner's permission to look to his/her data") (Muslukhov et al., 2013).

Unauthorized access does not only seem to be common, but also more common among population segments of concern. Younger people seem to be more likely to physically snoop on other people's devices, and to themselves be snooped on (Muslukhov et al., 2013). The so-called "digital natives" show more concern about insiders than about strangers; they are more aware of threats with a social context, like those arising from loss, theft, snooping or shoulder-surfing, than they are of threats with a technical connotation, like those arising from malware or network attacks (Kurkovsky and Syta, 2010). If it is the youngest that are more likely to engage in physical unauthorized access, then this phenomenon may grow as the current cohort ages, if nothing else is to change.

Furthermore, the true prevalence of these intrusions is likely higher than surveys suggest. One important limitation of current statistics on intrusion prevalence is their reliance on self-reports. Statistics on being a victim may be overly conservative, since intrusions can be silent; and statistics on being a perpetrator rely on self-admission to behaviors commonly deemed to be reprehensible.

2.2 Social Insiders And Privacy

We have indicated that intrusions by social insiders are concerning because, to a large degree, they inflict on people's sense of privacy. We next review prior research, unpacking how end-users understand privacy, and how such understanding intersects with the distinction between strangers and insiders.

2.2.1 Conceptions of Privacy

Privacy can be interpreted as people's ability to *control* what others can know, and cannot know, about them. Under this conception, privacy is a *process*, in the sense that people seek to exercise such control throughout time, in a changing environment, preferring to reveal, or

to withhold, more, or less, depending on the specific set of circumstances they find themselves in (e.g., Palen and Dourish, 2003).

People's views on privacy seem most consistent with this interpretation. For instance, in a recent Pew survey on attitudes about privacy, participants were prompted to think about several common conceptions of privacy and relay which they found to be important. Of the available options, "*being in control of who* can get information about you" was found to be very important by 79% of participants in a representative sample of US adults, surpassing all other conceptions, such as "not having someone watch or listen to without your permission" or "controlling what information is collected about you" (Pew Research Center, 2015a).

Palen and Dourish (2003) noted that it is particularly problematic to manage privacy in an age where data is accumulated in networked mediums. To participate in social life, we are compelled to selectively publicize some information. Simultaneously, we must contend with the fact that present privacy decisions deeply affect our future options. A conception of privacy as "static enforcement of rules" (Palen and Dourish, 2003) cannot accommodate such conflicting requirements.

2.2.2 Privacy Preferences

One implication of this conception of privacy, where moving boundaries must be continuously negotiated, is that privacy preferences are unstable. In a review of privacy behavior research, Acquisti et al. (2015) identified two strands of instability relevant to our work: *uncertainty* and *context-dependence* (a third, not discussed here, pertains to how privacy preferences are malleable and can be used to influence behavior). Acquisti et al. (2015) notes that uncertainty and context-dependence are interrelated, resulting in privacy preferences that are both highly fine-grained, and hard to adhere to.

Regarding uncertainty, Acquisti et al. (2015) noted that privacy preferences are not entirely explainable by cost-to-benefit considerations. In fact, people tend to have more stringent privacy preferences than their actual behavior would suggest, a phenomenon known as the "privacy paradox". Everyday privacy decisions may not conform to actual preferences for several reasons, including misconceptions of threats, a need to not infringe on social norms, and a need to share with others. These three factors, as we will discuss later in this chapter in the context of reviewing existing defenses, make it difficult to manage privacy boundaries with social insiders.

The other strand of instability identified in Acquisti et al. (2015) is context-dependence. Because privacy is continuously negotiated, concerns and sensitivities vary according to sets of circumstances. How each situation is perceived is informed by a wide variety of situational cues, including time, place, medium, or environment; and psychological cues, like past experiences, reciprocity in social relationships or social desirability. The context of social interactions with insiders is therefore expected to play a role on privacy sensitivity, a theme we will also return often to in our analysis of existing defenses.

Empirical research indicates that this instability in privacy preferences is particularly challenging to navigate in what relates to personal mobile devices. Users have expressed varying levels of sensitivity for different kinds of data, resources or functionality available on their devices (e.g., Ben-Asher et al., 2011; Felt et al., 2012; Karlson et al., 2009; Mazurek et al., 2010; Muslukhov et al., 2012), for different kinds of people that could have access to that data (e.g., Felt et al., 2012; Mazurek et al., 2010; Muslukhov et al., 2012, 2013), and even to individual

assets, like single pictures (Mazurek et al., 2010; Muslukhov et al., 2012).

2.2.3 Conception of Adversaries

The lagging realization of social insiders as a threat may be in part motivated by deficient understandings of what privacy means for end-users. Palen and Dourish (2003) noted that much of the discourse about privacy protection is predicated on simplistic, and static, conceptions of privacy, where preferences are stable and coherent. In such conceptions, threats to privacy originate in similarly simplistic sources. Those sources, often dubbed malicious, intend, also in stable and coherent fashion, to intrude and abuse one's privacy. Thus, concerns over surveillance or "hacking", which come from such sources, tend to be amplified, in relation to the more seemingly mundane concerns, that actually involve negotiating dynamic privacy preferences (Palen and Dourish, 2003). We note that the "malicious" sources of privacy threats tend to be associated with strangers, whereas the "mundane" tend to be associated with insiders.

Empirical work suggests people experience additional hurdles when trying to negotiate privacy boundaries with insiders through technology. Egelman et al. (2014) found, in an interview study on smartphone unlock authentication, that although participants expressed worries about strangers, they reported being able to manage such worries using unlock authentication to block them out. However, participants also expressed worries about insiders, but in that case, they had difficulties reconciling their concerns with their desire to allow some access to some of the people they knew, depending on who the person was, and in what context. These findings are mirrored in research about how users navigate Facebook's privacy settings. Johnson et al. (2012) observed that users were able to prevent strangers from accessing their content through simple binary privacy settings; however, they were less able to deal with placing access restrictions on "friends", some of whom could fit in several, fluid, and non-mutually exclusive categories, such as friends who are also co-workers. In an examination of device sharing practices within households, Mazurek et al. (2010) similarly found that access policies for insiders were difficult to configure beforehand, as there were factors other than type of relationship and asset at risk (e.g., presence, location, time of day), which influenced people's concerns. Mazurek et al. (2010) further observed that participants tended to manage insider access in ways that maximized their sense of *awareness and control*, with a complex mix of social norms and technological defenses. Instead of setting policies, participants showed a preference for being present when guest access were to happen, and to always be asked for permission. Some indicated wanting to make decisions to grant or reject access only at the last minute, and to revoke those decisions at any time. While these preferences are predictable from a conception of privacy as control, they do not conform to simplistic assumptions of maliciousness that are commonly applied to the design of security technologies and regimens.

2.3 An Analysis of Existing Defenses

We devote the remainder of this chapter to an analysis of existing defenses against physical intrusion on smartphones. Our objective was to gather a preliminary understanding of the adequateness of such defenses, especially when the actors of intrusion are social insiders.

To that end, we surveyed technological defenses which are commercially-available and those proposed in the literature, as well as non-technological coping strategies documented in

Table 2.1: Types and subtypes of surveyed defenses.

Types	**Subtypes**
Unlock Authentication	Secret-based Biometric Token-based
Risk-Aware Unlock Authentication	Asset-based Behavior-based Context-Based
Access Authorization	Explicit authorization Pre-lock access Risk-based access
Coping Strategies	

empirical studies, which we also consider to be *de facto* defenses. We classified the defenses we found into types and subtypes, listed in Table 2.1.

Since we analyzed defenses operated by end-users, our focus was on *usability*. We take a broad conception of *usability*, as encompassing all aspects of the experience of adopting technologies and practices. This stands in contrast with some security literature, where considerations of usability are conflated with "amount of work", expressed, for instance, as how long it takes for users to accomplish a task, or how frequently they make mistakes. We find such conceptions uninformative and impractical. Users may prefer tasks which take a longer time to accomplish if they are, for instance, playful; and they may not mind tasks in which they make mistakes frequently, if error recovery is easy. For instance, Android users, on average, authenticate faster with PINs, but perceive Pattern Unlock to be more usable (von Zezschwitz et al., 2013).

For user-facing security systems and regimens, usability, under this conception, is not only a concern in itself, but a determinant of security. If defenses are not usable in practice, they are not adopted, leaving users open to intrusions. Our analysis is thus centered on practical adoption, considering how the experiences of users are determined by the context in which technologies and practices are put to use. An important part of such consideration is the social context, and in particular the social constraints related to privacy management, which we outlined in the previous section, and will revisit frequently in our analysis.

We also placed special emphasis on usability constraints imposed by competing demands for attention. We note that any added efforts users have to make for the sake of security must be justified, at the risk that, given the opportunity, users will avert them. This issue has been framed within the rational choice framework of economics, where agents are assumed to make choices between several options available to them, taking into consideration the resources they can expend, and the costs and benefits associated with possible choices (e.g., Beautement et al., 2008; Herley, 2009). To follow a security regimen, individuals incur an opportunity cost, as they deprive themselves from exercising other choices, since the resources they have spent, such as effort and attention, cannot be spent elsewhere. If their relative preference is not to deprive themselves from other options, they may rationally reject security regimens which could be

advantageous. Herley (2009) noted that many security regimens burden users with continued and certain costs, which, over time, tend to be greater than the improbable and rare costs associated with occasions when they are victims of attacks *which could have been prevented by such regimens.* Beautement et al. (2008) advanced that, in organizations, under a rational choice model, failure of users in complying with certain security policies and recommendations is to be expected. Users can be seen as having a limited "compliance budget", and, when following all mandates would exceed it, they make choices to either abide or bypass, according to what they perceive to be costs and benefits of such mandates.

This section is organized as follows. Since unlock authentication has been disproportionally researched before, we divided its analysis into two subsections. Subsection 2.3.1 is devoted to standard forms of unlock authentication, which are widely available and used, such as PIN, password, and biometric unlock. Subsection 2.3.2 is devoted to more recent proposals for unlock methods that attempt to reduce the authentication burden by taking into consideration the risk of intrusion. Subsection 2.3.3 deals with technological defenses that are not based in outright locking, but instead on limiting access. Finally, Subsection 2.3.4 reviews coping strategies users employ to deal with social insiders.

2.3.1 Standard Unlock Authentication

2.3.1.1 Classification

We mirror the traditional taxonomy of authentication systems, as based on something you know, are, or have. We limit our analysis to instantiations of such approaches available for smartphone unlocking, thus classifying it as secret-based, biometric, or token-based.

Secret-based. In secret-based unlock authentication, users enter, at the beginning of each session *something they know*, and, presumably, only they know. Secrets are often sequences of characters, sometimes called passwords, passcodes, or PINs, depending on the character set allowed, and vendor preference. Notoriously, Android also offers a graphical "password", in which users draw a pattern by connecting dots. Since smartphones became widely available, secret-based unlocking has been offered in virtually every smartphone model.

Biometric. Recently, smartphones have started to provide users the ability to authenticate with *something they are*, through biometric unlock methods. Apple iPhone devices, since the 2013's 5S, are equipped with fingerprint authentication (TouchID). Some Android devices also offer fingerprint authentication, as well as facial recognition (Face Unlock), and, more recently, voice recognition (Trusted Voice). These biometric methods have fallbacks to secret-based methods. For instance, in TouchID, users must authenticate with a passcode at every reboot. Even if they never reboot, passcodes are required periodically, presumably to avoid users forgetting them. Empirical studies indicate that TouchID users do not choose better passcodes for fallback authentication (Cherapau et al., 2015). Biometric unlocking is thus subject to many of the same security limitations that apply to secret-based unlocking.

Token-based. Authentication with *something you have*, or physical tokens of identity, is also becoming increasingly available to end users. Special-purpose tokens have found a niche in multi-factor authentication systems, where they are used in conjunction with *something you know* (e.g., a password) and/or *something you are* (e.g., a fingerprint). Multi-factor authentication has traditionally been used for counteracting demanding threat models, but, more recently, under

the new Universal 2nd Factor standard, small USB and NFC tokens such as the YubiKey[1] have gained some notoriety. They can allow end-users, including smartphone users, to access their devices and logging into several online services without using other forms of authentication. Mobile operating systems have also started allowing wearable devices like smartwatches, or any other Bluetooth devices, to serve as tokens for unlocking smartphones. Android's Smartlock, for instance, provides such functionality (Google.com, 2017d).

2.3.1.2 Inconvenience

In standard secret-based authentication methods, security is achieved at the cost of convenience. Authentication secrets are as secure to guessing attacks as they are long and random. The human mind is extraordinary apt at some tasks, but far less effective in generating long and random character sequences, memorizing them, and later retrieving them accurately. Secret-generation schemes that are based on conservative models of the human mind have been proposed (e.g., Blum and Vempala, 2015), but so far there is no evidence that they can find application to device unlocking.

Part of the problem is that unlock authentication can be very frequent. Field studies have found smartphone users to initiate, on average, 46 to 48 sessions per day (Harbach et al., 2014; Mahfouz et al., 2016). Sessions not only tend to be frequent, but also short, and often with a single, well-defined, objective (Harbach et al., 2016, 2014; Mahfouz et al., 2016). For instance, Mahfouz et al. (2016) found that 85% of sessions involved 3 applications at most. The time spent authenticating with secrets is itself non-negligible, and compounds over these sessions. For instance, participants in a field study were found to, on average, over 27 days, expend 1.17 hours, of the 43.0 hours of smartphone use, unlocking their devices with either a PIN or a pattern (Harbach et al., 2014). Mahfouz et al. (2016) found that, in short sessions, users who authenticated with text passwords spent as much as 80% of their use time authenticating.

Regardless of time spent, secret-based unlocking also demands two shifts of attention. First, from an initial intent to accomplish a task, focus is shifted to overcoming authentication; then, focus again must be shifted back to the task. To a certain degree, users could, through repetition, internalize the authentication ceremonies, so as to devote less explicit attention to authenticating. Biometric authentication could also alleviate the cognitive load of attention shifting. Indeed, those who use fingerprint unlocking, perceive its ease and speed of use to be comparable to swiping to unlock (Cherapau et al., 2015; De Luca et al., 2015). However, even swiping to unlock cannot completely eliminate the demand for attention.

The usability of unlock authentication can also be impeded by situational factors. Unlock authentication has been reported to be perceived as inconvenient when users are driving (Harbach et al., 2016), when there is a time-sensitive task, like taking a picture (Harbach et al., 2016), or in emergency situations, where users themselves may not be able to authenticate (Cherapau et al., 2015; Egelman et al., 2014). Fingerprint authentication has also been regarded as inconvenient in some situations: when fingers are sweaty, or dirty, when wearing gloves, and when grasping the device while walking (Bhagavatula et al., 2015; Buschek et al., 2016).

Another possible source of inconvenience is lack of reliability. Some biometric authentication methods, such as face recognition, are often plagued with a high rate of false rejections, due both to situational factors, and to the probabilistic properties of biometric matching (Bhagavatula

[1] `https://www.yubico.com`

et al., 2015; De Luca et al., 2015).

2.3.1.3 Social Context Constraints

The all-or-nothing access model afforded by unlock authentication conflicts with common practices of device sharing. Users have been noted to want to be able to share access to their smartphones with some trusted people, or for some limited purpose and/or time (Hang et al., 2012; Hayashi et al., 2012; Karlson et al., 2009; Matthews et al., 2016). Users who want to have unlock authentication enabled, and still share their devices, face a variety of hurdles. For instance, some biometric unlock methods such as TouchID allow adding secondary users. However, configuration of biometrics is often perceived as being cumbersome (Bhagavatula et al., 2015), and many users do not even realize they have that option (Cherapau et al., 2015; De Luca et al., 2015). For impromptu sharing, the primary owner must authenticate on behalf of the other person, and eventually re-authenticate if they let the session expire. Users of secret-based authentication methods have the additional option of communicating their secrets, which they have been documented to exercise (Cherapau et al., 2015; Egelman et al., 2014; Matthews et al., 2016; Van Bruggen et al., 2013). Such practices are not only inconvenient, but also potentially expand attack vectors available to insiders.

Visible security mechanisms, like unlock authentication, also intersect with the signaling of trust between individuals. Because there are social relationships between device owners and potential adversaries, any indication of lack of openess can be fraught. For instance, in an interview study, Muslukhov et al. (2012) reported that some people did not like to lock their smartphones because of the social discomfort it caused, since it implied a distrust of people they knew. Mazurek et al. (2010), studying access control policies imposed on devices within households, found that available policies often could not conform to accepted norms of politeness and permission between cohabitants. Matthews et al. (2016) reported that, to implicitly communicate trust, participants sometimes abstained from supervising others while they are using their personal devices, despite the discomfort it caused.

2.3.1.4 Practical Security Limitations

Security mechanisms or regimes can only be qualified in relation to a specified threat. Unlock authentication is supposed to offer protection against physical unauthorized access to devices. Unpacking this threat, one first aspect to consider is the capabilities of potential adversaries. As Mickens (2014) put it, it is important if they are Mossad, or non-Mossad.

A sufficiently motivated and capable adversary may be able to surpass unlock authentication. That was made abundantly evident in recent media reports that a US policing organization, in seeking to access the contents of a suspect's smartphone, was able to overcome an unlock authentication method previously thought to be foolproof (Washington Post, 2016). Users are, however, unlikely to encounter such resourceful adversaries. Instead, a vastly more likely threat is unauthorized access by opportunistic and non-sophisticated adversaries, be them stranger or insiders.

As a defense against non-sophisticated strangers, unlock authentication can be effective, but still unattractive to end-users. When strangers obtain lost or stolen devices, owners now can, in the most popular platforms, remotely disable or wipe their devices (e.g., Chin et al., 2012; Google.com, 2017a). If data and functionality that might interest strangers, like access to

banking apps, is defended, for instance, with app-level authentication, users can rationally reject unlock authentication as a defense against unsophisticated strangers.

Unlock authentication seems to be much less effective to secure against unauthorized access by insiders. For secret-based authentication methods, or for biometric authentication methods with secret-based fallbacks, insiders are well positioned to conduct observation attacks, which do not require technical skill. One class of observation attacks that has received attention are shoulder-surfing attacks, in which adversaries visually observe the authentication ceremony. For instance, under default settings, Android's default graphic unlock password is highly guessable with a single direct visual observation (von Zezschwitz et al., 2015b). A steady stream of observation-resistant authentication methods have been proposed (e.g., De Luca et al., 2012, 2014; Marques et al., 2013; von Zezschwitz et al., 2015a). However, even observation-resistant authentication methods can be inadequate to deal with repeated observation, which insiders have opportunities to conduct. Wiese and Roth (2016), for instance, showed that for SwiPIN (von Zezschwitz et al., 2015a), an observation-resistant authentication method, the ability to replicate codes increased quickly and predictably with repeated partial observation.

One limitation in understanding the degree of security afforded by unlock authentication is the lack of empirical evidence that it actually prevents intrusions that would otherwise occur. For instance, Harbach et al. (2016), in a field study, failed to find direct evidence for screen locks having prevented intrusions. Herley (2009) noted that such lack of evidence is a common problem with security technologies and interventions, making it difficult to assess whether security regimens are cost-effective.

2.3.1.5 Adoption

Uptake of unlock authentication has been, since its inception, far from universal. Already in the pre-smartphone era, when authentication to mobile devices consisted of entering PIN codes, Clarke and Furnell (2005) found that 34% of participants in a survey did not set up authentication, and, of those who did, 45% never changed their code. Although it seems that unlock authentication has gained more adoption as smartphones became common, a large portion of users still rejects it. A 2010 study of 18- to 25-year-olds, found that in that age group, despite 80% of participants being aware of the availability of unlock authentication, only 29% chose to use it (Kurkovsky and Syta, 2010). In a 2013 online survey of 320 smartphone users, 57% indicated not locking their devices (Harbach et al., 2014); and a in larger, 2014 survey of 2,518 smartphone users, 42% indicated not locking them (Egelman et al., 2014).

Empirical studies indicate that one of the reasons for low uptake is that users indeed perceive unlock authentication to be inconvenient. Egelman et al. (2014) reported that, in a sample of 500 users who did not use secret-based unlock authentication, 34% indicated doing so because it was "too much of a hassle"; and Harbach et al. (2014) reported that, in a sample of 320 smartphone users, 47% indicated somewhat or fully agreeing that unlocking their devices could be annoying. Users who have fingerprint authentication on their devices, however, find it considerably more usable than secret-based authentication (Bhagavatula et al., 2015; Cherapau et al., 2015), and are rarely concerned with privacy issues arising from biometrics, such as unwanted identification (Bhagavatula et al., 2015; De Luca et al., 2015; Prabhakar et al., 2003).

2.3.1.6 Issues or Features?

To conclude our analysis of standard unlock authentication, we offer a reflection from a design perspective. We ask: is unlock authentication designed to fulfill what users actually want and need? It is worth to critically examine unlock authentication as a design, when so much effort is put into "righting its wrongs" (Maguire and Renaud, 2012).

Human authentication to computers is an instance of accidental design. Authenticating to computer systems came about with the Compatible Time Sharing System (Corbató et al., 1962), one of the first computers to allow multiple simultaneous users. Passwords were employed as a straightforward way to distinguish users, and thus their files (Wired, 2012). Authentication was a hack, and, as many other seminal hacks, it remains with us today.

But although it was not purposefully designed, human authentication has an implicit design, with four distinguishing features:

1. Authentication is a *ceremony*, which happens in a discrete time-span.

2. The ceremony starts by *prompting users to establish authenticity*, which can be done through a variety of mechanisms.

3. The ceremony ends with a *binary access decision*, to either grant access, or not.

4. User *authenticity is assumed to be maintained*, and access retained, as long as the session is not interrupted.

These implicit design features extended to standard unlock authentication on personal computing devices. Many of the issues we identified can be mapped onto design features:

1. Since unlocking is a ceremony, it requires expending time and shifting attention, thus being inconvenient;

2. Since users are prompted for authenticity, unlocking is susceptible to impersonation, which unsophisticated insiders can often achieve with observation attacks;

3. Since standard unlocking enforces a binary access model, it stands against user preferences for fine-grained control; and,

4. Since standard unlocking assumes authenticity is maintained throughout a session, it fails to accommodate common social practices, like impromptu sharing.

From our perspective, the limitations of standard unlock authentication are thus not issues that can be patched one by one; they are consequences of the design. It is now accepted that people are unlikely to accept new authentication methods that are more taxing than existing alternatives (e.g., Harbach et al., 2016; Herley, 2014; Mahfouz et al., 2016). Efforts to improve unlock authentication have therefore focused on reducing user effort. However, the effort required in the ceremony is only one of the design features of unlock authentication, and without addressing the others, there might not be substantial increases in uptake to be had.

2.3.2 Risk-Aware Unlock Authentication

Next, we review recent proposals to improve unlock authentication through risk-awareness. In these proposals, depending on risk, the authentication requirements, or authentication frequency, are adjusted. We found that proposals of risk-aware unlock authentication tweak features of the standard design in three ways: 1) by simply reducing the effort required for the authentication ceremony; and/or 2) by removing the need of prompting users for authentication; and/or 3) by removing the assumption that authentication is maintained throughout sessions. In all cases we found, however, when users are considered to be authenticated, access to devices is still all-or-nothing by default.

2.3.2.1 Classification

We classified proposals of risk-aware authentication by the indicators of risk they have employed. Risk has been assessed by the sensitivity of assets being accessed, or by the degree of confidence in user authentication, or by the operating context, or a combination of the aforementioned factors.

Asset-based. In asset-based approaches, authentication requirements are varied depending what functionality, applications, or data users want to access. In many proposals, users are to assess risk themselves, pre-configuring which assets are sensitive (e.g., Ben-Asher et al., 2011; Buschek et al., 2016; Muslukhov et al., 2012). For instance, a recent concept, SnapApp (Buschek et al., 2016), would have some applications accessible without authentication for a limited time window, or a "snap", provided they were not configured to be on a black list. Another notable, and commercially available, asset-based approach is to complement unlock authentication with app-level authentication. A variety of security suites offer the ability to select a set of apps which are subject to additional authentication, and some apps themselves offer locks (e.g., encrypted messaging app Signal[2]).

Behavior-based. A recent trend in proposals of improved unlock authentication is calculating risk through side-channels which track user behavior (e.g., De Luca et al., 2012; Jakobsson et al., 2009; Riva et al., 2012; Shi et al., 2010). Those approaches are probabilistic, and rely on identity detection either at the beginning of a session (sometimes referred to as *implicit authentication*), or during operation (in which case, it is often called *continuous authentication*, although technically it could be considered an intrusion detection mechanism). De Luca et al. (2012), for instance, explored asserting authentication by analyzing touch screen streams, such as touch coordinates and speed, while users unlocked their devices with simple swipe motions. Since there are individual variations on touch behavior, individuals could be granted or denied access based on this behaviormetric. An example of continuous authentication was a proposal to infer behavioral cues by continuously sensing motion, touch and proximity to other devices; and to only require explicit authentication when confidence was below a threshold (Riva et al., 2012). Variations on these approaches are a feature of several patents (e.g., Chow et al., 2012; Jakobsson et al., 2012a,b), and are increasingly available to end-users. For instance, Android's Smartlock can be set to continuously assess risk through on-body detection (Google.com, 2017d).

Context-based. Context-based approaches to unlock authentication use one or more available contextual indicators for a more holistic assessment of risk. Such indicators may be situational, like the location (e.g., Hayashi and Hong, 2015), but may also include indicators

[2] `https://whispersystems.org/`

of authenticity, or of asset value (e.g., Koved et al., 2016), subsuming the previously discussed approaches. For instance, Android's Smartlock allows for defining "trusted" locations, where explicit unlock authentication is required less often. Under Project Abacus (Business Insider, 2016), Google was reportedly working on a contextual unlock authentication mechanism that went beyond situational factors, and attempted to assert identity through behavior. IBM Research also proposed a prototypical risk-based mobile authentication system, mostly combining asset value, situational factors, and information correlated to such factors, as risk indicators, and imposing different authentication requirements depending on a combined risk score (Koved et al., 2016).

2.3.2.2 Limitations

The previously-discussed limitations of standard unlock authentication also apply to risk-aware approaches, although conceivably to a lesser degree. Even with risk-based approaches, users still must authenticate explicitly at least occasionally. Thus, risk-aware authentication by itself does little, for instance, to improve issues of memorability, or to support device sharing practices, or to reduce the social cost of signaling mistrust. We next discuss some limitations that are specific to risk-aware authentication.

A notable usability issue in risk-aware approaches is that they can remove consistency. To reduce cumulative user effort, many proposals aim at reducing authentication frequency. However, when users are required to authenticate with the same ceremony each session, there is a degree of predictability which might allow them to reduce the attention needed. When authentication is risk-based, users may face unpredictable authentication requirements at the beginning of each session (e.g., Koved et al., 2016; Riva et al., 2012), or may have to make an explicit decision as to which assets they want to access, which not only requires initial attention, but can be regretted later (e.g., Buschek et al., 2016). The degree to which users prefer authenticating less often, over having consistency, remains largely unaddressed. Ultimately, however, the choice is between the lesser of two evils.

Another factor that could affect user experience is reliability. There is intrinsic uncertainty associated with statistical approaches to determining risk (e.g., De Luca et al., 2012; Koved et al., 2016; Riva et al., 2012). Classification of risk is bound to produce false positives and false negatives, and, in some instances, to take a non-negligible amount of time. Users can become frustrated when, for instance, systems impose stricter authentication requirements than necessary, or reject access when it should have been granted, or terminate sessions prematurely. For instance, De Luca et al. (2012) found that, in their best performing behaviormetric identification variant for simple swipe gestures, legitimate users would be denied access in 6% of attempts, and attackers would be granted access in 50% of attempts. Probabilistic uncertainty can thus not only be another source of inconsistency, but also hinder both actual security, and the sense of security. A similar phenomenon has been observed with Android's Face Unlock, where high rates of false rejection have been associated with user drop-off and a decreased perception of security (Bhagavatula et al., 2015).

Another challenge with some approaches to risk-based authentication is configuration. Defining beforehand which security level should be assigned to each asset tends not to be a user-friendly task. If end-users are to make those decisions as configurations (e.g., Buschek et al., 2016; Riva et al., 2012), they could make them under misconceptions of security, which are not

uncommon (e.g., Bhagavatula et al., 2015; Cherapau et al., 2015; Egelman et al., 2014; Mazurek et al., 2010). They could also be jarred by the quantity of choices, which are hard to make given their fine-grained levels of sensitivity to assets and potential adversaries. The amount of possible combinations between factors is likely to make configuration unusable in any practical sense. Buschek et al. (2016), for instance, reported that 13 out of 18 participants in a field study rejected doing even the one-time task of defining a black list of applications that would require explicit authentication. If, on the other hand, choices of security level are not made by users, but instead pre-set, then individual variation in level of concern, which is frequently found (e.g., Hayashi et al., 2012; Muslukhov et al., 2013), can hardly be accommodated.

Context-based approaches can also overly rely on assumptions about what the risk level associated with situational indicators like location is. The common assumption is that there is less risk at home or at the office than elsewhere. For instance, Harbach et al. (2014) suggested that, in private environments, unlock authentication requests could be made less frequently, and Riva et al. (2012) suggested less authentication frequency when in proximity with other known devices. However, when considering social insiders, those locations might be exactly the ones where users crave for access control (Egelman et al., 2014; Muslukhov et al., 2013).

2.3.3 Access Authorization

The last set of technological defenses we reviewed are access authorization mechanisms. In contrast with standard or risk-aware unlock authentication approaches, access authorization mechanisms avoid the all-or-nothing access model, providing limited access to insiders or other non-owners. Such mechanisms are often motivated by the need to accommodate sharing practices.

2.3.3.1 Classification

We divided access authorization mechanisms in three classes according to the model of operation.

Explicit authorization. Commercial operating systems now offer users the option of, before sharing their devices, explicitly engaging non-locking access controls. One way to do so is with compartmented *guest accounts*. Recent Android devices, for instance, provide such functionality, in addition to full-fledged multi-user support (Google.com, 2017b), and restricted profiles, in which parental controls can be enforced (Google.com, 2017e). Both iOS and Android also offer mechanisms of *temporary limitation*, where device access is blocked to a single app, without having to change user accounts, known as "pinning" in Android (Google.com, 2017c), and "guided access" in iOS (Apple.com, 2017).

Pre-lock access. Another possible access authorization mechanism is having non-sensitive assets available prior to unlock authentication. Hayashi et al. (2012), for instance, suggested such approach, making it possible for users to configure which assets were sensitive. Hayashi et al. (2012) further suggested that more than two access levels (pre- and post-unlock) could be created, by grouping applications under several labels. Some less powerful variations of the two-layer approach are currently available to end-users. Operating systems such as Android and iOS provide small interactive components, such as audio play controllers, that work over the lock screen. Some more involved applications are also sometimes available prior to authentication, albeit with access to stored data being restricted. For instance, some many devices make the

camera application available without authentication, but users can only navigate through pictures taken in the same session.

Risk-based access. Mirroring some of the proposals for dynamic authentication, one approach for access authorization that has been explored is limiting access based on detected risk. These approaches are akin to intrusion detection mechanisms, with the response to possible intrusions being access limitation, instead of outright rejection. Karlson et al. (2009) explored the potential for such approaches by inquiring study participants on whether they would be open to automatic account switching upon detection of a non-owner operator, and 9 out of 12 participants indicated they were. TreasurePhone (Seifert et al., 2010) implemented a risk-based approach, which allowed users to configure what access to data and services was allowed under several system-detected *"spheres"*, such as "Home", "Work", or "Closed". Spheres were detectable by geo-location or proximity to other devices.

2.3.3.2 Limitations

Access authorization mechanisms mirror some of the limitations that affect the other previously discussed defenses. We again limit our discussion to particularly notable issues. We further note that the empirical basis for this discussion is limited, due to lack of sources. Despite commercial availability of some options for access authorization, there is little research illuminating on issues such as technology uptake, usability, and suitability to defend against social insiders.

One salient issue with access authorization mechanisms, and especially those of explicit authorization, is that they can stand against social expectations. Karlson et al. (2009) noted that taking explicit action to switch profiles prior to sharing a device may signal mistrust. Hayashi et al. (2012) documented that, faced with the option of having access restrictions signaled by a lock over certain apps, some participants indicated such could irritate, or entice, their guest users. Mazurek et al. (2010) warned that designs for access control mechanisms should take into consideration social conventions users already abide by, including not signaling secretiveness.

Explicitly switching profiles or accounts also requires added effort and attention. Since users are already using attention at their fullest, any additional requirements for effort and attention are likely to stand in obstacle to substantial adoption (e.g., Herley, 2014). In interviews, users have indicated being favorable to the existence of sophisticated access authorization mechanisms, but only under the condition that imposing restrictions is fast and easy (Hayashi et al., 2012; Karlson et al., 2009; Mazurek et al., 2010). Technology that conforms to such requirements seems to be elusive.

Not unlike risk-aware unlock authentication, risk-aware access authorization mechanisms also pose configuration challenges. If access levels are to be define on a per-application basis, there is no easy way to define defaults, because there's high variability in personal preferences regarding which applications should have more restricted access (e.g., Hayashi et al., 2012). It would be left to users to configure their preferences, which could itself be a substantial inconvenience, as interviewees in Hayashi et al. (2012) expressed, or could be done under misconceptions of which apps are sensitive (e.g., Egelman et al., 2014; Mazurek et al., 2010). If access level is defined beforehand on a per-guest or per- type of guest basis, the dynamic nature of privacy preferences can also be difficult to accommodate. This is especially the case for insiders, who could fit in a number of overlapping access groups, depending on the current status of relationship or other

contextual factors (e.g., Johnson et al., 2012; Mazurek et al., 2010).

Risk-based approaches to access authorization also run into the previously discussed issues of probabilistic reliability as do risk-based authentication (as noted by e.g., Karlson et al., 2009). Where it applies, reliance on trusted locations (e.g., Seifert et al., 2010) may also be problematic.

2.3.4 Coping Strategies

The last type of defense we analyzed are non-technological coping strategies. When coping strategies are widely used, it is an indication that security mechanisms do not meet user preferences or needs. For instance, when users face security mandates that stand as obstacles to their intended goals, they can cope with the inconvenience by not complying. In such cases, coping strategies are ways for users to lower the demands placed on them, often subverting the intended security goals. This observation was the genesis of the field of Usable Security (Adams and Sasse, 1999; Whitten and Tygar, 1999; Zurko and Simon, 1996), and has been repeatedly confirmed since, for instance in studies of password management practices (e.g., Chiasson et al., 2009). Another way coping strategies are used is not to lower demands, but instead to provide security, or a sense of security, that is not offered by other defenses, even if at the cost of added inconvenience.

To cope with the inconveniences of unlock authentication, users have resorted to sharing their unlock authentication secrets with others (Cherapau et al., 2015; Egelman et al., 2014; Harbach et al., 2014; Matthews et al., 2016; Van Bruggen et al., 2013), selecting codes that are easy to remember (Cherapau et al., 2015), or re-using secrets, like ATM PIN codes or bicycle lock codes for device authentication (Cherapau et al., 2015; Egelman et al., 2014). Such practices are not marginal. In a recent survey of 374 iPhone users, only 40% indicated not having shared unlock codes with anyone (Cherapau et al., 2015).

To cope with the possibility of unauthorized access to their devices, user have resorted to strategies like keeping devices in close proximity at all times (Harbach et al., 2014), keeping devices shut off or in locations considered safe when unattended (Harbach et al., 2014; Mazurek et al., 2010), keeping close supervision when other people are using devices (Hang et al., 2012; Karlson et al., 2009; Mazurek et al., 2010; Muslukhov et al., 2012), limiting physical access by others (Hang et al., 2012; Mazurek et al., 2010), obfuscating entry of authentication secrets by shielding or speed (De Luca et al., 2014; Harbach et al., 2014), delaying use until finding a physical location considered safe (De Luca et al., 2014), under-using some of the functionality provided, so as not to leave traces of potentially sensitive information (Ben-Asher et al., 2011), or actively deleting or hiding information (Mazurek et al., 2010).

3

Quantifying Social Insider Intrusions

The key challenge in measuring social insider intrusions is collecting accurate data from respondents. Asking people about incidents of intrusion runs into several issues. If we ask people whether they have engaged in intrusions, they may be reluctant to self-incriminate. If, instead, we ask people whether they were targets of intrusions, they may not be able to respond accurately, since there may have been incidents which they never learned about.

In this chapter, we report on how we addressed this challenge. In Section 3.1.1 we contextualize the issue of measurement error in survey questions which participants may deem sensitive, describe the types of survey instruments that are often deployed in such circumstances, and justify our selection of the *list experiment* technique to measure the prevalence of social insider intrusions. We then report on two empirical studies which we conducted to validate that list experiments can appropriately measure intrusion behaviors. First, as described in Section 3.2, we conducted an in-person pilot list experiment with a small sample, and compared its results to direct questioning. Second, as described in Section 3.3, to validate the appropriateness of conducting list experiments online, we ran a list experiment on Amazon Mechanical Turk, in which we measured behaviors for which we knew the true prevalence in advance.

3.1 Survey Instrument Selection

3.1.1 Sensitive Questions In Surveys

When participants give systematically untruthful responses, studies of self-reported attitudes, opinions and behaviors can run into measurement error. One classic example of measurement error can be found in studies of self-reported number of sexual intercourse partners: men consistently report having had many more partners than women. Under any reasonable set of assumptions about the populations in which these studies are conducted, such discrepancy is illogical (Wiederman, 1997).

One common source of measurement error is social desirability bias (Tourangeau and Yan, 2007). When the questions posed to participants are sensitive, they tend to give answers that are the most socially acceptable, and not necessarily the truth. Questions that pertain to protecting one's privacy are known to be subject to that bias. It has been shown that the mere addition of privacy wording in surveys makes respondents much more likely to give socially desirable responses (Braunstein et al., 2011).

To counteract the effects of social desirability bias, some indirect survey techniques have been developed. Indirect survey techniques can offer assurances of response confidentiality by

design, not policy. Even assuming an adversarial researcher, respondents have strict guarantees that their individual answers cannot be revealed, by the way the survey instrument is designed. Having that assurance, respondents are expected to be more likely to answer truthfully. The cost to researchers is that they lose the ability to know the truthful response of each individual participant, and can only rely on aggregate estimates. Two main types of indirect survey techniques have longstanding traditions as instruments to address measurement error stemming from social desirability bias: the randomized response technique, and the list experiment technique.

3.1.2 The Randomized Response Technique

With the *randomized response technique* (RRT) (Warner, 1965), in its simplest form, respondents are shown a sensitive question and asked to privately flip a coin. If it lands on one side, participants must answer "yes", regardless of truthfulness; and if it lands on the other side, they must answer truthfully, "yes" or "no". Each individual respondent is thus assured that answering "yes" does not reveal their true response, as long as no one else knows on which side the coin landed. But, knowing that the probability of a coin landing heads or tails is equal, researchers can calculate the proportion of positive responses by assuming that half the positive responses are a consequence of the coin toss, and the remaining are truthful. For a comprehensive analysis of RRT designs, see e.g., Blair et al. (2015). In the field of Human-Computer Interaction, the RRT technique has been used, for instance, to study attitudes towards differential privacy (Bullek et al., 2017).

3.1.3 The List Experiment Technique

The other technique, which we have ultimately selected, is the *list experiment*, sometimes called unmatched count technique, or item count technique, or unmatched block design (e.g., Raghavarao and Federer, 1979).

List experiments are a type of survey experiment (Mutz, 2011), which involve assigning some participants to a control group, and some to a treatment group. The groups are then administered different variants of a survey instrument. In a list experiment, individuals in the control group are shown a list of items and asked how many (not which) they identify with. Individuals in the treatment group are shown a similar list, with an extra item, referred to as a treatment item, and similarly asked to indicate how many items they identify with. If there is a difference in the mean number of items each group selects, it follows that the presence of the treatment item can explain it. Thus, the difference in mean number of items selected by each group is an estimate of the proportion of identification with the treatment item.

This technique has been shown to produce better estimates of a wide array of sensitive behaviors, including drug use, sexual practices and racial discrimination (Coutts and Jann, 2008). List experiments have also previously been used in the field of Human-Computer Interaction. For instance, Antin and Shaw (2012) used a list experiment to compare self-reported motivations to perform crowdwork in different geographies.

3.1.3.1 Discussion

It has been shown that both the list experiment and the RRT reduce response bias. In a comparative study, which tested both approaches, Rosenfeld et al. (2015) found that an RRT

survey predicted almost exactly the outcome of a vote which public opinion polls had failed to predict. A list experiment considerably reduced the bias of public opinion poll, but still underestimated the actual vote share.

However, since we intended to deploy surveys to large, online samples, application of the RRT was problematic. The RRT procedure is complex, requiring respondents to expend considerable time understanding it. Furthermore, participants in RRT surveys can have trouble believing their true answers are not revealed (Coutts and Jann, 2008). The attrition created by a high degree of complexity, and a reduced sense of anonymity, could lead to high rates of non-response and/or measurement error. List experiments, in contrast, are easier to interpret, and their guarantees of anonymity more obvious. Even if list experiments were not as effective in reducing bias, the direction of the bias would lead to conservative estimates of the kinds of behaviors we were interested in. Our judgement was that estimates erring on the side of caution would still be consequential.

3.2 A Pilot List Experiment

To validate that list experiments were an adequate method to measure the prevalence of intrusions, we first conducted a pilot study with 90 participants. Our goal was to explore whether asking for direct self-reports of intrusions would prompt social desirability bias, and whether list experiments could reduce that bias.

We thus conducted two parallel surveys: one in which a question about unauthorized access was posed directly (n = 30), and one were we administered the same question through the list experiment procedure (n = 60, divided evenly in n = 30 control and treatment groups).

Both surveys were administered in-person, so we could more closely observe how respondents engaged with the survey instruments. Participants were recruited among students of a nearby university. We next provide detail on the study design and the findings.

3.2.1 Study Design

3.2.1.1 Instrument

For the list experiment, we used the instrument depicted in Figure 3.1. The treatment item was worded as follows: “Since Jan. 1, I have used a device from someone I know without their permission to look into personal data (for instance, look through texts or call history).” The wording of the treatment item was adapted from Muslukhov et al. (2013). We intended the wording to be as descriptive and non-judgmental as possible, and thus did not use terms such as “intrusion”.

The control items were adapted from Felt et al. (2012), which includes a survey of security-relevant user behaviors. We chose the first and fourth items because they were common behaviors among mobile device users; and the second and fifth items because they were uncommon behaviors. Having a mix of common and uncommon behaviors minimizes the chances of respondents perceiving disclosure of whether they identify with the treatment item or not (see Section 4.1 for more detailed considerations on selecting items for list experiments).

For the direct-question survey, the wording was equivalent to that of the list experiment, but posed in the form of a yes or no question: *“Since Jan. 1, have you used a smartphone/tablet*

LIST EXPERIMENT Consider the following sentences relating to personal mobile device (smartphone or tablet) usage:

- Since Jan. 1, I have received on my mobile device at least one unsolicited advertising instant message (SMS or similar).
- Since Jan. 1, I have purposefully made phone calls or sent text messages to value-added numbers (excluding costumer support numbers and regular international calls/texts).
- **Since Jan. 1, I have used a device from someone I know without their permission to look into personal data (for instance, look through texts or call history).**
- Since Jan. 1, I have shared photographs taken with my mobile device with other people (for instance, with Instagram, Facebook, Twitter, SnapChat, WhatsApp, email, MMS, etc.).
- Since Jan. 1, I have lost or had a mobile device stolen from me (even if you I have recovered it later).

How many of the previous sentences apply to you?

Figure 3.1: List experiment survey instrument used in pilot study. List question as shown in the version administered to participants randomly assigned to the treatment group. Participants in the control group received the same question without the treatment item. The treatment item is marked here in bold for illustrative purposes only.

from someone you know without their knowledge, to look into personal data (for instance, look through texts or call history)?"

We prefixed all list experiment items with a 1-year temporal frame — we prompted participants to recall experiences which had happened since January 1, and collected responses in December. By indicating a temporal window of one year, we intended to focus participants on relatively recent experiences, which they could recall with a higher degree of certainty. Had we not constrained the question to a limited temporal frame, measurement error could be compounded by participants not immediately remembering experiences they had had in a more distant past.

3.2.1.2 Procedure

To gather responses, a group of three interviewers went to a public area in a neighboring university, and intercepted potential participants. Upon interception, interviewers presented themselves and asked for volunteer participation in a survey about mobile device use. Prior to administering the survey, interviewers screened participants for the following inclusion criteria: a) being between 18 and 34 years old, b) being students, and c) being regular smartphone or tablet users. When potential respondents accepted to participate and fit within the inclusion criteria, interviewers administered either the direct question or one of the list questions, and recorded responses in pen and paper. Participants were not offered any compensation.

We recruited three interviewers (2 men, 1 woman) to administer the surveys. Interviewers were from the same age group as participants, but from a different university, to avoid participants knowing the interviewers. We trained interviewers during a half-day session, immediately before conducting the survey. In the training session, topics covered included ethical treatment of participants and interview techniques, such as interception. For this study, since the questions could be sensitive, we instructed interviewers to avoid intercepting individuals walking in groups, and to gather responses out of earshot of other people. Following a matrix, each interviewer collected 30 responses, balanced between genders (15 men, 15 women) and experimental groups

Response	Control group	Treatment group
0	1	0
1	2	1
2	14	9
3	12	11
4	1	9
5	-	0
Total:	**30**	**30**

Table 3.1: Frequency of participant responses to the list experiment question administered in the pilot study.

(10 direct questions, 10 list experiment control, 10 list experiment treatment). Interviewers were compensated for their work.

3.2.2 Findings

Defining intrusions as self-identification with the behavior of *having used a device from a known person without their knowledge, to look into personal data*, our surveys provided the following estimates:

- 10% of participants (3 in 30) acknowledged having engaged in intrusions in the year before the survey was conducted, when answering a direct question.
- 60% of participants in a list experiment were estimated to have had engaged in intrusions in the year before the survey was conducted.

The list experiment estimate was calculated as follows: in the control group, the mean number of items selected was 2.33 (SD = 0.80), whereas in the treatment group it was 2.93 (SD = 0.87). The difference in means is, therefore, 0.6, or 60%. Table 3.1 shows the distribution of responses for both groups. The estimate of 10% obtained with direct questioning approximates the one previously found in the study of crowdworkers (Muslukhov et al., 2013).

3.2.3 Discussion

The results indicate that list experiments are a promising method to measure intrusions. The discrepancy in estimates when measuring the behavior through direct questioning, versus with a list experiment procedure, indicates that intrusion behaviors are indeed perceived as sensitive. Measuring intrusion behaviors directly is thus likely to induce social desirability bias, resulting in underestimation of the true prevalence. List experiments, it appears, can reduce that measurement error, and result in significantly larger estimates of prevalence.

Having an encompassing estimate of this behavior requires, however, studies at a larger scale. In the next section, we report on a larger-scale online study, in which we compared list experiment estimates to true prevalence rates known in advance.

3.3 Validating Online List Experiments

To measure the prevalence of intrusions more broadly, we considered administering list experiments online. Online administration would allow us to target larger samples than in-person

interception. To recruit participants, we elected to use the Amazon Mechanical Turk (MTurk) service[1]. MTurk is commonly used to target large participant pools for behavioral research (e.g., Paolacci and Chandler, 2014). MTurk mediates the process of engaging and compensating remote workers to perform short tasks online.

However, there are challenges in conducting survey research on MTurk, since participants may engage in satisficing (Gadiraju et al., 2015; Kapelner and Chandler, 2010; Oppenheimer et al., 2009; Peer et al., 2014). Satisficing refers to a pattern of behavior in which survey respondents select answers that require the least effort, and are still formally acceptable, regardless of their truthfulness (e.g., Müller et al., 2014, p. 244). For instance, in a multiple-choice survey question, participants may, instead of spending time considering answer options, satisfice by selecting an option at random.

There was reason to suspect that satisficing could particularly impact list experiments. List questions are cognitively more demanding than short, direct ones (Coutts and Jann, 2008), taking more time and effort to answer thoughtfully. Yet, MTurk workers have incentives to maximize compensation per time unit (Paolacci and Chandler, 2014). For studies in which groups of observational units are compared, as is the case of list experiments, there are concerns that MTurk samples, especially those with non-naive participants, may provide measurements with greater error, leading to underestimation of effect sizes (Chandler et al., 2015) and, at worst, to not finding effects when they are present (type II error).

One popular way to counteract satisficing is to use attention check questions (ACQs) (Oppenheimer et al., 2009; Peer et al., 2014). ACQs are questions whose right answer is known in advance, such as logic puzzles, trick questions, and direct instructions to answer a certain way. Although their use is well accepted and built on evidence (e.g., Peer et al., 2014), MTurk workers are now very much aware of this practice, and may therefore satisfice ACQs. It has been suggested that some workers may scan for ACQs, answer them attentively, and satisfice through the remaining questions (Hauser and Schwarz, 2016).

Another way to mitigate satisficing is restricting participation to high-reputation workers. When posting a task to MTurk, it is possible to restrict participation on a set of criteria. Two such criteria are commonly used as proxies for reputation: the total number of tasks that participants have completed in the past, and the proportion of their submitted work that was accepted by requesters. Previous research indicates that filtering participation to MTurk workers with at least 95% acceptance rate is sufficient to obtain good quality data (Peer et al., 2014). But, based on our own experience conducting studies on MTurk, and expert opinions we solicited, we came to believe that a 95% acceptance rate had become relatively easy to attain. One reason may be that requesters have grown weary of refusing work, as it might affect their own reputations, which are disseminated in platforms like Turkopticon (Irani and Silberman, 2013), and thus the ability to recruit workers for their tasks.

Since satisficing could affect the reliability of the estimates we wanted to obtain with list experiments, we aimed to understand if measurement error in MTurk list experiments could be reduced by a) restricting participation based on reputation, and b) using ACQs. We devised a between-subjects experiment in which we administered surveys to MTurk workers with distinct degrees of reputation (3 levels), and randomly assigned them to experimental groups (3 groups). One experimental group would receive a list question with only the control items, and the other

[1] `https://www.mturk.com/`

LIST EXPERIMENT Below is a list of experiences you might have had in the past 12 months. To preserve your anonymity, we ask you only to indicate HOW MANY, not WHICH ONES, apply to you.

- In the past 12 months, I've shared pictures from my cell phone on social media.
- In the past 12 months, I've purposefully made phone calls or sent text messages to 1-900 numbers.
- **[Treatment-0 ACQ] In the past 12 months, I've been to space, aboard an interplanetary vessel that I built myself.**
- **[Treatment-1 ACQ] In the past 12 months, I've opened my eyes in the morning at least once (for instance, after waking up).**
- In the past 12 months, I've received at least one text message with unsolicited advertising (spam) on my cell phone.
- In the past 12 months, I've lost or had my cell phone stolen from me.

Please count how many you have had and indicate below.

◯ 0 (None) ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 (All)

Figure 3.2: List experiment survey instrument used to estimate prevalence of 2 treatment items with known true prevalence. The list question includes 4 control items, and, to participants randomly assigned to one of the two treatment groups, one of the alternative treatment items. The treatment items are marked here in bold for illustrative purposes only. Participants in the control group received the same question without treatment items.

two would receive the control items plus one of two alternative treatment items. These treatment items were, in reality, attention checks for which we knew the true prevalence in advance. One item referred to the participant having had travelled in interplanetary space, and therefore had expected prevalence of ~0%; the other item referred to the participant having had opened their eyes in the morning, and therefore had expected prevalence of ~100%. With this design, we could compare the true prevalence to the one estimated by the difference-in-means between experimental groups.

3.3.1 Study Design

3.3.1.1 Instrument

Figure 3.2 shows the list experiment instrument we used for this study. There were three variations of the list experiment question, which were administered to three different groups of participants:

Control The control group was administered a list question with only the 4 control items.

Treatment-0 One treatment group was administered a 5-item list question, with the control items, plus the item "In the past 12 months, I've been to space, aboard an interplanetary vessel that I built myself." The expected, true prevalence of participants identifying with this item, is 0%.

Treatment-1 Another treatment group was administered a 5-item list question, with the control items, plus the item "In the past 12 months, I've opened my eyes in the morning at least once (for instance, after waking up)." The expected, true prevalence of participants identifying with this item is 100%.

	Control		Treatment-0		Treatment-1	
	$n_{Control}$	Mean	$n_{Treatment\text{-}0}$	Mean	$n_{Treatment\text{-}1}$	Mean
Low	51	1.71	54	1.61	44	2.59
Medium	46	1.13	47	1.51	42	2.43
High	57	1.46	33	1.45	60	2.50
Overall	154	1.44	134	1.54	146	2.51

Table 3.2: Number of participants, and mean number of items selected in response to list experiments administered to MTurk workers. Participants were recruited in three levels of reputation (Low, Medium, High), and randomly assigned to one of three experimental groups (Control, Treatment-0, Treatment-1).

These ACQ items were created by us, and, as far as we know, not previously used in MTurk surveys. In this way, we intended to minimize the effect of respondents detecting them without expending much effort, or using automated tools.

We selected the 4 control items for this study simultaneously with selecting items for the study reported in Section 4.2. The process of selecting the items is reported in Section 4.1.

3.3.1.2 Procedure

The study advertisement on MTurk instructed participants to follow a link to an online survey hosted on a private web server. We configured the survey to randomly assign participants to one of the experimental groups, and thus receive their respective list questions.

We posted the survey as a task on MTurk 3 times, and prevented repeated participation by the custom qualifications method (Amazon Mechanical Turk, 2017). Each time we posted it, we enforced system-level qualifications that created the following three reputation groups:

High Reputation Approval rate of 98% or higher, and at least 10,000 completed tasks.

Medium Reputation Approval rate of 95% or higher; at least 5,000, and no more than 10,000 completed tasks.

Low Reputation No minimum approval rate, and at most 5,000 completed tasks.

We targeted 150 participants per reputation group, with randomization expected to assign approximately 50 to each list experiment group. Because control items were derived from previous surveys of participants in the United States (see Section 4.1 for details), we restricted participation to workers in the US, through MTurk's own filtering feature.

3.3.2 Findings

3.3.2.1 Effect of reputation

Table 3.2 shows the mean number of items that participants selected, discriminated by levels of reputation and experimental group. We found no evidence that the mean number of selected items was different depending on reputation, when the list question was either the Treatment-0 or Treatment-1 versions (columns 5 and 7; one-way ANOVA for Treatment-0: $F(2) = 0.305$, $p = 0.737$; for Treatment-1: $F(2) = 0.292$, $p = 0.747$). Only those participants who received the Control version, which did not have attention check items, were found to

	Treatment-0 - Control ("been to space")		**Treatment-1 - Control** ("opened eyes")		**Treatment-1 - Treatment-0** ("opened eyes")	
	Estimate	SE	Estimate	SE	Estimate	SE
Low	-9 %	0.190	88 %	0.186	98 %	0.189
Medium	38 %	0.195	130 %	0.201	92 %	0.206
High	-0.2 %	0.182	104 %	0.177	105 %	0.201
Overall	10 %	0.110	107 %	0.110	97 %	0.115

Table 3.3: Prevalence estimates of treatment items by the difference-in-means between groups, and respective standard errors, from list experiments administered to MTurk workers. Participants were recruited in three levels of reputation (Low, Medium, High), and randomly assigned to one of three experimental groups (Control, Treatment-0, Treatment-1).

have answered differently according to reputation level (column 3, $F(2) = 5.053$, $p = 0.00751$). Particularly, those in the (*Medium reputation x Control version*) condition selected, on average, 1.13 items, which was the lowest among those that received either the Control version or the Treatment-0 version.

3.3.2.2 Effect of ACQs

Table 3.3 shows the estimates, by the difference-in-means, of positive answers to the "been to space" (Treatment-0) and "opened eyes in the morning" (Treatment-1) ACQ items.

The difference between the means of the Treatment-0 group and the Control group was expected to be 0 if participants were answering attentively, since they had the same number of items they could identify with. If, on the other hand, participants were choosing at random, those that received the Treatment-0 version would have selected, on average, more items, because there is one more option — a truly random response pattern in both groups would yield a difference-in-means of 0.5. The difference we actually found, not taking into account level of reputation, was 0.1, which is non-negligible, as it would mean that 10% of our sample had travelled in space. We also observed an inconsistent pattern across reputation groups, with the abnormally low mean in the (*Medium reputation x Control version*) condition inducing a difference-in-means of 0.38, thus closer to 0.5 than the expected 0.

For differences between Treatment-1 and the two possible baselines, Control and Treatment-0, the same principle applies: attentive participation should yield a difference-in-means of 1.0, and random response 0.5. Either the Control or Treatment-0 can be baselines because one item in the Treatment-0 version has true prevalence of 0%. What we found was that when the baseline was Control, the overall difference-in-means, regardless of reputation, was 1.07, and when the baseline was Treatment-0, it was 0.97. The comparison between the two groups that received attention checks, Treatment-0 and Treatment-1, was the closest to yield the expected proportion of 1.0. Furthermore, that comparison did not overestimate the true proportion, as did the comparison between Treatment-1 and Control.

3.3.3 Discussion

The attention checks we crafted seemed to have elicited enough attention from participants as to prevent degrees of satisficing that would jeopardize the validity of difference-in-means

estimates. The feedback form that we included in the task provided some anecdotal indication that, even when participants recognized some items as ACQs, they generally responded positively to their presence. For instance, participant 208 (low reputation group, Treatment-0 version) commented: *"That was a funny attention check. I wish I could have answered as having done that."*

Although we could not exclude that there were workers who engaged in satisficing, we did not uncover evidence of a pattern of misreporting that could be attributed to reputation, as measured by work history. The estimates by difference-in-means generally approached the expected 0% and 100% proportions. However, participants in the Control group, who were not exposed to ACQ-type items, appeared to be less consistent.

The differences-in-means between Treatment-1 and Treatment-0, both of which contained attention checks, were very close to the expected 100% prevalence rate, suggesting that the attention checks indeed mitigated the effect of satisficing.

The implication of these findings is that using reputation criteria to exclude participants may not be an optimal strategy to decrease satisficing. Instead, adding attention check items appears to positively affect data quality, reducing measurement error, and, in particular, overestimation. Thus, all the remaining list experiments that we conducted and report in the following chapters, include attention check items in both control and treatment groups.

3.4 Conclusion

3.4.1 Summary

In this chapter we sought to address an important challenge: how to quantify the prevalence of social insider intrusions. We elected to use a survey technique, called the list experiment, which reduces social desirability bias, while, at the same time, being easy to administer at scale to online participants.

We validated the use of list experiments to measure the prevalence of intrusions with two studies. We conducted a pilot study with 90 participants, from which we concluded that list experiments were indeed a promising approach to reduce measurement error. Then, we conducted a second study, with 434 participants, in which we validated the use of list experiments in MTurk. We concluded that list experiments in MTurk produce reliable data, as long as there are appropriate attention checks.

3.4.2 Limitations

Our quantitative measurements of intrusion prevalence must be interpreted in light of some limitations, which apply to studies reported in this and subsequent chapters.

First, although the list experiment technique reduces the effect of social desirability bias, we do not precisely know the degree to which it reduced it in our studies. Previous work (Rosenfeld et al., 2015) suggests that list experiments do not remove the effect entirely. Our estimates of prevalence of intrusions may thus be a lower bound to the actual prevalence of these behaviors.

Second, there are limits to the generalizability of our measurements, stemming from the approach we took to sampling. Our studies were conducted with convenience samples, and not with representative cross-sections of broader populations. Furthermore, as is the case of our pilot

study, our sample sizes were sometimes relatively small. Thus, for instance, the list experiment estimate of 60% prevalence that we found in our pilot study should be interpreted with caution. In that study, we sampled from a young and educated participant pool, which can be of special interest in itself when studying evolving technologies, but is certainly not representative of a broader population.

4

Prevalence of "Snooping" Intrusions

In this chapter, we quantify intrusions to mobile phones among U.S. adults. Following Muslukhov et al. (2013), we focus on "snooping" intrusions — those aimed at inspecting personal data which is accessible through mobile phones. We aimed to measure the proportion of people, in a population with a large degree of mobile device adoption, who snooped on someone else's device, and to explore the pervasiveness of the phenomenon, or lack thereof, across population groups. We targeted the U.S. adult population because it is easy to sample from for survey research, and well characterized regarding mobile device adoption. To measure prevalence, we used the list experiment procedure, which is understood to provide less biased estimates of response to sensitive questions, in comparison with direct self-reporting (see e.g. Chapter 3).

We report on three empirical studies:

In a first study, we selected the control and treatment items to include in the list experiment instrument. Our choice of items was informed by responses to direct questions, administered online through Google Consumer Surveys (GCS). To select control items, we prompted participants to self-report behaviors that relate to privacy and security. Based on 1,140 responses, we selected a mix of behaviors with varying degrees of occurrence, thus lowering the likelihood of ceiling or floor effects in the list experiment that would follow. To select the treatment item, we prompted participants to self-report on 4 alternative ways of wording the concept of a snooping intrusion. Our goal was to select a wording choice that was easy to understand, while still functioning as the operational definition of a snooping intrusion. Based on 1,086 responses, we concluded that the most adequate wording, among the alternatives, was "looked through someone else's phone without their permission". This study is reported in Section 4.1.

In a second study, we deployed a list experiment to Amazon Mechanical Turk (MTurk) to measure the prevalence of snooping intrusions (n = 1,381). We measured the proportion of people who, in 1 year, "looked through someone else's phone without their permission". Our analysis includes both a point estimate of prevalence, and predictors of such behavior. We provide estimates for the MTurk sample, which is often taken as being representative of the Internet population, and further project it into the U.S. adult population, by post-stratification weighting. We describe the final design of this study, the data collection process, and the results, in Section 4.2.

Finally, in a third study, with a similar design to the previous, we explored whether, among smartphone users, individuals' own depth of adoption of smartphones affected the likelihood of snooping on other people's mobile phones (n = 653). We examined depth of adoption because it could be the common factor between being young and having a smartphone — both characteristics

	Survey question	Yes	No	Participants
1	**In the past 12 months, have you purposefully made phone calls or sent text messages to 1-900 numbers from your cell?**	6%	94%	183
2	**In the past 12 months, have you lost or had your cell phone stolen from you?**	11%	89%	191
3	In the past 12 months, have you sent a text message to the wrong person by mistake?	17%	83%	155
4	**In the past 12 months, have you shared pictures from your cell phone on social media (for instance, Facebook or Twitter)?**	27%	73%	108
5	**In the past 12 months, have you received at least one text message/IM with unsolicited advertising (spam) on your cell?**	42%	58%	173
6	In the past 12 months, have you been asked to create a new password for an online service?	37%	63%	110
7	In the past 12 months, have you at least once cleared your cookies or browsing history?	54%	46%	113
8	In the past 12 months, have you at least once deleted / edited something you posted online?	26%	74%	107
9	In the past 12 months, have you used someone else's cell phone without their knowledge?	9%	91%	250
10	In the past 12 months, have you used someone else's cell phone without their permission?	11%	89%	335
11	In the past 12 months, have you looked through someone else's cell phone without their knowledge?	10%	90%	250
12	**In the past 12 months, have you looked through someone else's cell phone without their permission?**	15%	85%	251

Table 4.1: Questions administered through 12 single-question surveys in Google Consumer Surveys, and respective response rates and number of participants. Questions 1 to 8 were candidate control items for the list experiment question, with 1 to 5 standing for behaviors related to mobile security, and 6 to 8 behaviors related to online security. Questions 9 to 12 were several ways to phrase the treatment item. The items which ultimately appeared in the list experiment are marked in bold.

that in the previous study predicted higher likelihood of snooping on others. This hypothesis was strengthened by previous observations that smartphone users often engage in a pattern of adoption in which the phone mediates important aspects of their private social life (e.g., Chin et al., 2012; Pew Research Center, 2015b). This final study is reported in Section 4.3.

4.1 Item Selection

List experiments aim to reduce the measurement error that would occur if sensitive questions were asked directly. For them to be effective, careful consideration has to be given to the composition of the list. The perception of confidentiality can be jeopardized when lists are not credible, or when truthful answers would reveal that respondents had answered positively to the treatment item. With this first empirical study, we aimed to compose a list of items that would minimize the chances of obtaining unreliable measurements from a full-scale list experiment.

The risk of unreliable measurement may be mitigated by following common advice on designing list experiments (e.g., Blair and Imai, 2012; Coutts and Jann, 2008; Glynn, 2013; McNeeley, 2012)), which includes:

1. **Avoid ceiling effects** A ceiling effect happens when all the control items are so common that many participants would, if answering truthfully, identify with all of them, thus revealing their positive answer to the treatment item.

2. **Avoid floor effects** A floor effect occurs when the control items are so uncommon that, for many participants, the only item they could credibly report as identifying with would be the treatment item.

3. **Avoid lists that are too short** Short lists increase the likelihood of a ceiling or floor effect.

4. **Avoid lists that are too long** Long lists increase variance and demand more attention from participants.

5. **Avoid contrast effects** If the treatment item is too salient, respondents might worry that any non-zero answer to the list is indicative of identification. The list should therefore include control items that are on the same topic as the treatment item. The treatment item should itself be worded in neutral language.

Taking this advice into account, we decided to run surveys on individual behaviors to obtain prevalence estimates, so we could select a combination of control items, and a wording for the item pertaining to snooping intrusions, that would make confidentiality plausible.

4.1.1 Study Design

To build the list of items, we ran direct question surveys on several candidate items using Google Consumer Surveys (GCS). GCS lets people answer short surveys when they first land on participating websites, in exchange for access to paywalled content. For each candidate control item, we aimed at a target sample of 100 participants. For candidate treatment items, we targeted a sample of 250 participants, as we expected lower sensitivity, due to social desirability bias. The actual number of participants was often different than the target, because of the particular way in which GCS samples from its participant pool. McDonald et al. (2012) and Pew Research Center (2012a) provide additional detail on how GCS samples from the internet population, and how it compares to other participant pools.

Table 4.1, rows 1 to 8, shows the candidate control items, and respective responses to direct questions in GCS. To avoid contrast effects with the treatment item, we selected candidates among previously documented behaviors or situations related to mobile privacy (Felt et al., 2012) (rows 1 to 5) and online privacy (Pew Research Center, 2014) (rows 6 to 8).

Table 4.1, rows 9 to 12, shows the 4 options we tested to act as the treatment item pertaining to snooping intrusions. We tested four ways of wording the behavior without using the word “snooping”, which we deemed to have a too-negative connotation. Instead, we tested a maliciousness dimension, with “used” vs. “looked through” wording, and an egregiousness dimension, with “without knowledge” vs. “without permission” wording.

4.1.2 Findings

4.1.2.1 Control items

Our surveys did not find privacy-relevant behaviors or situations that can be said to be of high prevalence. Behaviors of low prevalence were, however, abundant. In part, such could be explained by the existence of social desirability bias for many of the behaviors we asked about.

Nevertheless, taking the measured prevalences for candidate items as indicative of true differences in the population, results indicated it would be trivial to avoid ceiling effects (advice 1) even with a short list, by selecting among the items with very low prevalence.

Avoiding floor effects (advice 2) was more challenging, as we did not find highly prevalent items. We decided to include 4 control items in the final list, at the cost of possible lower precision in estimates (advice 4). With 4 control items rather than 2 or 3, there were, we reasoned, enough guarantees of confidentiality. Even if respondents answered "1" it would be plausible enough that they were referring to one of the controls that is not abundantly privacy-sensitive, such as receiving spam.

We finally selected questions 1, 2, 4 and 5 as control items for the list experiment. These were the items with the highest and lowest prevalences as measured by responses to GCS surveys, among those pertaining to mobile security, and thus less contrasting with the treatment item (advice 5).

4.1.2.2 Treatment item

For the 4 candidate items conveying the "snooping intrusion" construct, the surveys we conducted did not provide evidence of appreciable differences as a result of different wording. A Chi-squared test did not provide evidence that the wording had an overall effect on the rate of positive answers ($\chi^2(3) = 5.36$, $p = 0.1471$, Cramer's V = 0.07), nor that wording conveying either egregiousness or maliciousness had significant effects in isolation ($\chi^2(1) = 2.610$, $p = 0.1062$, Cramer's V = 0.05, and $\chi^2(1) = 1.192$, $p = 0.2749$, Cramer's V = 0.04, respectively). In a logistic regression model of positive or negative answer as a function of egregiousness or maliciousness wording, we also did not find either factor to be a significant predictor at the 0.05 significance level, and the model accounted for very little of the deviance (null deviance 751 on 1085 d.f. vs. residual deviance 746 on 1083 d.f.).

We could have expanded the sample to get more precise estimates and possibly establish minute differences between wording choices, but given the observed effect sizes, and the likelihood that social desirability bias was already introducing measurement error, any differences, even if statistically significant, were unlikely to be of practical importance. We thus concluded that, for the purpose of our main survey, we should use the wording that, on its face, represented an egregious violation of an access policy with malicious intent, namely: having *looked through* someone else's cell phone without their *permission*.

4.1.3 Discussion

Based on the results of direct question surveys, we selected a list of items that included a mix of controls which were low to medium prevalence, and an item of interest that referred to a "snooping intrusion" with language as neutral as possible. We crafted this list of items into a list experiment instrument (Figure 4.1), which we used in the remaining studies of this chapter.

4.2 Prevalence Estimation

4.2.1 Study Design

4.2.1.1 Instrument

Having selected the list of items, we proceeded to design and deploy a large-scale list experiment.

We crafted a short questionnaire, with only the list question, and six other questions on personal characteristics, none of them open-ended. We only asked about age, gender, education, geographical location, and whether participants owned smartphones. Figure 4.1 shows the questions. The decision to not include more questions was made for two reasons. First, we had started with a very concise research question, and broadening the scope before that question was answered could be a waste of time. Second, with more questions, or questions that were more probing, there was a risk that participants might feel that anonymity was reduced. For instance, they could reasonably suspect that their identity could be triangulated with responses to other surveys.

To further reinforce a sense of anonymity, questions on personal characteristics were carefully chosen. We did not include, as is usual, questions about level of income or ethnicity, which participants could have felt to be very personal. For geographical location, we asked participants for US state of residency, but not city; and for level of education, we asked participants to select among broad categories.

Another survey design consideration we paid special attention to was the ordering of questions. To maximize attention and minimize incomplete responses, we chose to show the list question at the beginning of the survey. Since the question is cognitively heavy, it would be more frustrating to answer it after having quickly gone through simple demographics questions. We also inquired about personal characteristics in what we reasoned to be an increasing level of identifiability, to keep the sense of anonymity as strong as possible for as long as possible.

The list question included the control items and the treatment item selected in the study reported in the earlier section of this chapter. It also included the two attention checks used as treatment manipulations in the study reported in Section 3.3. As per the findings of that study, the main purpose of including the attention checks was not to seek out inattentive participants, but to engage participants when thinking of the answer.

4.2.1.2 Procedure

We placed the questionnaire online on a private web server, and configured it to randomly assign participants to either the treatment or the control group, each receiving the corresponding version of the list question. The survey proper was preceded by an informed consent form. We posted the survey several times as a task in MTurk, so that it would re-appear on the front page. Repeated participation was prevented by the custom qualification method (Amazon Mechanical Turk, 2017). MTurk qualifications were also used to restrict participation to residents in the United States. No other restrictions regarding past performance were enforced, as we found them to be superfluous in the study reported in Section 3.3. Participants were paid $0.20, regardless of them giving valid responses or not. The survey took 1 to 2 minutes to complete attentively.

LIST EXPERIMENT Below is a list of experiences you might have had in the past 12 months. To preserve your anonymity, we ask you only to indicate HOW MANY, not WHICH ONES, apply to you.

- In the past 12 months, I've shared pictures from my cell phone on social media.
- In the past 12 months, I've opened my eyes in the morning at least once (for instance, after waking up).
- In the past 12 months, I've purposefully made phone calls or sent text messages to 1-900 numbers.
- In the past 12 months, I've received at least one text message with unsolicited advertising (spam) on my cell phone.
- **In the past 12 months, I've looked through someone else's cell phone without their permission.**
- In the past 12 months, I've been to space, aboard and interplanetary vessel that I built myself.
- In the past 12 months, I've lost or had my cell phone stolen from me.

Please count how many you have had and indicate below.

◯ 0 (None) ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 (All)

AGE How old are you (years)? ⌊_⌋⌊_⌋⌊_⌋

GENDER What is your gender? ◯ Male ◯ Female ◯ Other

EDUCATION What is your highest level completed education?

◯ Less than High School ◯ High School
◯ Community College or Professional School (College degree)
◯ University (Bachelor's) ◯ Graduate School (Master or PhD)
◯ Other: _____

STATE In which state do you reside? ◯ Alabama ◯ Alaska ◯ Arizona ◯ Arkansas [...]

SMART1 Some cell phones are called "smartphones" because of certain features they have. Is your cell phone, if you have one, a smartphone?

◯ Yes, it is a smartphone. ◯ No, it is not a smartphone.
◯ Not sure if it is a smartphone or not. ◯ I do not have a cell phone.

SMART2 Which of the following best describes the type of cell phone you have, if you have one?

◯ iPhone ◯ Android ◯ Windows Phone ◯ Blackberry ◯ Something else
◯ I do not have a cell phone

Figure 4.1: Survey instrument used to estimate prevalence of "snooping". The first question is a list experiment question, here shown in the version administered to participants randomly assigned to the treatment group. The second and sixth items in the list question are attention checks. Participants in the control group received the same question without the treatment item. The treatment item is marked here in bold for illustrative purposes only.

4.2.2 Findings

4.2.2.1 Data Preparation

We received a total of 1,481 responses to the survey. Of those, 84 (6%) were incomplete, and were removed from the dataset. Additionally, 16 responses (1%) were eliminated for being obviously invalid: 8 for responding "none" to the list question, and 8 for responding "all". The following analysis is based on the remaining 1,381 responses.

Following Pew's approach (Pew Research Center, 2015b), we computed smartphone ownership status combining responses from two questions on ownership (SMART1 and SMART2 on Figure 4.1). Whenever the response to the question "Is your cell phone, if you have one, a smartphone?" was "Not sure", or "No, it is not a smartphone", we referred to the next question, "Which of the following best describes the type of cell phone you have", and classified participants to be smartphone users if they selected either "iPhone", "Android", "Windows Phone" or "Blackberry". There were 12 (1%) such cases.

Responses to the question about state of residency were binned into the 4 statistic regions defined by the US Census Bureau: Northeast, Midwest, South and West. For part of the analysis, ages were binned into age groups.

4.2.2.2 Participants

Table 4.2 summarizes the personal characteristics of the sample, segregated by control and treatment groups. A logistic regression of demographic characteristics as predictors, and membership to either control or treatment group as outcome, did not reveal any significant differences between groups. Applying stepwise elimination of variables, starting with a model with AIC = 1926.1 and no significant predictors, the final model marginally improved AIC to 1916.45, with the elimination of all variables. In the final model, the remaining term was not a significant predictor ($Z = 0.135$, $p = 0.893$).

Therefore, as expected from randomized assignment, there was no evidence to suggest existence of a priori differences between the control and treatment groups, which would hurt the validity of the prevalence estimates obtained through this list experiment. The demographics were similar across experimental groups, and any possible confounds could reasonably be expected to be equally distributed among them.

4.2.2.3 Attentive Participation

We investigated if there were any indications that answers were inattentive. For that we looked at the relationship between how much time it took to answer the list question, and the actual response. If participants were rushing through the question, it would be expected that they had selected one of the first options, and hence that there would be a negative correlation between the time to complete the task and the number of behaviors that participants reported as having engaged in.

The correlations for either group were close to 0 (treatment: r = -0.0015 with 95% CI -0.0760 to 0.0730; control: $r = 0.0185$ with 95% CI: -0.0563 to 0.0931), and, for both, the hypothesis of the true correlation being 0 could not be excluded (treatment: t (691) = -0.402, p = 0.968; control: t (686) = 0.484, $p = 0.6284$). We therefore concluded there was no evidence of systematic inattentive participation.

		Control (n_c = 688)	Treatment (n_t = 693)	Total (n = 1381)
By gender				
	Female	43.2 %	42.3 %	42.7 %
	Male	56.4 %	57.6 %	57 %
	Other	0.4 %	0.1 %	0.3 %
By age group				
	18-24	26 %	26 %	26 %
	25-34	46.2 %	47.3 %	46.8 %
	35-44	15.4 %	14.6 %	15 %
	45-54	6.8 %	8.5 %	7.7 %
	55-64	5.4 %	3 %	4.2 %
	65 +	0.1 %	0.6 %	0.4 %
By level of education				
	Less than high school	0.6 %	0.9 %	0.7 %
	High school	28.3 %	27.4 %	27.9 %
	Other college degree	18.8 %	19.9 %	19.3 %
	Bachelor's degree	41.4 %	39 %	40.2 %
	Masters or PhD	9.6 %	11.4 %	10.5 %
	Other	1.3 %	1.4 %	1.4 %
By region				
	Midwest	23 %	21.1 %	22 %
	Northeast	19.5 %	21.2 %	20.3 %
	South	35.2 %	33.8 %	34.5 %
	West	22.4 %	24 %	23.2 %
By ownership status				
	Does not own smartphone	12.4 %	10.1 %	11.2 %
	Owns smartphone	87.6 %	89.9 %	88.8 %

Table 4.2: Summary of participant demographics, overall and by experimental group, in the survey administered to estimate prevalence of "snooping" (n = 1,381).

Response	Control group	Treatment group
0	88 (12.8%)	76 (11%)
1	258 (37.5%)	204 (29.4%)
2	249 (36.2%)	239 (34.5%)
3	84 (12.2%)	122 (17.6%)
4	9 (1.3%)	43 (6.2%)
5	-	9 (1.3%)
Total:	**688 (100%)**	**693 (100%)**

Table 4.3: Frequency of participant responses to the list experiment question used to estimate prevalence of "snooping" (n = 1,381). Responses adjusted for 4 control items and 1 treatment item in the treatment group.

4.2.2.4 Instrument Design Effects

We inspected the data for possible ceiling and floor effects. Table 4.3 shows the raw distribution of responses to the list experiment question for both groups. The vast majority of participants selected an answer between 1 and 3 (85.9% in the control group, 81.5% in the treatment group). Thus, the presence of appreciable ceiling or floor effects was unlikely.

We then investigated the possibility that the treatment item changed how participants in the treatment group identified with the control items. For instance, participants could be more willing to identify with having called a 1-900 number because it appeared to be less censurable when compared to snooping. Blair and Imai (2012) describes a statistical procedure to check for such an effect. Following that procedure, we found no evidence of a design effect.

Taking all this evidence together, we concluded that the design of the study and its deployment yielded a sound dataset.

4.2.2.5 Prevalence Estimate

We defined (1-year) prevalence as the proportion of people in the population who internally identified as having had looked through someone else's cell phone without their permission. Prevalence was estimated by the difference-in-means between groups in a list experiment.

Table 4.4 summarizes the estimated 1-year prevalence for the sample and further breaks it down by segments of personal characteristics. For the overall sample (line 1), the 1-year estimate of prevalence was 31%.

Our sample was not, however, a fair reflection of the U.S. population. Participants, on average, were younger, attained a higher level of education, and predominately identified as being male, which is expected in MTurk convenience samples (Buhrmester et al., 2011). We adjusted the data to the U.S. population estimates from the 2010 Census, and obtained an estimate of 20% for the U.S. adult population (see Table 4.5).

The data was adjusted with cell-based post-stratification weighting. We created weights for strata which, from the sample subset summaries, we found to have appreciably different prevalence estimates between levels. Using every possible demographics criteria to stratify would create cells with two few observations. Even the combination of gender, age group and region yielded marginal frequencies of 0. Moreover, using demographic criteria for which there was little divergence between strata would have very limited impact on the overall prevalence estimate. We therefore decided to use weights based on the cross-tabulation of only age group and gender.

	Control group mean (SE)	Treatment group mean (SE)	Prevalence (SE)	*P*-value
Overall	**2.517 (0.035)**	**2.825 (0.042)**	**30.8 % (0.055)**	**<0.00001**
By gender				
Male	2.500 (0.046)	2.759 (0.057)	25.9 % (0.073)	**0.00043**
Female	2.542 (0.053)	2.918 (0.063)	37.6 % (0.083)	**0.00001**
By age group				
18-24	2.631 (0.067)	3.156 (0.086)	52.4 % (0.109)	**<0.00001**
25-34	2.522 (0.051)	2.820 (0.062)	29.8 % (0.080)	**0.00023**
35-44	2.509 (0.089)	2.644 (0.096)	13.4 % (0.131)	0.30730
45-54	2.362 (0.116)	2.407 (0.124)	4.5 % (0.169)	0.79038
55+	2.158 (0.158)	2.240 (0.202)	8.2 % (0.257)	0.75036
By level of education				
High school	2.482 (0.061)	2.789 (0.087)	30.7 % (0.106)	**0.00396**
Other college degree	2.667 (0.085)	2.949 (0.096)	28.3 % (0.129)	0.02889
Bachelor's degree	2.526 (0.054)	2.826 (0.067)	30.0 % (0.086)	**0.00053**
Masters or PhD	2.318 (0.110)	2.633 (0.105)	31.5 % (0.153)	0.04102
By region				
Midwest	2.494 (0.071)	2.699 (0.092)	20.5 % (0.117)	0.07989
Northeast	2.515 (0.078)	2.776 (0.093)	26.1 % (0.122)	0.03290
South	2.566 (0.060)	2.915 (0.072)	34.8 % (0.094)	**0.00024**
West	2.468 (0.073)	2.855 (0.086)	38.8 % (0.113)	**0.00067**
By ownership status				
Does not own smartphone	1.800 (0.093)	1.914 (0.093)	11.4 % (0.131)	0.38513
Owns smartphone	2.619 (0.036)	2.928 (0.044)	30.9 % (0.057)	**<0.00001**

Table 4.4: Estimated 1-year prevalence of engaging in "snooping" intrusions, calculated by the difference in means between experimental groups in the list experiment (n = 1,381). The table shows estimates for overall sample and for subsets based on personal characteristics. We do not provide estimates for subsets in which there were less than 20 observations in either experimental group, except for the age 65+ subset, which we binned with the 54-65 subset into the 55+ level. *P*-values from a t-test with the null hypothesis that there was no difference between experimental groups, with alpha set at 0.05. Bonferroni-adjusted significant differences in bold. Standard error of measures shown in parentheses.

	Control group	Treatment group	Prevalence	*P*-value
Adjusted mean	2.41	2.61	20%	**0.01515**
SE	0.055	0.061	0.081	

Table 4.5: Estimated 1-year prevalence of U.S. adults engaging in "snooping" intrusions, calculated by the difference in means between experimental groups in the list experiment (n = 1,381), adjusted by cell-based post-stratification weighting to the 2010 Census by age and gender. *P*-value from a design-based t-test (Lumley, 2004) of the difference in means. Table 4.6 shows the weights that were applied.

Gender	Age group	Proportion of US population	Proportion of respondents	Weight
Female	18-24	6.4%	10.4%	0.6162
Female	25-34	8.7%	19.0%	0.4596
Female	35-44	8.8%	6.5%	1.3459
Female	45+	27.6%	7.0%	3.9534
Male	18-24	6.7%	15.6%	0.4276
Male	25-34	8.8%	27.7%	0.3171
Male	35-44	8.7%	8.5%	1.0254
Male	45+	24.3%	5.3%	4.5923

Table 4.6: Weights applied to adjust the sample of participants who responded to the survey used to estimate prevalence of "snooping" (n = 1,381) to the U.S. adult population. Weights reflect the differences between subsets of the sample and corresponding subsets of the U.S. adult population, as measured by the 2010 Census. The sample was younger and had a greater proportion of males than the general population.

Predictor variables	R^2	ΔR^2	*F*	D.f.	*P*-value
L.E. group	0.022				
L.E. group + gender	0.025	0.003	1.87	2	0.1542
L.E. group + age	0.053	0.031	44.78	1	**<0.0001**
L.E. group + level of education	0.031	0.009	2.47	5	**0.0306**
L.E. group + region	0.025	0.003	1.32	3	0.2671
L.E. group + ownership status	0.100	0.077	118.38	1	**<0.0001**

Table 4.7: Summary of generalized linear regression models of number of items selected in the list experiment question in the survey administered to estimate prevalence of "snooping" (n = 1,381), as a function of demographic variables, controlling for experimental group membership. The first row shows a reduced model, in which responses to the list experiment question are modelled only as a function of experimental group membership, and indicates the proportion of variance explained by the model (R^2). The remaining rows show models in which demographic variables are added to the reduced model, and indicates the difference in explained variance (ΔR^2), and the F-statistic from an ANOVA of the reduced and larger models, with corresponding degrees of freedom and P-value. Table 4.8 shows the coefficients for each model.

At that granularity, the number of observations for some *age x gender* subsets was still too low to obtain reasonable weights. Re-leveling the 3 older age groups into one *45+* group, we were able to obtain more adequate weights, shown in Table 4.6. As with any adjustment of this type, we obtained a more representative estimate, at the cost of increasing standard error (SE). The national population statistics and diagnostics are shown in Table 4.5, and were computed with the R "survey" package, which implements the weighted analysis instruments described in Lumley (2004).

4.2.2.6 Predictors

Although the overall 1-year estimates were informative by themselves, having a large sample allowed us to look at differences between sample subsets, which could help explain snooping behaviors. Table 4.4 suggests that in all demographic criteria, except for level of education, the estimates of prevalence were considerably different between subsets, but more detailed analysis was required to discern if demographic criteria could *predict* lower or higher prevalence. It would have been, however, impractical and uninformative to try to understand the underlying demographics of snooping behavior based on all possible criteria. We therefore

Model / Variables	Estimate	SE	*t*	*p*
List Experiment group				
(Intercept)	2.51744	0.03885	64.806	<0.00001
Treatment group	0.30795	0.05484	5.616	**<0.00001**
List Experiment group + gender				
(Intercept)	2.57587	0.04999	51.531	<0.00001
Treatment group	0.30797	0.05483	5.617	**<0.00001**
Male	-0.10050	0.05545	-1.812	0.0702
Other gender id.	-0.40287	0.51104	-0.788	0.4306
List Experiment group + age				
(Intercept)	3.08289	0.09275	33.24	<0.00001
Treatment group	0.30305	0.05399	5.613	**<0.00001**
Age	-0.01784	0.00267	-6.692	**<0.00001**
List Experiment group + level of education				
(Intercept)	2.32081	0.08951	25.929	<0.00001
Treatment group	0.30991	0.05474	5.662	**<0.00001**
Bachelor's degree	0.20050	0.09483	2.114	0.0347
Other college degree	0.33175	0.10484	3.164	0.0016
High school	0.16002	0.09906	1.615	0.1065
Less than high school	-0.00675	0.33226	-0.02	0.9838
Other level of education	0.46345	0.24794	1.869	0.0618
List Experiment group + region				
(Intercept)	2.44412	0.06408	38.14	<0.00001
Treatment group	0.30814	0.05485	5.618	**<0.00001**
Northeast	0.08432	0.545	0.586	0.5580
South	0.1418	0.07478	1.896	0.0582
West	0.06479	0.0816	0.794	0.4274
List Experiment group + ownership status				
(Intercept)	1.72178	0.08209	20.975	<0.00001
Treatment group	0.2875	0.05268	5.458	**<0.00001**
Owns smartphone	0.90782	0.08344	10.88	**<0.00001**

Table 4.8: Coefficients of generalized linear regression models of number of items selected in the list experiment question in the survey administered to estimate prevalence of "snooping" (n = 1,381), as a function of demographic variables, controlling for experimental group membership. The first model has a single predictor: assignment to either treatment or control group. The remaining models add each of the other variables (gender, age, level of education, region, and smartphone ownership), controlling for assignment to control or treatment group. All demographic variables, except for age, modelled as categorical. Table 4.7 shows differences between models.

sought to find the demographic variables that better explained the list experiment outcomes, and only then to model the prevalence according to those variables.

To find relationships between demographic criteria and likelihood of snooping, we first constructed linear regression models of the number of items participants selected as a function of each available variable (gender, age, level of education, region, and smartphone ownership), controlling for assignment to control or treatment group. Table 4.7 summarizes those models with the R-squared and F statistics, and shows comparisons to a smaller model in which group assignment is the only predictor. Coefficients of each model are reproduced in Table 4.8.

Regarding gender, for respondents who identified as being female, the prevalence estimate in the sample was 38%, whereas for those who identified as male, it was 26% — a difference of more than 10 percentage points (Table 4.4, lines 2 and 3). However, the model with both gender and experimental group variables as predictors, indicated that the gender variable explained very little of the variance in either group. This model did not significantly improve on the smaller model, with just the experimental group as predictor, explaining only an additional 0.003 of the variance (Table 4.7, line 2). Gender, therefore, did not seem to have a strong enough relationship with snooping behavior to justify including it in a model with other predictors.

Age (modelled as continuous variable, not by age group), on the contrary, significantly contributed to selecting more items. Each additional 10 years of age predicted selecting, on average, less 0.18 items ($p < 0.0001$), in addition to the effect of group membership. Age, was therefore, considered a good candidate variable for a larger model.

The results of the model of level of education were mixed. Level of education can be thought of as an ordered variable, raising the question of whether more education could predict selecting a greater or lower number of items. Looking into the estimates of that regression, we found no clear evidence. Taking post-graduate education as a baseline, the model indicated that those with a college or Bachelor's degree selected a higher number of items (+ 0.33 with $p = 0.0016$, and + 0.20 with $p = 0.0347$, respectively), but there was no evidence of an effect for other levels of education. We expected to find that greater predicted difference in number of selected items would be associated with the greater differences in level of education, but that was not the case. Without an interpretation for that pattern, we concluded that this variable was not a good candidate for a larger model, despite the fact that adding it modestly improved the smaller model.

Region, like gender, did not seem to have a relationship with prevalence, on the basis that the model including it as a predictor did not significantly improve on the smaller model. We found it, therefore, to not be a good candidate.

Finally, regarding ownership status, the model suggested that those who owned smartphones selected more items from the list, even when controlling for membership in either control or treatment group. Adding ownership status to a model of only group membership explained 7.7% more of the variance, the greatest difference we found. Looking at the estimates of the model, we found the additional effect of owning a smartphone to be selecting 0.91 more items ($p < 0.0001$). Thus, ownership was judged a promising candidate variable for a larger model.

4.2.2.7 A Model of the Likelihood of Snooping

Having identified gender and smartphone ownership status as variables of interest, we finally aimed to understand how they predicted the likelihood of engaging in snooping intrusions.

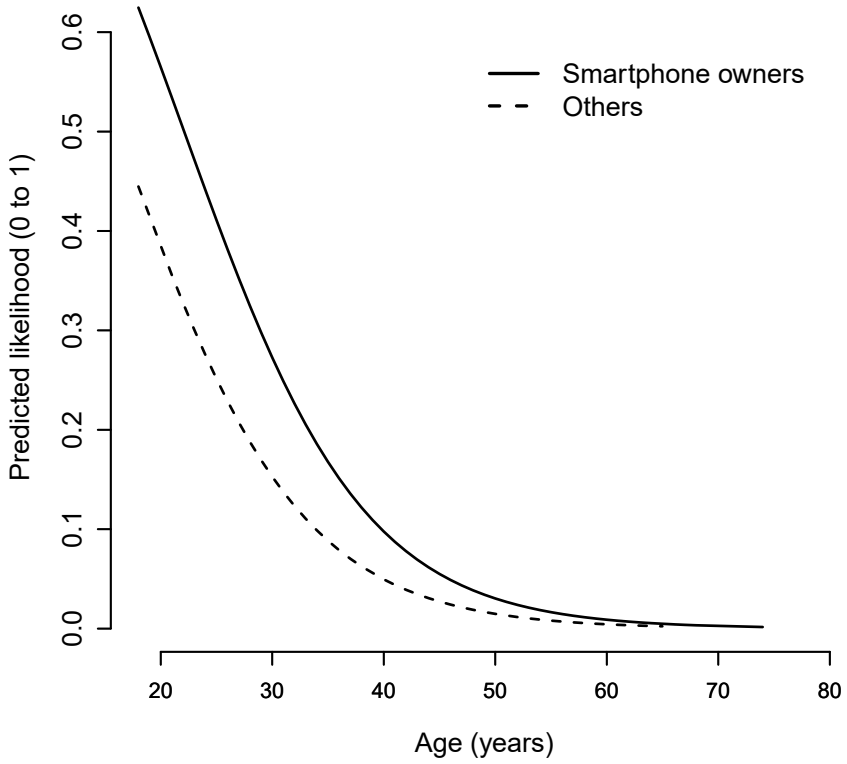

Figure 4.2: Likelihood of having engaged in "snooping" intrusions in the preceding year, by age and smartphone ownership status. Likelihood predicted with a list experiment regression model (Blair and Imai, 2012) of responses to the survey administered to estimate prevalence of "snooping" (n = 1,381). Regression coefficients are shown in Table 4.9.

Variables	**Treatment item**		**Control items** $\mathbf{h_0(y;\, x, \psi_0)}$		**Control items** $\mathbf{h_1(y;\, x, \psi_1)}$	
	Estimate	SE	Estimate	SE	Estimate	SE
(Intercept)	2.014	1.714	-1.167	0.194	-3.529	4.567
Age	-0.124	0.057	-0.002	0.004	-0.024	0.018
Owns smartphone	0.732	0.953	0.832	0.122	3.824	4.542

Table 4.9: List experiment regression, modelling likelihood of having engaged in "snooping" intrusions in the preceding year, as a function of age and having a smartphone or not. Coefficients from a regression using Maximum Likelihood estimation with the Expectation-Maximization algorithm, as described in Blair and Imai (2012), with control group parameters not constrained to be equal. Model constructed from responses to the survey administered to estimate prevalence of "snooping" (n = 1,381), in which the treatment item in the list experiment was whether someone had "looked through someone else's cell phone without their permission" in the preceding year.

For variable selection, we had used number of items selected, controlled by group membership, as an indicator of higher probability. For the final model, we wanted to look at predicted probability, while using both variables as predictors, and accounting for possible non-linear relationships. Thus, linear least squares regression models of participant responses would have not been adequate.

Recently, it has been noted that although list experiments cannot reveal what each participant responded to the treatment item, it is still possible to estimate conditional and joint proportions (Corstange, 2008; Glynn, 2013), and thus model the joint probability distribution (Blair and Imai, 2012; Imai, 2011). Using the R *list* package (Blair and Imai, 2010), we created such a model, and were thus able to summarize the proportion of respondents identifying with the treatment item, as a function of age and ownership status.

Table 4.9 shows the coefficients of that model, and Figure 4.2 depicts it graphically. It shows two clear trends:

- There is a sharp, concave decline in likelihood of snooping as people get older. Each additional year of age disproportionately decreases the likelihood of snooping on others.

- Those who own smartphones are more likely to engage in snooping. The difference decreases, and eventually disappears, as people get older.

The model also suggests that the youngest participants who are smartphone owners are more likely to have snooped on others than to have abstained from it. Thus, for some groups, conducting snooping intrusions, as we have defined them, may be the norm, not the exception.

4.2.3 Discussion

Through a list experiment, we estimated the 1-year prevalence of what we called snooping intrusions to be 30.8% in an online sample. With post-stratification weighting, we generalized that finding to a national population, estimating that 20% of US adults had engaged in snooping in a 1-year period. Looking at specific subsets of the sample, some apparent trends emerged but, due to the nature of list experiment data, comparisons between raw subsets can be misleading. Expanding our analysis, we did not find gender, level of education, or geographical sub-region to be strongly related to snooping behavior. We did however find that being younger, and owning a smartphone, was independently linked to the likelihood of engaging in snooping. In the sample, those that did not own smartphones were, indeed, much less likely to have engaged in snooping (11% 1-year prevalence), while those that were younger were more likely (52% 1-year prevalence in the 18-24 age group).

We note, however, that being young and owning a smartphone is very much related. In 2015, in the US, 85% of those between 18 and 29 owned a smartphone, whereas for those that were 65 or older, the proportion was 27% (Pew Research Center, 2015b). In our sample, there was also a notable relationship between the two variables ($r_{\text{point-biserial}} = 0.28$). Such suggests there may be other variables, which we did not examine, which could be a common explanatory factor associated with both age and ownership.

4.3 Snooping And Depth of Adoption

Being young and owning a smartphone, variables which the model suggests to be indicative of higher likelihood of engaging in snooping, are also typical characteristics of "digital natives." This population is known to be much more aware and concerned about threats to security involving people in close proximity, such as unauthorized physical access (Kurkovsky and Syta, 2010). Where does that concern stem from? We hypothesized that those who use smartphones intensively as a gateway to their social lives, thus producing privacy-sensitive information, become, by their own experiences, more aware of what one may learn from snooping. Thus, they would not only be more concerned about others snooping on them, but also more likely to snoop on others.

To explore whether this hypothesis was plausible, we ran a follow-up list experiment. We again examined the likelihood of engaging in snooping, but this time restricted the target population to smartphone users only. We explored how that likelihood is influenced by age, as well as by the degree to which people use their devices for personal purposes, in ways that may leave traces of potentially privacy-sensitive data.

4.3.1 Study Design

4.3.1.1 Instrument

We created a new online survey, similar to the one used in the previous study. We removed the questions about gender, level of education, geographical region, and the first question on smartphone ownership (SMART1); and kept the list experiment question, the question about age, and the question about the kind of smartphone the participant had (SMART2).

We added an additional question, shown in a second page. This question was a Likert-type scale of depth of adoption for privacy-sensitive purposes, with 10 items to be answered in 7-levels rating scales. As an example, one item was "I use my smartphone to look up information about health conditions". For each item, participants rated their perceived degree of frequency of use, from "Never" (1) to "All the time" (7). Items were based on behaviors of smartphone users that were reported in a Pew survey (Pew Research Center, 2015b). The scale is reproduced in Table 4.3.

4.3.1.2 Procedure

The survey was deployed to MTurk, following the same procedure as in the previous study. The advertisement asked specifically for smartphone users, both in the title ("Survey of smartphone users") and the description ("[...] Do not accept this task if you do not regularly use a smartphone"). Data was prepared in the same way as the previous study, resulting in the exclusion of 7 responses (1%). All participants were paid $0.25.

4.3.2 Findings

4.3.2.1 Participants

There were 653 valid responses, 314 of which in the control group, and 339 in the treatment group. The majority of participants (56%) reported having an Android smartphone, followed by an iPhone (41%), Windows Phone (3%) and Blackberry (<1%). No participants

PROMPT Here are some statements about smartphone usage for personal purposes.

Please answer on a scale from 1 to 7, where a 1 means that the statement indicates something you *feel like* you never do, and a 7 means that the statement indicates something you *feel like* you do all the time.

You can also use the values in-between to indicate where you fall on the scale.

Item-1 I use my smartphone to check my personal email account.

	1	2	3	4	5	6	7	
Never	◯	◯	◯	◯	◯	◯	◯	All the time

Item-2 I use my smartphone to take pictures of myself or of people close to me.

	1	2	3	4	5	6	7	
Never	◯	◯	◯	◯	◯	◯	◯	All the time

Item-3 I use my smartphone to go on social networks (like Facebook, Twitter, Snapchat) with my personal account.

	1	2	3	4	5	6	7	
Never	◯	◯	◯	◯	◯	◯	◯	All the time

Item-4 I use my smartphone to exchange instant messages with people who are close to me.

	1	2	3	4	5	6	7	
Never	◯	◯	◯	◯	◯	◯	◯	All the time

Item-5 I use my smartphone to to look up information about health conditions.

	1	2	3	4	5	6	7	
Never	◯	◯	◯	◯	◯	◯	◯	All the time

Item-6 I use my smartphone to do online banking on my personal accounts.

	1	2	3	4	5	6	7	
Never	◯	◯	◯	◯	◯	◯	◯	All the time

Item-7 I use my smartphone to look up jobs or submit job applications.

	1	2	3	4	5	6	7	
Never	◯	◯	◯	◯	◯	◯	◯	All the time

Item-8 I use my smartphone to look up government services or information.

	1	2	3	4	5	6	7	
Never	◯	◯	◯	◯	◯	◯	◯	All the time

Item-9 I use my smartphone to look up directions to places, or to get turn-by-turn navigation.

	1	2	3	4	5	6	7	
Never	◯	◯	◯	◯	◯	◯	◯	All the time

Item-10 I use my smartphone to organize personal affairs (for instance, access personal notes, calendar or shopping list).

	1	2	3	4	5	6	7	
Never	◯	◯	◯	◯	◯	◯	◯	All the time

Figure 4.3: Scale used to measure degree of privacy-sensitive adoption of smartphones ("depth of adoption"). The scale aims to quantify, in a range from 10 to 70, the degree of privacy-sensitive smartphone use. Each item in the scale refers to a type of smartphone use that can leave potentially sensitive information on the device. Participants were asked to rate their perceived frequency of use in scales from 1 to 7. Item presentation was randomized.

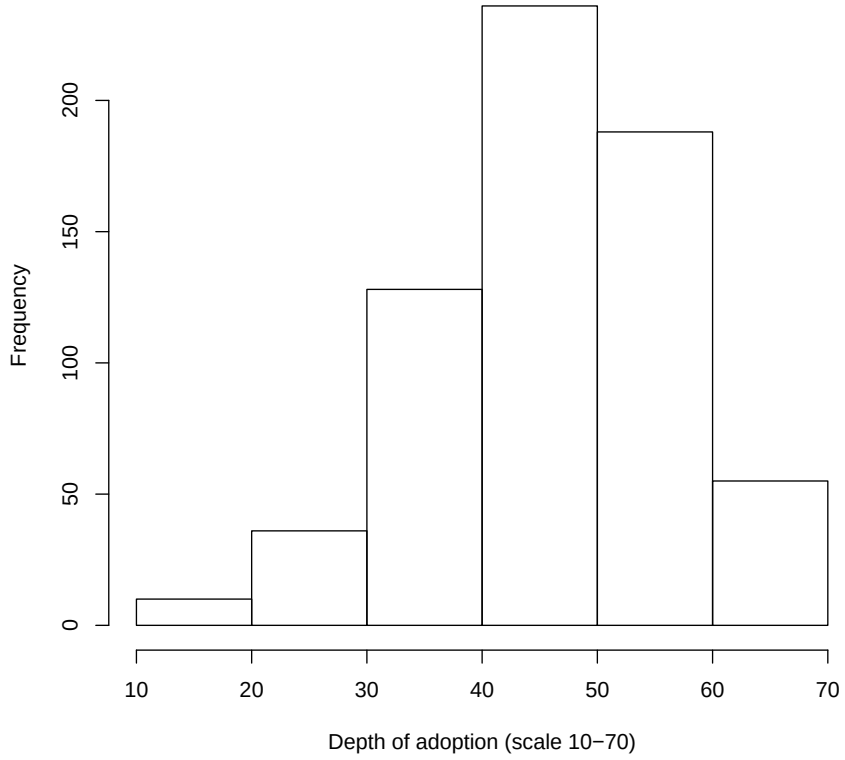

Figure 4.4: Distribution of responses to a scale of depth of privacy-sensitive adoption of smartphones, administered in a survey to explore the relationship between "snooping" and depth of adoption (n = 653). The score for each participant is the sum of ratings to individual items, and can thus range from 10, from a participant responding 1 (Never) to the 10 items in the scale, to 70, from responding 7 (All the time) to the 10 items. Table 4.5 shows the distribution of responses to individual items.

selected the option "I do not have a cell phone", that was kept to exclude responses in case of inattentive reading of the advertisement.

4.3.2.2 Depth of Adoption

Responses to the depth of adoption scale, which can range from 10 to 70, were from 16 to 70, and somewhat skewed toward the higher end. The middle point of the scale is 40, and the mean response was 44.66 (SD [standard deviation] = 10.6). The distribution of responses, for the scale and individual items, is shown in Figures 4.4 and 4.5, respectively.

Responses to the depth of adoption scale were, as expected, negatively correlated with age (r = -0.18, $t(651)$ = -4.78, $p < 0.00001$). This correlation, however, was not strong (according to Cohen's effect size criteria, it falls between small, 0.1, and medium, 0.3). Because depth of adoption, as it was measured, was relatively independent of age, it could therefore more easily be interpreted as a predictor of likelihood of engaging in snooping.

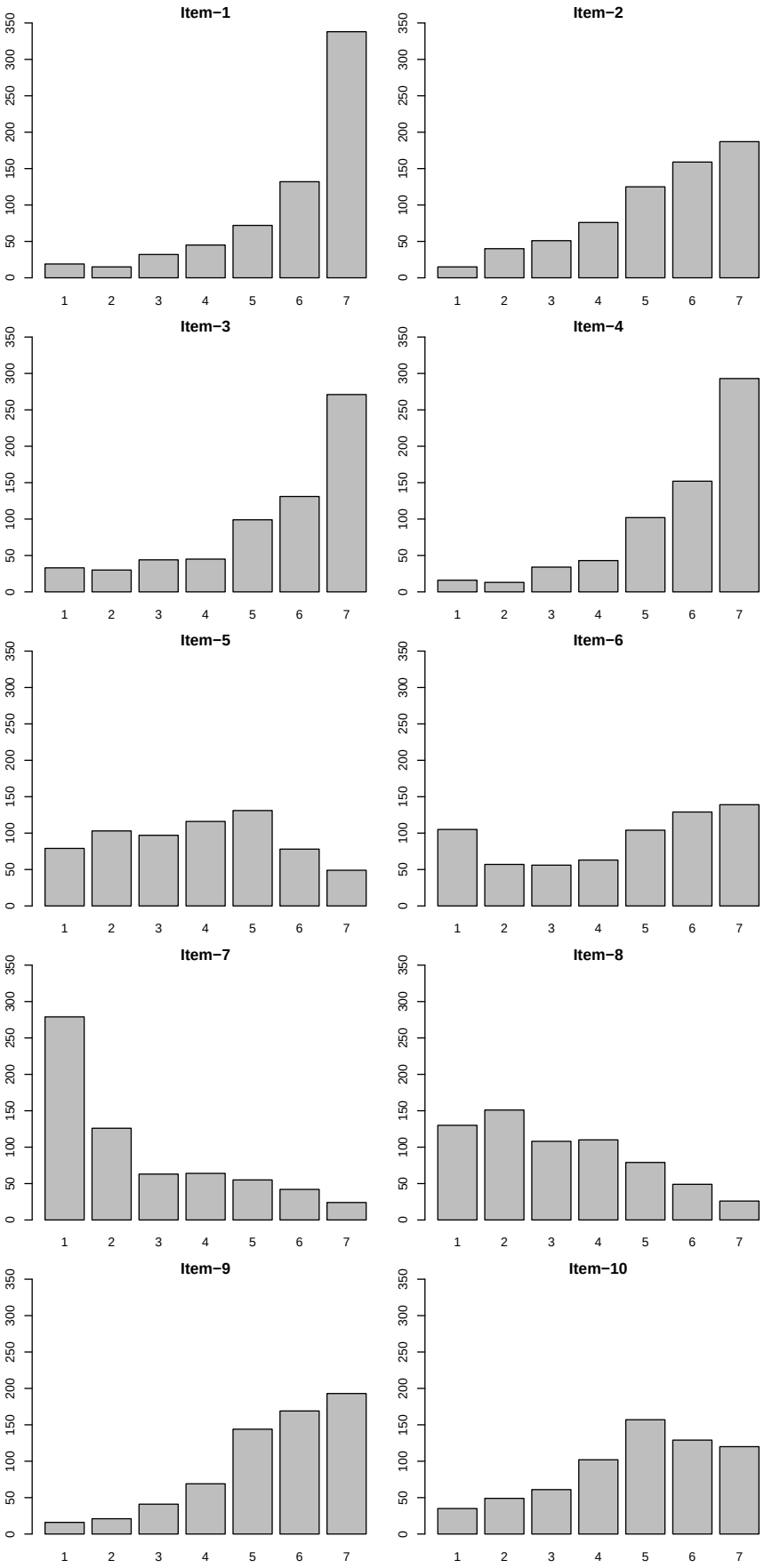

Figure 4.5: Distribution of responses to the 10 items in a scale of depth of privacy-sensitive adoption of smartphones, administered in a survey to explore the relationship between "snooping" and depth of adoption (n = 653). For each item, participants responded on a rating scale from 1 (Never) to 7 (All the time).

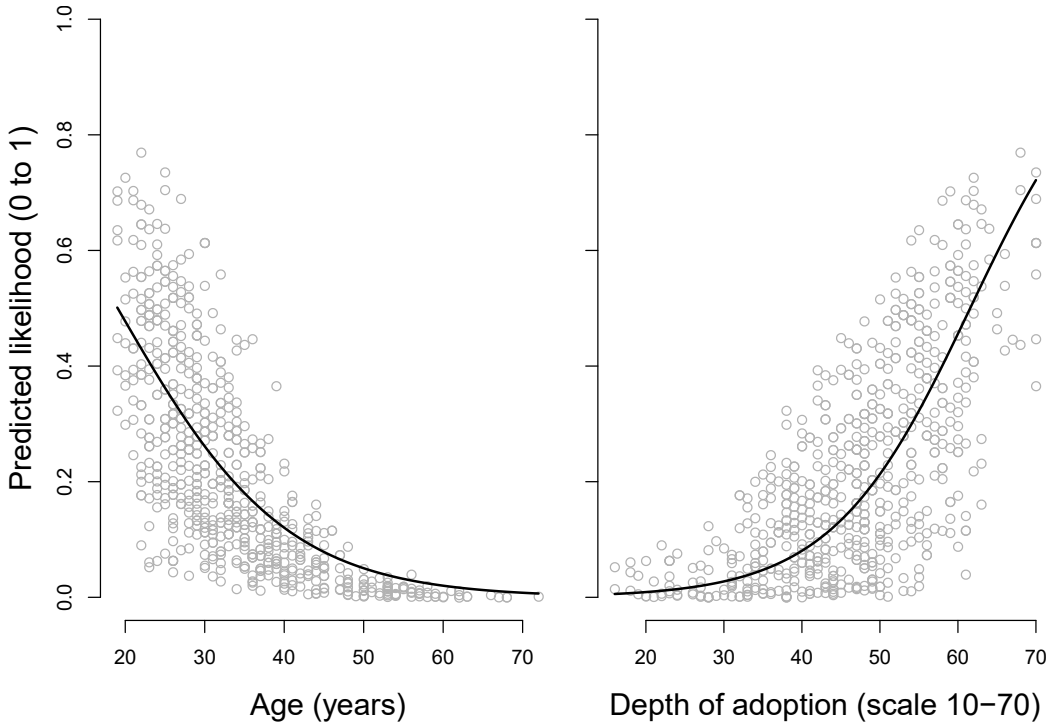

Figure 4.6: Likelihood of having engaged in "snooping" intrusions in the preceding year, by age (left panel) and depth of privacy-sensitive adoption (right panel). Likelihood predicted with a list experiment regression model (Blair and Imai, 2012) of responses to the survey administered to explore the relationship between "snooping" and depth of adoption (n = 653). Dots represent per-participant predicted likelihood based on a model with both age and depth of adoption as predictors. Trend lines represent the respective single-predictor regression models. Regression coefficients are shown in Table 4.10.

4.3.2.3 Prevalence Estimate

The difference-in-means in responses to the list experiment, between treatment and control group, was 27.9%, slightly lower than in the previous study, but not significantly so ($\chi^2(1) = 1.47$, $p = 0.23$, in a test for equality of proportions with continuity correction). Although this was not the focus of this study, it indicates that the prevalence estimates obtained in the previous study were stable.

4.3.2.4 Another Model of the Likelihood of Snooping

Using the same procedure as in the previous study, we created a model of likelihood of having engaged in snooping, based on the age and depth of adoption variables. Figure 4.6 depicts the model predictions, and Table 4.10 shows the regression coefficients. In the left panel of the figure, the dots represent model predictions for each participant, on an x-axis of age, and the line represents a reduced model, with only age as predictor. In the right panel, the dots represent model predictions on an x-axis of the depth of adoption scale ratings, and the line represents a reduced model with just the scale as predictor.

If there were noticeable differences in the pattern of dispersion of dots in relation to the

Model / Variables	Treatment item		Control items	
	Estimate	SE	Estimate	SE
Likelihood by age				
(Intercept)	1.80821	1.48669	-0.17064	0.17876
Age	-0.09492	0.05080	-0.00728	0.00474
Likelihood by "depth of adoption"				
(Intercept)	-6.95467	4.36000	-1.00872	0.21807
Depth of adoption	0.11296	0.07714	0.01315	0.00446
Likelihood by age + "depth of adoption"				
(Intercept)	-1.48857	4.23927	-0.88936	0.37492
Age	-0.11248	0.06047	-0.00457	0.00536
Depth of adoption	0.07617	0.06999	0.01360	0.00505

Table 4.10: Three list experiment regressions, modelling likelihood of having engaged in "snooping" intrusions in the preceding year, as a function of a) age; b) depth of privacy-sensitive adoption of smartphones, and c) both. Coefficients from regressions models using Maximum Likelihood estimation with the Expectation-Maximization algorithm, as described in Blair and Imai (2012). Model constructed from responses to the survey used to explore the relationship between "snooping" and depth of adoption (n = 653), in which the treatment item in the list experiment was whether someone had "looked through someone else's cell phone without their permission" in the preceding year.

lines, such could be interpreted as one variable being a stronger predictor than the other (the stronger predictor should show less dispersion, or none at all). What is observable, however, is that neither the age or depth of adoption variables explain the other away.

The model with both variables as predictors has Log-likelihood of -868.786, which is higher than either the reduced models for only age (-880.458) and only depth of adoption (-873.834), indicating better fit. Predictions of both reduced models are strongly correlated to the ones of the larger model (age: $r = 0.75$, $t(651) = 28.6$, $p < 0.00001$; depth of adoption: $r = 0.71$, $t(651) = 25.7$, $p < 0.00001$). They are also correlated amongst themselves, as would be expected from the correlation of the variables, but not strongly ($r = 0.14$, $t(651) = 3.7$, $p = 0.00022$). Again, these correlations indicate that neither variable explains the other away, and both contribute independently to the larger model.

4.3.3 Discussion

We found evidence consistent with the theory that people who use their smartphones in ways that generate privacy-sensitive records are more likely to snoop on others. Higher depth of adoption, as measured by a short scale we developed, predicted higher likelihood of identifying with the list experiment item indicating having "looked through someone else's cell phone without their permission" in the preceding year, even when controlling for age. However, in our models, depth of adoption did not explain away the effect of age that we had found.

It is also likely that there are other factors related to age which we did not measure, but play a role in predicting higher likelihood of snooping, such as tech-savviness, or degree of volatility of social relationships.

4.4 Conclusion

4.4.1 Summary

In this chapter, we have explored the prevalence of snooping intrusions on mobile devices, and found evidence that it is considerably higher than previously estimated. Our findings suggest that snooping intrusions are associated with increased adoption of mobile devices, and thus, that it is the youngest, those who use smartphone, and particularly those who use smartphones in ways that it stores privacy-sensitive data, that are more likely to snoop on others. In some segments of the population, people were more likely to "have gone through someone else's phone without permission", than not, in a period of one year.

To gather these findings, we conducted a series of empirical studies. In a first study, we crafted items for a list experiment instrument. Then, we conducted two list experiment studies, that inform on the prevalence of snooping intrusions. Employing conservative design choices, that may have had the effect of underestimating prevalence, we were still able to estimate 1-year prevalence rates for the MTurk population, and, by weighting, for the U.S. adult population, which are much higher than previous lifetime prevalence indicators. Furthermore, we uncovered predictors of the likelihood of engaging in snooping, and discerned independent population trends related to age and adoption of smartphones. We hypothesize that one mechanism for the observed trends is that users learn by their own experiences the kinds of valuable information kept on smartphones, which makes them more capable of engaging in snooping intrusions.

4.4.2 Limitations

We have referenced some limitations of our application of the list experiment method in Section 3.4.2, namely limited generalizability, and limited reduction of measurement bias. The findings of the studies reported in this chapter should, again, be interpreted in light of those limitations.

In this chapter, we report on how we selected items for the list experiment through an empirical process (Section 4.1). This process highlighted method limitations we had not previously considered. One such limitation stems from us not measuring direct responses to list items when shown together to participants. Because of that, we could not know before administering the list experiment if there were interactions between items that could increase the chances of ceiling or floor effects. Another limitation was that our process for item selection was not informative as to the possibility of contrast effects between the selected controls and the treatment items, which could hurt the credibility of the list. We understood that these limitations could potentially result in further underestimating the true prevalence of intrusions. We did however consider it an acceptable risk, as it represented a conservative design choice. After obtaining the list experiment data, we tested for a list design effect, and did not find evidence of one being present (see Section 4.2.2.4).

The findings of the follow-up study reported in this chapter, looking at the relationship between snooping intrusions and depth of privacy-sensitive adoption (Section 4.3), should also be interpreted with caution, since the scale we used was not validated. Because our intent was exploratory, we crafted an *ad hoc* scale, based on previous public opinion polls, which *prima facie* could give some sense of the distribution of our construct of interest. We did however not validate that it did so, or how. Notably, the scale items did not measure the actual frequency of

certain behaviors, but how people *perceived* that frequency, which may be a weaker proxy for the construct of depth of privacy-sensitive adoption.

5

Prevalence of "Facejacking" Intrusions

In this chapter, we quantify intrusions to Facebook accounts by social insiders, known as *facejacking*. Facebook users often share and maintain personal and potentially sensitive information on their accounts, including messages, pictures and videos (e.g., Krishnamurthy and Wills, 2008). This information can entice others to try to access accounts without the owner's consent. Those who have a social relationship with the account owner are of special concern — the proximity between parties can make it easier for them to access devices without permission, and, through them, Facebook accounts.

For the purposes of this research, we define facejacking intrusions as *situations in which a person accesses the Facebook account of someone else, using Facebook's end-user interfaces, like the web or mobile application, on the victim's device, and without the victim's permission.* We consider a victim's device to be one that is regularly controlled by the victim. This includes not only personal devices, but also work computers and shared devices in a household.

Facejacking, as we have defined it, is not limited to instances of unauthorized access with the intent of playing pranks. Confining the understanding of facejacking to well-meaning pranks can lead to underestimating the potential adverse effects of such incidents, which may include unauthorized access to private data for a variety of reasons. For instance, the act of posting potentially embarrassing material using the victim's account can, in some cases, be a well-meaning prank, but there have been situations in which these actions have been regarded as defacement and resulted in criminal prosecution (e.g., Business Insider, 2014).

We next report on empirical studies we conducted to estimate the prevalence of these intrusions. As in the previous chapter, we estimated the prevalence of intrusions with a survey (n = 1,308) conducted on Amazon's Mechanical Turk service (MTurk). Our study targeted the U.S. Facebook users population. According to a 2014 Pew survey, 62% of U.S. adults use Facebook (Pew Research Center, 2015a). This made it easy to find Facebook users among U.S. MTurk workers.

Since direct questions about intrusions are sensitive, we again opted to use the list experiment technique. We followed, for the most part, the same method as in Chapter 4, since facejacking intrusions are in many ways similar to snooping intrusions. Both involve unauthorized physical access to devices, and in both cases the parties are likely to be known to each other. We did, however, introduce one significant deviation, using two treatment groups instead of just one. One of the two treatment group was shown a treatment item which identified the participant as having been a victim of a facejacking intrusion, while the other identified the participant as a perpetrator. The difference between those two estimates was expected to offer some insight into how common it is for people to be unaware they have been facejacked.

We next report on how we crafted items for the list experiment instrument, in Section 5.1; and provide detail about the facejacking list experiment, in Section 5.2.

5.1 Item Selection

To be able to quantify facejacking intrusions, we first tasked ourselves with crafting the set of items that would compose the list experiment instrument. A central design consideration in list experiments is the composition of the list question. Common recommendations when building list questions include: avoiding floor and ceiling effects (that is, many participants identifying with none or all statements in the list), avoiding lists that are too long or too short, and avoiding items that stand out in relation to the others (for more detail on these design considerations, see Section 4.1). We thus aimed to create a set of control items for the list question, with statements related to Facebook usage; and to create two treatment items, one referring to being facejacked, and another referring to facejacking another person.

We followed distinct procedures to create treatment and control items. Treatment items were the result of a series of discussions among research team members; whereas control items, as in Chapter 4, were the result of an empirical process.

5.1.1 Treatment Items

For treatment items, our goal was to capture the construct of interest, with two neutrally-worded formulations: a statement that would identify participants as victims of facejacking, and a statement that would identify them as perpetrators of facejacking. To that end, we brainstormed the wording among the research team, and went through multiple rounds of refinement until we were satisfied the statements were understandable, neutral, and not excessively verbose.

We settled on the following wording for self-identification statements:

- I have used a device of someone I know to access their Facebook account without permission. (Perpetrator)
- Somebody I know has used my device to access my Facebook account without permission. (Victim)

We avoided, as much as possible, using terms with security connotations, like "perpetrator", "intrusion", "victim", or "insider", both to avoid biasing participants, and to reduce contrast with control items. We used "my device" to imply physical unauthorized access, "someone/somebody I know" to imply that it was perpetrated by an insider, and "access without permission" to refer to the intrusion.

5.1.2 Control Items

5.1.2.1 Study Design

To create control items, we opted for an empirical approach, running a direct question survey with MTurk workers. Our goal was to find a combination of control items that would minimize the chances of ceiling and floor effects. In particular, we wanted to find a set of four statements for which participants would rarely agree with all of the statements, or none of the statements.

Our task advertisement asked for participants who had a Facebook account, and again avoided charged terms such as “privacy” or “intrusion”. The survey consisted of demographic questions including age, level of education, and state of residence. We also explicitly asked participants to indicate whether or not they had a Facebook account.

Following these questions, participants responded to a list of 22 check-box items with the prompt “Please check all statements that apply to you”. Table 5.1, rows 1 to 20, reproduces the statements participants could check, which were the candidate control items. The statements were drawn from previous research on motivations for Facebook use (Spiliotopoulos and Oakley, 2013) and common Facebook use cases developed by the research team in brainstorming sessions. We also included the two treatment items in the survey, so that we could have estimates both from direct questioning and from the list experiment. The ordering of the statements was randomized when presented to each participant.

Participants were paid $0.20 for completing the task. Only workers with location set to U.S. were allowed to participate. At the beginning of the survey, a filter based on IP addresses further prevented participation from non-U.S. locations.

5.1.2.2 Findings

We collected 202 complete responses, and excluded 28 in which participants either indicated not using Facebook, or took less than 40 seconds to complete the survey. We selected the threshold of 40 seconds based on a pilot study with five native English speakers in which we measured how long it took them to read through the survey. The remaining 174 participants reported an age range from 19 to 69 (mean = 33.7, SD = 10.6), and a gender distribution of 43% male, and 57% female. Table 5.1 shows the percentage and number of respondents who checked each statement.

To select the control items, we wrote a script that computed all possible combinations of four statements, and ordered them by how many cases of floor and ceiling effects they would cause if they had been administered in conjunction to the same sample. Statements 7, 8, 13, and 16, also shown in Figure 5.1, were selected for being those that would create the least such cases.

5.1.3 Discussion

To design a list experiment instrument aimed at measuring facejacking, we selected four control items from an empirical process, and two treatment items from brainstorming exercises. Our choice of items is intended to reduce the possibility of ceiling, floor, and contrast effects, thereby minimizing the effects of social desirability bias.

Although that was not the focus of the item selection process, we can observe in the data how questions about facejacking are sensitive. Having included the treatment items in the direct-question survey, we estimated that, under direct questioning, 8.6% of participants identified as perpetrators of facejacking, and 9.2% as victims. Peeking at the results of the list experiment (described in the next section), the estimates we obtained through direct questioning were less than half of those we obtained with the list experiment.

	Statement	%
1	I have posted a message in a group on Facebook and received a reply	62.6%
2	Someone I know has posted content on my Facebook wall	57.5%
3	I have received 5 or more unsolicited messages from strangers on Facebook	32.4%
4	One of my relatives has sent me a friend request on Facebook	65.4%
5	I have posted a picture of myself on Facebook	66.5%
6	Someone liked one of the pictures I posted on Facebook	65.9%
7	**I have more than 300 friends on Facebook**	45.3%
8	**I am friends with one of my parents on Facebook**	43.6%
9	I check Facebook every day	79.3%
10	On average, I spend more than 30 minutes on Facebook every day	55.9%
11	I have changed my Facebook profile picture in the last 12 months	60.9%
12	In the last week, I have clicked on a link posted on my Facebook newsfeed	50.8%
13	**I have commented or liked a post in the last month on Facebook**	68.7%
14	I am a member of a Facebook group	76.0%
15	In the last week, I have checked Facebook while at work	57.5%
16	**I have reported an account on Facebook**	26.8%
17	I re-shared someone else's post on Facebook	62.0%
18	I have made my birth date publicly visible on Facebook	50.3%
19	I have clicked on an advertisement on Facebook	58.7%
20	I have responded to an event invitation on Facebook	55.3%
21	**I have used a device of someone I know to access their Facebook account without permission**	8.6%
22	**Somebody I know has used my device to access my Facebook account without permission**	9.2%

Table 5.1: Statements in a question administered to 174 MTurk workers, and respective percentages of participants who identified with them. Participants were prompted with "Please check all statements that apply to you", and each statement had a corresponding checkbox. Statements 1 to 20 were candidate control items for the list experiment question. Statements 21 and 22 correspond with the treatment items of the list experiment question, which had been selected in advance. The items which ultimately appeared in the list experiment are marked in bold.

5.2 Prevalence Estimation

5.2.1 Study Design

5.2.1.1 Instrument

Following our item selection exercise, we crafted the list experiment instrument, reproduced in Figure 5.1. The list question contained the 4 control items selected through the empirical study reported in the earlier section, in addition to one attention check question item. We included an attention check following our previous finding that attention checks reduce measurement error (Section 3.3). We selected a statement which we did not expect any participants to agree with ("I have had dinner with the founder of Facebook, Mark Zuckerberg"). In the versions of the list question administered to participants in one of the two treatment groups, the respective treatment items were also included. Other than the list question, the survey contained questions on participant age, gender, level of education, U.S. state of residence, and which online social networks participants used.

5.2.1.2 Procedure

We again recruited among U.S. MTurk workers and limited participation to those who were accessing our survey from U.S. IP addresses. As before, we avoided words like "privacy" or "intrusion" in the task advertisement and consent form, informing participants that they were filling out a survey about their behavior on social media, and that being a Facebook user was a requirement for participation.

After providing consent, participants were randomly assigned to either the *Control* group, the *Treatment-Perpetrator* group (which included the treatment item indicating that participants had been perpetrators of facejacking), or the *Treatment-Victim* group (which included the treatment item indicating they had been victims of facejacking). Each worker was paid $0.20 for completing the survey.

5.2.2 Findings

5.2.2.1 Data Preparation

We received 1,512 complete responses and cleaned up the data by applying the following exclusion criteria:

- Responses in which participants had agreed with all statements, including the attention check.
- Responses in which participants failed to confirm they used Facebook.
- Responses that took less than 30 seconds to complete (based on a prior pilot with 5 native English speakers in which we measured how long it took them to read through the survey).
- Responses in which the reported age was below 18.

This left us with 1,308 responses, on which the following analysis is based.

LIST EXPERIMENT Below is a list of statements that describe various experiences that you may have encountered in the past year. To preserve your anonymity, select **how many** statements that apply to you, not **which ones**. [RANDOMIZE:]

- I have more than 300 friends on Facebook.
- I am friends with one of my parents on Facebook.
- I have commented or liked a post in the last month on Facebook.
- I have reported an account on Facebook.
- I have had dinner with the founder of Facebook, Mark Zuckerberg.
- **I have used a device of someone I know to access their Facebook account without permission.** [Treatment-Perpetrator group]
  Somebody I know has used my device to access my Facebook account without permission. [Treatment-Victim group]

Statements that apply to you:

	0	1	2	3	4	5	6	
None	◯	◯	◯	◯	◯	◯	◯	All

AGE How old are you? ⌊_⌋⌊_⌋⌊_⌋

GENDER What is your gender? ◯ Male ◯ Female ◯ Other

EDUCATION What is your highest level of education?

◯ High School
◯ College degree
◯ Graduate School
◯ Other: _____

COUNTRY In which country do you reside? ◯ Alabama ◯ Afghanistan ◯ Albania [...]

STATE In which state do you reside? ◯ Alabama ◯ Alaska ◯ Arizona ◯ Arkansas [...]

OSN Which of the following social networking sites do you use?

☐ Facebook
☐ Twitter
☐ Reddit
☐ Pinterest
☐ Tumblr
☐ LinkedIn
☐ Other: _____
☐ None

Figure 5.1: Survey instrument used to estimate prevalence of "facejacking" . The first question is a list experiment question, with items presented to participants in random order. The list question includes 4 control items, and, to participants randomly assigned to one of the two treatment groups, one of the two alternative treatment items. The fifth item in the list question is an attention check. The treatment items are marked here in bold for illustrative purposes only. Participants in the control group received the same question without any of the treatment items.

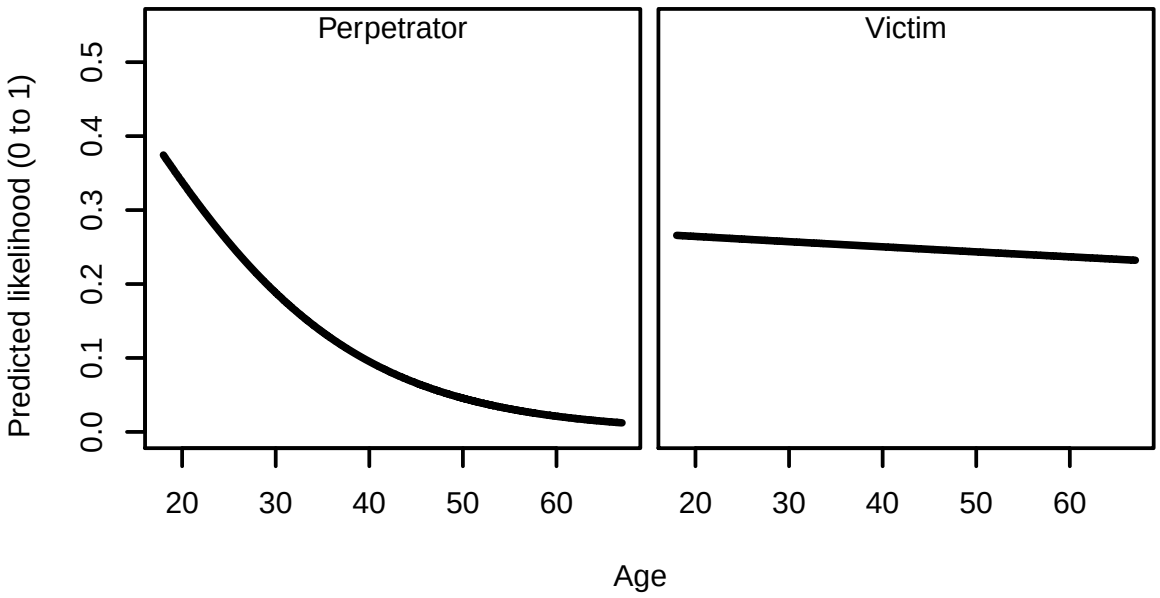

Figure 5.2: Likelihood of having been a perpetrator or a knowing victim of "facejacking", by age of participants. Likelihood predicted with a list experiment regression model (Blair and Imai, 2012) of responses to the survey administered to estimate prevalence of "facejacking" (n = 1,308). Regression coefficients are shown in Table 5.2.

5.2.2.2 Participants

Out of the 1,308 participants who provided acceptable responses, 440 had been assigned to the control group, 423 to the Treatment-Perpetrator group, and 445 to the Treatment-Victim group. Overall, reported ages ranged from 18 to 72, with the mean being 32.9 (SD = 10.16). Reported gender identification was 49% female, and 51% male. Most participants indicated being college graduates (52%), followed by those indicating being high school graduates (29%), and those indicating having post-graduate degrees (16%). Grouping reported states of residency into census regions, the geographical distribution was 32% South, 21% West, 21% Midwest, and 18% Northeast. On average, participants reported being on 3.29 online social networks (SD = 1.38), with only 9% reporting being only on Facebook. Reddit (65%), Twitter (56%), Pinterest (37%), LinkedIn (23%), Tumblr (19%), and Instagram (9%) were the most popular online social networks co-occurring with Facebook among participants.

To test for a priori demographic differences between the control and the treatment groups, we ran a logistical regression of group assignment per all available demographic variables, and then applied the stepwise procedure for variable elimination. The selected model had no demographic variables, which indicates a lack of evidence for a priori demographic differences between groups.

5.2.2.3 Prevalence Estimates

The mean number of items selected was 2.334 (SE = 0.046) in the control group, 2.574 (SE = 0.053) in Treatment-Perpetrator group, and 2.546 (SE = 0.053) in Treatment-Victim group. The estimates of participants identifying with the treatment items, based on the differences in means, are thus:

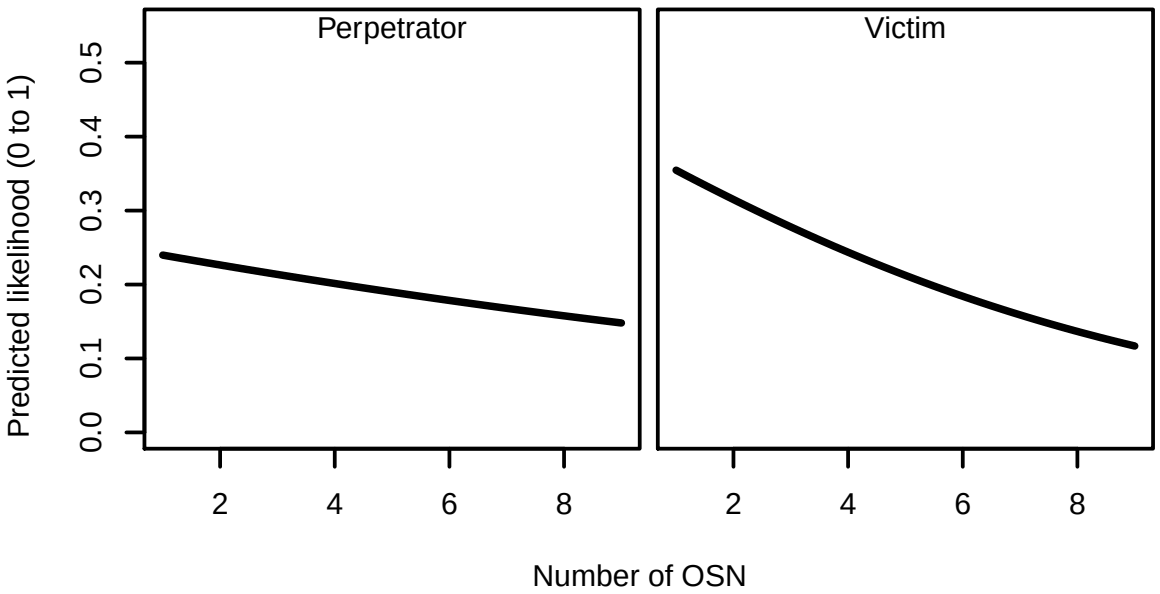

Figure 5.3: Likelihood of having been a perpetrator or a knowing victim of "facejacking", by count of online social networks participants reported using. Likelihood predicted with a list experiment regression model (Blair and Imai, 2012) of responses to the survey administered to estimate prevalence of "facejacking" (n = 1,308). Regression coefficients are shown in Table 5.2.

Model / Variables	**Treatment item**		**Control items**	
	Estimate	SE	Estimate	SE
Likelihood of being perpetrator by age				
(Intercept)	0.90911	1.78994	0.84997	0.15176
Age	-0.07903	0.06450	-0.01575	0.00424
Likelihood of being perpetrator by OSN count				
(Intercept)	-1.07957	1.00454	-0.12703	0.12999
# OSN	-0.07452	0.30095	0.13738	0.03767
Likelihood of being victim by age				
(Intercept)	-0.95046	0.95651	0.91442	0.15598
Age	-0.00367	0.02685	-0.01853	0.00438
Likelihood of being victim by OSN count				
(Intercept)	-0.42123	0.93241	-0.15046	0.13387
# OSN	-0.17790	0.27317	0.13477	0.03778

Table 5.2: Four list experiment regressions, modelling likelihood of having been either a perpetrator or a knowing victim of facejacking, as a function of a) age of participant and b) count of online social networks participants reported using. Coefficients from regressions models using Maximum Likelihood estimation with the Expectation-Maximization algorithm, as described in Blair and Imai (2012). Model constructed from responses to the survey administered to estimate prevalence of "facejacking" (n = 1,308).

- **Perpetrator** 24.0% (SE = 0.070)
- **Victim** 21.2% (SE = 0.070)

In Chapter 4, we found evidence suggesting that snooping on mobile phones was more prevalent among younger people, and among people who adopted smartphones in ways that retained more private data. To verify if similar effects could be found among victims and perpetrators of facejacking intrusions, we ran list experiment regression models (Blair and Imai, 2012) on the age variable, and lacking a specific measure of depth of adoption, on the number of OSNs that participants reported using. Figures 5.2 and 5.3 depict those regression models graphically, and Table 5.2 shows the regression coefficients.

Regarding the effect of age (Figure 5.2), there was a visible pattern of decreasing likelihood of being a perpetrator of facejacking as age increased. However, for the likelihood of being a victim, the dependency on age was less pronounced, and nearly flat.

Regarding the effects of participation in online social networks (Figure 5.3), a different pattern seems to prevail than the one we found in Chapter 4. Using a greater number of online social networks was a weak predictor of perpetrating facejacking intrusions on others, and at best predicted a slight decrease of the likelihood. For being a victim, however, the pattern appeared to be clearer: the more online social networks participants used, the less likely they were to be victims of facejacking.

5.2.3 Discussion

Contrasting the estimates obtained through the list experiment (24% perpetrators/21% victims) with the ones obtained through direct questioning (9% perpetrators/9% victims), lead us to two observations.

First, questions about facejacking appear to induce strong social desirability bias. List experiment estimates were more than double the self-reported prevalence. We expected social desirability bias to affect questions about perpetrating facejacking intrusions, as people are generally unwilling to openly admit to behaviors which can be seen as censurable. However, for questions about being a victim of facejacking, possible effects of social desirability bias were more surprising. One possible explanation for such effects is that victims of facejacking may assign themselves responsibility for intrusions, as has been observed in previous work on account hijacking (Shay et al., 2014), and thus be reluctant to answer questions truthfully.

Second, facejacking appears to rarely be silent. The list experiment estimate for being a victim of facejacking is very close to that of being a perpetrator. Such pattern is consistent with victims most often learning about intrusions when they happen.

The regression models we fit also indicate two trends. First, younger people appear more likely to engage in facejacking, mirroring the findings of Chapter 4. Second, people who use more online social networks seem less likely to be victims of facejacking. One possible explanation for this trend is that those who use more online social networks may be more tech savvy and/or more aware of the private information which is retained on online social networks, and thus more motivated to protect themselves.

5.3 Conclusion

5.3.1 Summary

In this chapter, we have reported on an empirical quantification of facejacking intrusions through a list experiment. Our results suggest facejacking happens frequently, with 24% of participants estimated to have implicitly identified with the statement "I have used a device of someone I know to access their Facebook account without permission", and 21% with the statement "Somebody I know had used my device to access my Facebook account without permission".

5.3.2 Limitations

The findings we report in this chapter are, again, not without limitations. Some limitations, as we already mentioned in sections 4.4.2 and 3.4.2, arise from applying the list experiment method. As we noted, the degree to which our estimate of facejacking can be generalized is limited by our sample not being representative; and the degree to which list experiments reduce measurement error is unclear.

As in Chapter 4, our process for selecting items may have further decreased the ability of the list experiment to reduce measurement error, since we do not know whether our approach effectively preempted ceiling, floor, and contrast effects. For control items, it is possible that some candidate control statements might have been perceived as sensitive by some participants and thus subject to the same bias as the treatment statements. For example, some might consider the number of friends they have on Facebook a sensitive subject, if they feel it is correlated with their popularity. Additionally, the wording used for the control items was crafted not only to minimize the likelihood of participants perceiving them as sensitive, but also to limit their contrast with the treatment items. However, the data we collected does not inform on whether contrast effects remained or not. As in Chapter 4, we reasoned that the most likely outcome of possible measurement errors resulting from these limitations would be underestimating the true prevalence, and judged that risk to be acceptable.

Unlike in Chapter 4, we did not attempt to project our sample estimate into a population estimate. The reason for that was the absence of reliable tabulations of the Facebook users population. From informal data that others have collected (Statista, 2018), it appears our sample was younger and slightly more skewed to males than the U.S. Facebook user population.

The extent to which this research applies to other online social networks is also unclear. There is indication that accounts on other online social networks, such as Twitter, are also common targets of intrusions (e.g., Shay et al., 2014), which suggests our findings may not be unique to Facebook.

Outlook

A larger limitation of our efforts to quantify intrusions is the types of findings that are available from our methodological approach. Both in this chapter and in Chapter 4, our results are informative on intrusion constructs which can be both too narrow, and too broad, for our goal of understanding social insider intrusions.

They are too narrow in the sense that "snooping" and "facejacking" are two kinds of security incidents which we can identify from our own frames of reference. Other researchers have looked at intrusions as part of the lived experiences among user groups of interest. Findings which intersect with ours may thus be found in the growing body of research taking the perspective of users of personal computing devices affected by intimate partner violence (e.g., Freed et al., 2018, 2017; Leitão, 2019; Matthews et al., 2017; Woodlock, 2017).

On the other hand, our quantitative findings are too broad, in the sense that we were unable to explore gradations or variations within each of the two types of security incidents we quantified. The prevalence estimates we obtained apply to any incidents which participants may have perceived as instances of snooping, or facejacking, as we have defined them. But there can be substantial variation in the severity of these intrusions. For instance, we do not know what proportion of facejacking incidents were harmless pranks, as opposed to intrusions with graver consequences. The method we chose did not allow us to understand these phenomena at a more granular level.

In the next chapter, we work towards addressing this latter limitation, through a qualitative exploration of *why* and *how* intrusions to personal computing devices happen, and how they are experienced by the people involved.

6

Experiences of Intrusion

Our analysis of unauthorized access to mobile phones (Chapter 4) and Facebook accounts (Chapter 5) indicates that social insider intrusions are common occurrences. The possibility of unauthorized access, prior research has found, often spurs concerns in users of smartphones (see Chapter 2). However, there has been scarce examination of the ways in which incidents of unauthorized access are commonly experienced.

In this chapter, we ask: *how do people experience incidents of unauthorized access to smartphones*? To answer this question, we collected 102 first-person accounts of unauthorized access. We solicited accounts from both people who accessed the smartphone of someone they knew, and from people who had someone they knew access their smartphone. Next, in Section 6.1, we describe how we approached collecting these accounts. Then, in Section 6.2, we explore *what happened* in the incidents participants experienced, including the context leading up to them, the course of events, and the consequences. Finally, in Section 6.3, we explore *how participants describe these incidents*, through two orthogonal themes.

6.1 Method

To obtain detailed accounts of unauthorized access, we designed a study in which participants were asked to write open-ended accounts of past experiences. Our goal was to collect accounts from opposing perspectives: both experiences of having accessed a smartphone of a known person without permission, and experiences of having one's smartphone accessed. Reconciling accounts from opposing perspectives, we reasoned, could offer deeper, and more rigorous insights into the processes involved in incidents of unauthorized access.

We again opted for an online study, in which participants were to engage remotely. We expected self-administration of the data collection instrument to increase willingness to report on sensitive behaviors (see e.g., McNeeley, 2012). Furthermore, online administration offered practical advantages, such as the ability to reach a target sample within reasonable time and cost constraints (see e.g., Lazar et al., 2010, p. 99).

We next describe our data collection instrument, our procedure for collecting data, and how we analyzed it.

QUAL1 At least one of the following has happened:

- I have physically accessed someone else's smartphone without permission.
- I have had my smartphone physically accessed by someone else, without permission.

◯ Yes ◯ No

QUAL2 You and the other person knew one another personally. ◯ Yes ◯ No

AGE What is your age? ◯ 0-24 ◯ 25-44 ◯ 45-64 ◯ 65-74 ◯ 75+

GENDER To which gender identity do you most identify?
◯ Male ◯ Female ◯ _____

STORY Recall a situation where you have either physically accessed a smartphone of someone you know without their permission; or someone you know has physically accessed your smartphone without your permission.

Your task is to describe that situation in a story format.

Instead of using real names, use the following characters:

- **Ash**, the smartphone owner.
- **Val**, the person who accessed the smartphone without permission.

If there are other characters in your story, use fictional names for them as well. To maintain anonymity, use gender-neutral pronouns such as 'they' instead of 'he' or 'she', or 'their' instead of 'his' or 'her'. Do not include any personally identifiable information.

Your story should include details, such as:

- Where did the situation take place and when?
- What was the relationships between Ash and Val?
- Why did Val wanted to access Ash's smartphone?
- How was Val able to get access to Ash's device?
- What did Val do on Ash's smartphone?
- How, if at all, did Ash ever learn about Val having accessed the smartphone?
- Were there any consequences?

You do not need to answer every question above explicitly, but include enough detail so that a reader could understand the story and retell it to someone else.

Figure 6.1: Questions in online survey instrument. The first and second questions are quality checks. The last question is the story-writing prompt, crafted to help maintain a sense of anonymity while facilitating the story-writing process.

6.1.1 Study Design

6.1.1.1 Instrument

To collect first-person accounts of incidents of unauthorized access, we designed a storytelling task, delivered in the form of a qualitative survey question. We asked participants to write free-form stories about past experiences through the survey instrument shown in Figure 6.1.

We emphasized that stories were anonymous. To that end, we did not ask participants to convey their role. Instead, we asked them to write stories as if they were narrators not involved in the incident. We also suggested they use a set of names we selected in advance:

- **Ash**, for the person whose device was accessed; and
- **Val**, for the person who accessed it.

We further suggested participants use gender-neutral pronouns, and asked them to refrain from including any personally-identifiable information.

We also offered some suggestions to facilitate the story-writing process. We suggested participants to include key elements of narrative, such as *when* and *where* the incident took place, the *relationship* between Ash and Val, *what* happened, and *why*. We indicated a good length threshold was having "enough detail so that a reader could understand the story and retell it to someone else." These suggestions, as well as our framing of participant's role as a "narrator", and of the subjects in the stories as "characters", was also intended to accentuate the storytelling device we wanted participants to employ. Asking for narratives of past events is a well-established method in many disciplines, including HCI and security. The approach we took can be understood as an application of the Critical Incident Technique (CIT) (Flanagan, 1954). Unlike what is common in applications of the CIT, which usually rely on direct first-person accounts, we instead asked for stories. Our intention was to provide more anonymity to participants, muting some of the social desirability bias associated with admitting to unauthorized access (see e.g. Chapters 4 and 5), while still gathering rich details. Story-writing methods have been noted to have the potential to gather qualitative data on sensitive topics more effectively than other approaches (Braun et al., 2017).

6.1.1.2 Procedure

We set up our instrument as an online survey, and deployed it to a private web server. To recruit participants, we used Prolific[1]. Like the better-known Amazon Mechanical Turk service, Prolific recruits people for online tasks, and mediates their compensation. Prolific, however, was specifically created to recruit participants for online research, and has been found to provide better-quality data (Peer et al., 2017). We also believe Prolific treats participants better, imposing compensation minimums and, in our experience, being active in preventing abuse.

Using Prolific's screening questions feature, we were able to only invite participants who had indicated having a prior experience with unauthorized access. Once participants accepted the invitation, they were informed of the researchers' contact information, the purposes of the study, and asked for consent in our use of their responses for research purposes.

[1] `https://prolific.ac/`

Data was collected in two stages. We first collected a set of 35 responses and inspected the stories to verify that our instrument was working as intended. We were satisfied that it did, and observed that we had gathered a diversity of considerably unique kinds of stories. To increase the chances of capturing as much variation as possible, within reasonable time and cost, we decided to continue collecting data until obtaining a total of 100 responses. We ended up collecting 115 responses, but, after inspecting them, we excluded 13 which were either empty, nonsensical or not relevant to the prompt. Our analysis draws from the remaining 102 stories.

6.1.2 Participants

Participants whose stories we used identified themselves as female 61 times, and as male 40 times. They reported their ages as 18–24 years in 31 instances, 25–44 in 63 instances, and 45–64 in 8 instances. About 75% of participants were from Europe, and about 25% from the US or Canada. Only three participants were from elsewhere. On average, participants took about nine minutes to complete the task, and the average story was 151 words-long. Participants were compensated at an average hourly rate of £11 (GBP).

6.1.3 Analysis

We analyzed the data in two steps. First, we engaged in exploratory and descriptive analysis of the qualitative data. We built a codebook, coded all stories, verified inter-rater reliability, and summarized the domains and codes we found. After completing this analysis, we were not entirely satisfied with how our codebook captured the richness of the data. We thus engaged in a second step, which was a thematic analysis (e.g., Terry et al., 2017) of participants' stories. We approached the process of data re-examination mainly through close reading. Since the data was already coded, and thus easy to subset, we could explore latent aspects of participants' experiences from several vantage points. Close reading is an analytical procedure associated with the social sciences and the humanities (for a discussion of humanistic approaches to HCI, see Bardzell and Bardzell (2016)). Our process was therefore reflexive. As a result, this analysis cannot be detached from the researchers who were involved in this process.

In the next two sections, we report on each of the two steps of analysis. In the first, we examine *what* happens in incidents of unauthorized access, and, in the second, we examine *how* incidents are represented by participants. More detail on our analysis process is provided at the beginning of each section.

6.2 Exploratory Analysis

Having collected 102 anonymous stories of people accessing the smartphones of people they know, we next explored what happened in these incidents. To understand the salient features of incidents of unauthorized access, we encoded essential elements of circumstances described in the stories.

We thus created a codebook, comprising of eight categories of codes, and coded each of the stories. To build the codebook, one researcher (the author of this document) inductively created codes from textual evidence in stories. Using this codebook, that researcher, and a second researcher (a collaborator), both coded a subset of ten stories. The researchers agreed on

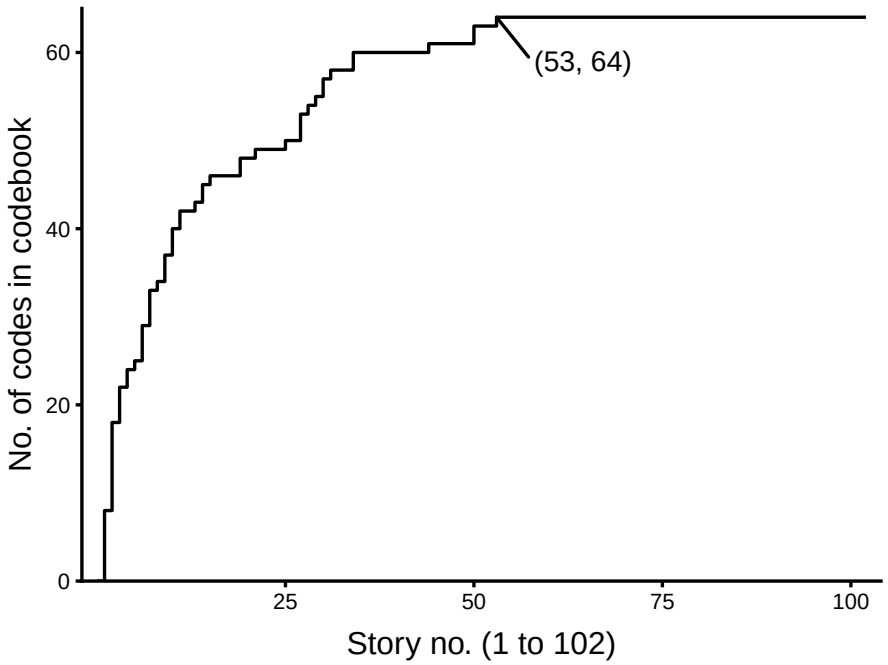

Figure 6.2: Step graph of codebook size. The vertical axis shows the cumulative number of codes in the codebook, and the horizontal axis shows the stories in which new codes were first attributed. In the subset of stories 1 through 53, all the 64 codes in the codebook were attributed at least once.

95% of coding decisions (Cohen's $\varkappa = 0.90$, z-score = 29.2, $p < 0.001$), indicating the coding was reliable. In the process of reaching consensus, we found most disagreements were lapses in code assignment by researchers. We resolved the remaining disagreements by disambiguating some code descriptions. The first researcher then re-coded all stories. Because inter-rater agreement had very little room to improve, we found it unnecessary to repeat the process of parallel rating and consensus-reaching.

In the codebook, we formulated code categories as questions about stories, such as *What was the primary motivation for unauthorized access?*; and formulated codes as possible answers, such as *Val wants to play a prank on Ash.* In six of the eight categories, questions called for classification, so we assigned, at most, one code per story. In the remaining two categories, questions called for enumeration, so we assigned as many codes as applicable. Categories are therefore dimensions of stories, and codes describe the variation within these dimensions.

Figure 6.2 shows a step chart of codebook growth per additional story. The final codebook had 64 codes. Story 53 of 102 was the last in which a new code was first attributed, indicating that the number of stories we collected was almost double what was needed to capture the variation in the dimensions we found.

The dimensions we captured describe a narrative chain, including the context in which incidents happened, the course of events, and the consequences. To capture context, we classified the **types of relationship** between Ash and Val, and Val's primary **motivation**. To capture the course of events, we classified how **opportunities** for access came about, how Val overcame the **lock** if it was set up, and enumerated Val's **actions** once they obtained access. Finally, to capture consequences, we classified whether and how Ash became **aware** of their device being accessed, enumerated expressions of **emotional aftermath** experienced by either party, and classified whether relationships **ended**. We next describe the codes we found.

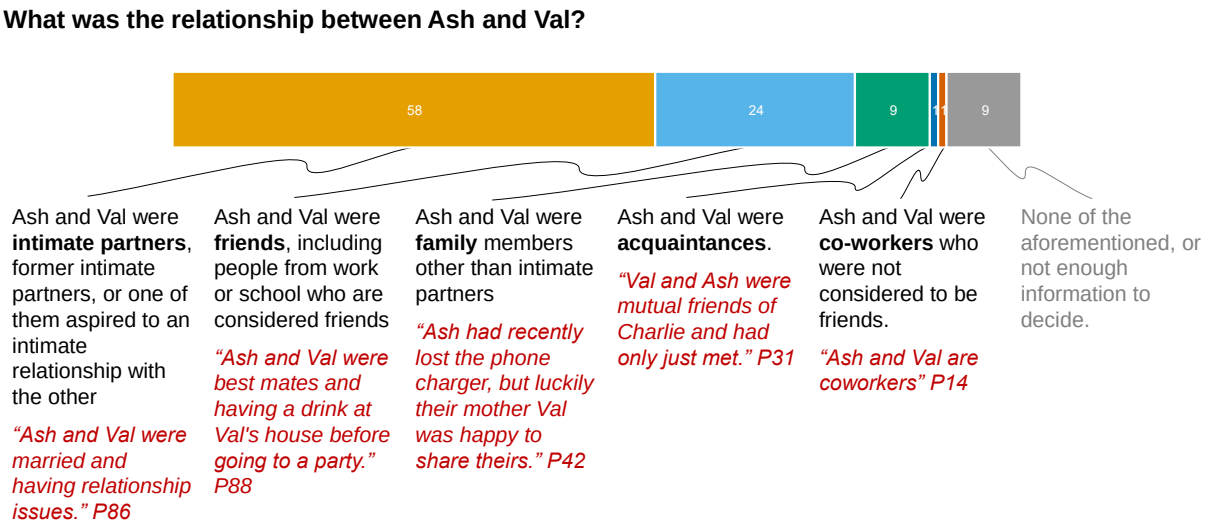

Figure 6.3: Distribution of types of relationship between parties, in 102 stories of unauthorized access to smartphones.

6.2.1 The Context Leading up to Incidents

6.2.1.1 Type of Relationship

We classified relationships between parties into five types: **intimate partners**, **friends**, **family** members (who are not intimate partners), **co-workers** (who are not also friends), and **acquaintances**. In the online survey, we prompted participants to write about incidents involving them and "someone they knew", without suggesting relationship types. The stories therefore reflect those relationships that participants judged to be non-strangers. We also suggested for participants to describe the nature of the relationship in their story. In all but 9 stories, participants provided enough evidence for us to classify relationships. Figure 6.3 shows the relative frequency of relationship types we identified. Two of the codes are outliers, appearing only once. These outliers were a story describing an attempt at unauthorized access by a co-worker, who was ultimately unable to unlock the device; and a story in which someone, by accident, accessed an acquaintance's smartphone of equal make and model to theirs. Despite these unique relationship types, we included these stories as they add diversity to our data.

Participants more often conveyed incidents involving people in an inner circle of close relationships. The outliers corresponded with the more distant types of relationships - acquaintances and co-workers. Even within the more common types of relationships in the data, upon closer inspection, we found patterns suggesting that most stories were associated with closer relationships. In the case of the largest type, intimate relationships, most stories described non-transient relationships. Stories signaled the non-transient nature of relationships with a combination of markers, including commitment labels (e.g., "married", "couple", "in a relationship"), indication of duration (e.g., "long-term relationship", "together for three years"), or reference to having children. We found the same pattern in stories describing incidents between friends: in most cases the relationships were qualified with markers of closeness (e.g., "best friends", "longtime friends", "childhood friends", "real friends"), or with reference to co-habitation. Relationships we coded as "family" only included very close ties: six parent-child relationships; two sibling relationships, and one pibling-child relationship.

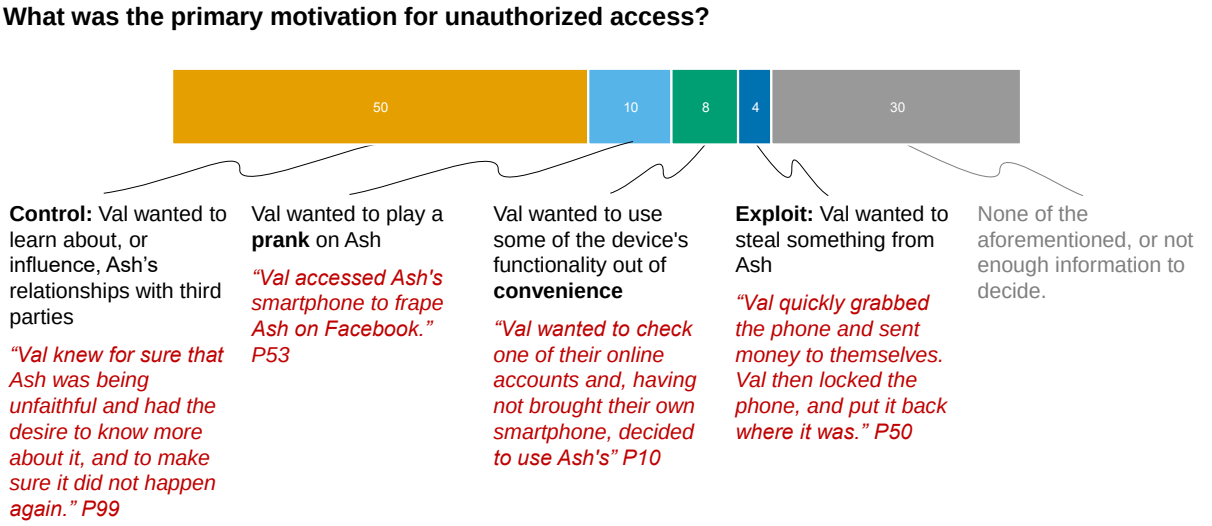

Figure 6.4: Distribution of types of motivation for unauthorized access, in 102 stories of unauthorized access to smartphones.

How can this pattern be explained? One possibility is that our sample represents a larger reality. In particular, the repetition of the pattern within subsets of data is consistent with unauthorized physical access being more common in close relationships. Our data is also consistent with previous observations that social proximity is associated with physical proximity, offering more opportunities for unauthorized access; and that socially-close people could be specially motivated to obtain access (e.g., Matthews et al., 2016; Mazurek et al., 2010; Muslukhov et al., 2013). However, the pattern can also be an artifact of our sample. Ours was a small, convenience sample, and the study was not designed to make quantitative generalizations. Another possible explanation for the pattern is that participants chose to recount the incidents that were significant to them. Our sense is that incidents involving people from an "inner circle" often carried a heavy emotional toll. Possibly, this made them easier to recall and reflect upon.

6.2.1.2 Motivation

We found four types of motivation for unauthorized access: to seek **control** over Ash's relationships with others, to pull a **prank** on Ash, to use some of the device's functionality for **convenience**, or to **exploit** access for personal (e.g., financial) gain. While we suggested participants to describe the motive for device access, we were not able to classify motivation in 30 stories. Figure 6.4 shows the relative frequency of motivation types we could identify from evidence in the text. Notably, in about two thirds of cases, unauthorized access was control-motivated.

The **control** code covered a wide range of incidents. We used a definition of seeking control which encompassed both surveillance and interference: "Val wants to learn about, or influence, Ash's relationships with third parties". The code was initially based on Stark's *coercive control* framework of intimate partner abuse (Stark, 2007), which constructs controlling behaviors, rather than episodes of violence, as markers of abuse. This framework has been previously used in investigating technology-mediated abuse between intimate partners (Woodlock, 2017). In our data, controlling behaviors were abundant. Many stories featured incidents between intimate partners, in which one party sought to verify compliance with expectations of monogamy, and

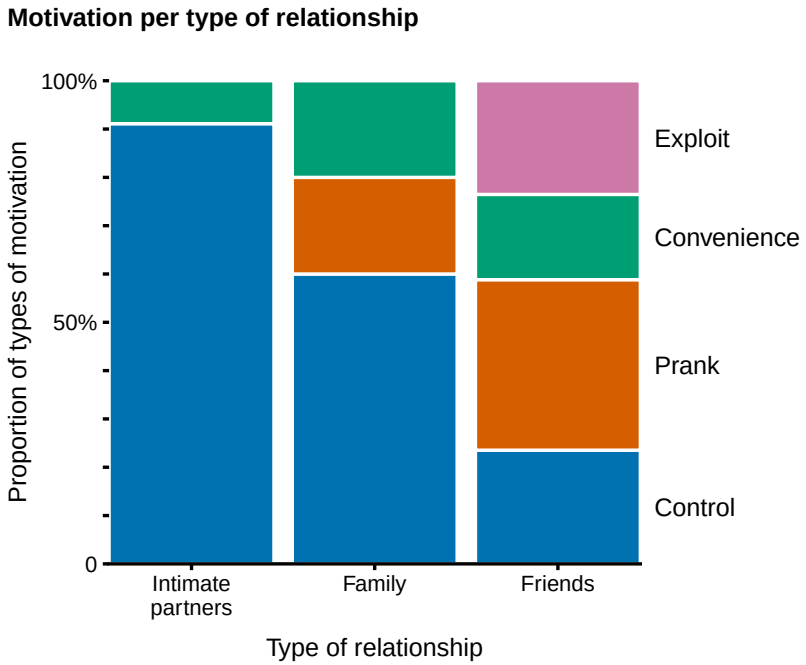

Figure 6.5: Relative distribution of types of motivation for unauthorized access per types of relationship between parties, in a subset of 102 stories of unauthorized access to smartphones for which both codes were attributed.

sometimes punish perceived infractions. However, since the code definition was merely descriptive of an intent, it also applied to other stories. For instance, there were stories describing incidents in which friends, or family members like parents, sought knowledge or influence over relationships with third parties.

The codes **prank** and **convenience** were used for stories featuring individuals seeking access to play pranks, or to use some of the device's functionality for practical purposes, respectively. Of the stories in which we could classify a motive, around one quarter were pranks or convenience-motivated access. The existence of such stories in the data suggests that at least some participants understood they could write stories about any experiences of unauthorized access, including those not involving stigmatizing behaviors.

We only classified four stories with the **exploit** code. The four stories are, however, unique: they portray a range of ways in which stealing of valued possessions — a concern often more associated with strangers (e.g., Muslukhov et al., 2013) — is sometimes sought by individuals known to each other. Three of those stories describe people exploiting unauthorized access to benefit financially — in one story by stealing a device, in another by stealing business contacts, and in the third by transferring currency out of a digital account. In the remaining story unauthorized access was a means to steal sexualized media.

The stories participants provided indicate a connection between the relationship type and the motivations for unauthorized access. Figure 6.5 shows the proportion of classified motivations in relation to the relationship type (excluding the two outliers). We found just two motivation types in stories involving intimate partners: convenience, and control. However, control-motivated unauthorized access was overwhelmingly prevalent. Among family members, the control motive was also prevalent, but playing pranks or convenience were also typical. Among friends, we found

all four types of motivation. Exploitation for personal gain occurred exclusively among friends.

The stories also indicate that people access smartphones without permission for many reasons. Some, such as stealing money or data, are clearly nefarious. Some, such as playing pranks or accessing a device for convenience, lean towards being benign. Control-motivated access was, however, often more difficult for us to judge as to its nefariousness. Participants, as prompted, most often described distinct episodes of unauthorized access, not sustained patterns of behavior which could be markers of abusive relationships. Furthermore, equal behaviors can be considered acceptable or not by parties depending on context (e.g., Burke et al., 2011). In exceptional cases, however, participants did describe what was unequivocally abuse. In our data, these cases appeared predominantly in stories in which parties were not intimate partners at the time of the incident. For instance, in one story, they had "just ended their relationship", yet Val, after accessing Ash's device, turned verbally abusive and threatening; and, in another, Val is described as aspiring to an intimate relationship, but the perception of rejection leads to bullying and harassment. A more rigorous examination of these matters can be found in the growing body of literature on the role of technology in intimate partner abuse (Burke et al., 2011; Dimond et al., 2011; Freed et al., 2018, 2017; Leitão, 2019; Matthews et al., 2017; Woodlock, 2017). Our analysis lends support to prior observations that unauthorized access can be a component of intimate partner abuse, but indicates a wider range of relationships, and relationship states, in which unauthorized access occurs.

6.2.2 How Events Unfolded

6.2.2.1 Opportunity

We classified how opportunities for unauthorized access came about with three codes, referring to situations in which devices were left **unattended**; situations in which access was obtained through **secondary devices** (i.e., not Ash's current smartphone); and situations in which the person accessing the device used **deception**. Having had suggested that participants provided details about how one person was "able to get access" to the other's device, we were able to classify opportunity in 85 stories from explicit evidence. Figure 6.6 shows the distribution of types of opportunities for unauthorized access that we could identify in the stories. Overwhelmingly, stories indicated that, when devices were accessed, they had been left unattended. We saw few stories with unauthorized access through secondary devices or through deception. The secondary devices mentioned in the stories were a tablet that was synced with a primary smartphone, and a smartphone that had not been reset after the owner started using a new one. The one case of deception refers to a story in which a person asked for access to "check something on the internet" and then accessed a social media account. Although these stories were outliers, we found that they provided diversity and mostly matched what was asked of participants.

When there was enough detail in the stories, we further classified cases of devices being left unattended into four notable sets of circumstances. Stories commonly indicated devices had been left unattended while their owners went to the **bathroom** (for instance, to take a shower); while they were **asleep**; and while they went **outside** of their homes (for instance, for shopping or going to class). We found one case of a device being left unattended at work, while the owner was attending a **meeting**. Noticeably, in all these circumstances, devices had been left unattended in locations often deemed trusted by some security software, such as homes or

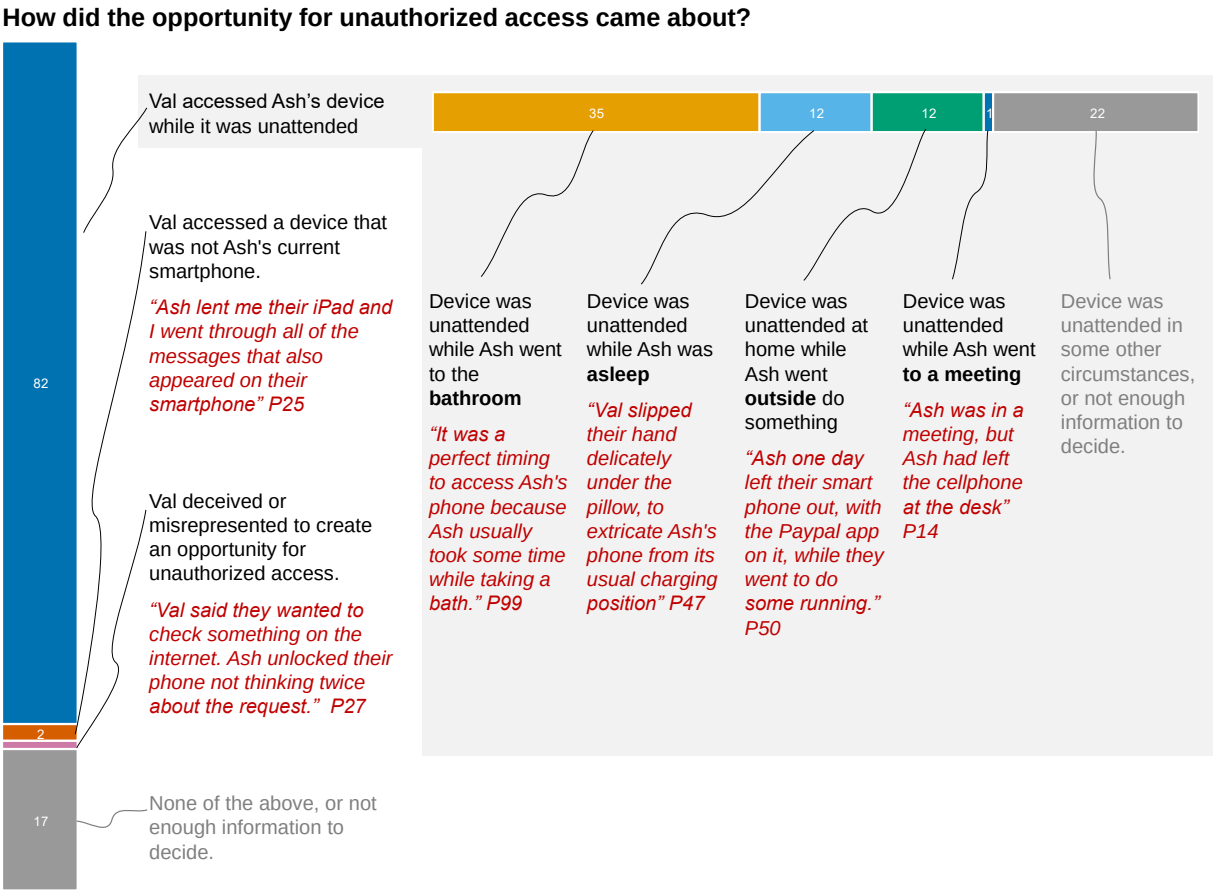

Figure 6.6: Distribution of circumstances in which devices were accessed, in 102 stories of unauthorized access to smartphones.

workplaces. For instance, Android's Smart Lock (Google, 2018) actively suggests users add their home's location to a set of trusted places where unlocking is required less often.

6.2.2.2 Locks

We found that a considerable number of stories referenced smartphone locks. Figure 6.7 shows the distribution of such occurrences. In 61 of the stories, we found explicit references to either **locks not being set up**, or to **people overcoming locks** that were set up. We encountered four notable ways in which locks, despite being set up, were ineffective in preventing unauthorized access. Most commonly, stories indicated that authentication **codes were already known**, for instance because they had been willingly shared previously. Sharing smartphone authentication secrets is a common behavior in interpersonal relationships (e.g., Cherapau et al., 2015; Egelman et al., 2014; Harbach et al., 2014; Matthews et al., 2016). In other cases, authentication **codes were discovered through visual observation**. Visual observation, or "shoulder-surfing", is another well-documented vector for unauthorized physical access (e.g., Eiband et al., 2017; Harbach et al., 2014). We also found stories in which characters were described as having **guessed authentication codes**; and stories in which **locks were set up but not active at the time** of unauthorized access, for instance because devices were not inactive long enough to lock.

Participants seemed to perceive smartphone locks as a key element in preventing unau-

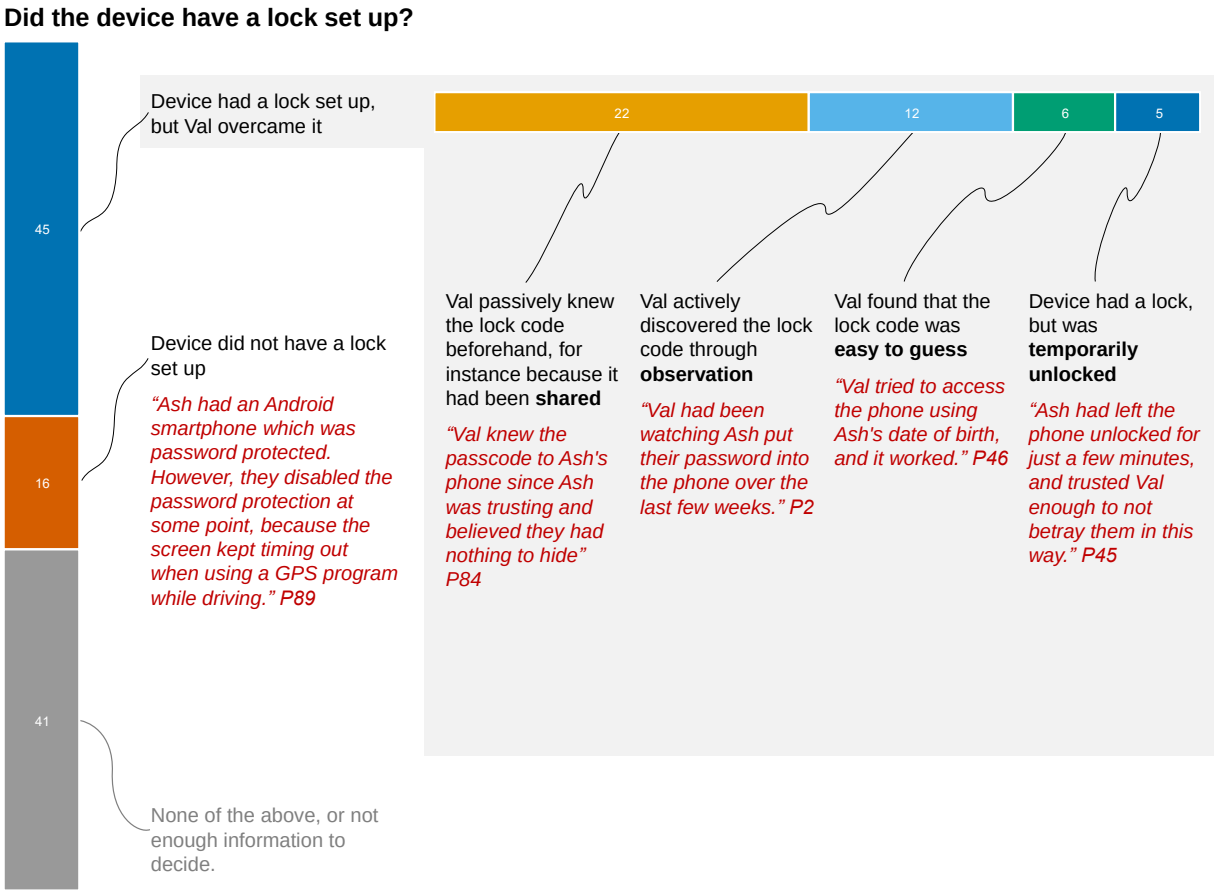

Figure 6.7: Distribution of the role of authentication locks, in 102 stories of unauthorized access to smartphones.

thorized access. We prompted participants to include information about how Val was "able to get access", but did not reference locks. The fact that stories provide such level of detail on locks suggests that participants considered them to be relevant for preventing access by known people, as previous work has documented (e.g., Cherapau et al., 2015; Egelman et al., 2014). Our data does not contradict that, absent smartphone locks, unauthorized access by known people would be even more common. In fact, upon closer inspection, we found five stories in which Ash counteracts the possibility of future incidents by setting up a lock or changing the authentication code. In these five stories, either the motivation for the incident had been to play a prank, or Ash was defending the device against family members. Changing of locks was not mentioned in other stories possibly because it was not seen as an effective strategy in other circumstances, such as control-motivated access among intimate partners.

6.2.2.3 Actions

For each story, we enumerated the actions performed by the person accessing the device. We categorized actions into four types: **gathering information** by visual inspection, **tampering** with devices by making changes to their state which are not easily observable, **impersonating** the device owner, and **exfiltrating** data. We further categorized actions by their object. After combining objects of actions with types of action, we ended up with 21 codes. We had suggested that participants included details about what active parties did after obtaining

What did Val do once they gained access?	Frequency
Information gathering:	
Val inspected archives of non-public conversations in text form, such as text messages, emails, instant messages, or chats.	61
Val inspected archives of visual media, such as photo galleries.	9
Val inspected social media activity.	8
Val inspected contents associated with device notifications, such previews of new text messages.	5
Val inspected lists of contacts.	4
Val inspected the device's call log.	3
Val inspected which apps were installed on the device.	2
Val inspected records of internet activity, such as internet searches and internet browsing history.	2
Val inspected non-specified data which was on the device.	2
Val inspected the calendar.	1
Tampering:	
Val captured new visual media with the device, for instance, taking photos.	5
Val deleted contents from the device.	3
Val changed contact records.	1
Val implanted surveillance software.	1
Val changed settings on device, including settings internal to apps.	1
Impersonation:	
Val impersonated the Ash on social media activity.	4
Val impersonated the Ash in non-public textual conversations.	3
Val impersonated the Ash in financial / banking services.	1
Exfiltration:	
Val extracted contact information to another medium.	1
Val extracted contents of non-public textual conversations to another medium.	1
Val extracted visual media, such as photos, to another medium.	1
	119 codings

Table 6.1: Frequency of actions executed by the person accessing the device, in 102 stories of unauthorized access to smartphones.

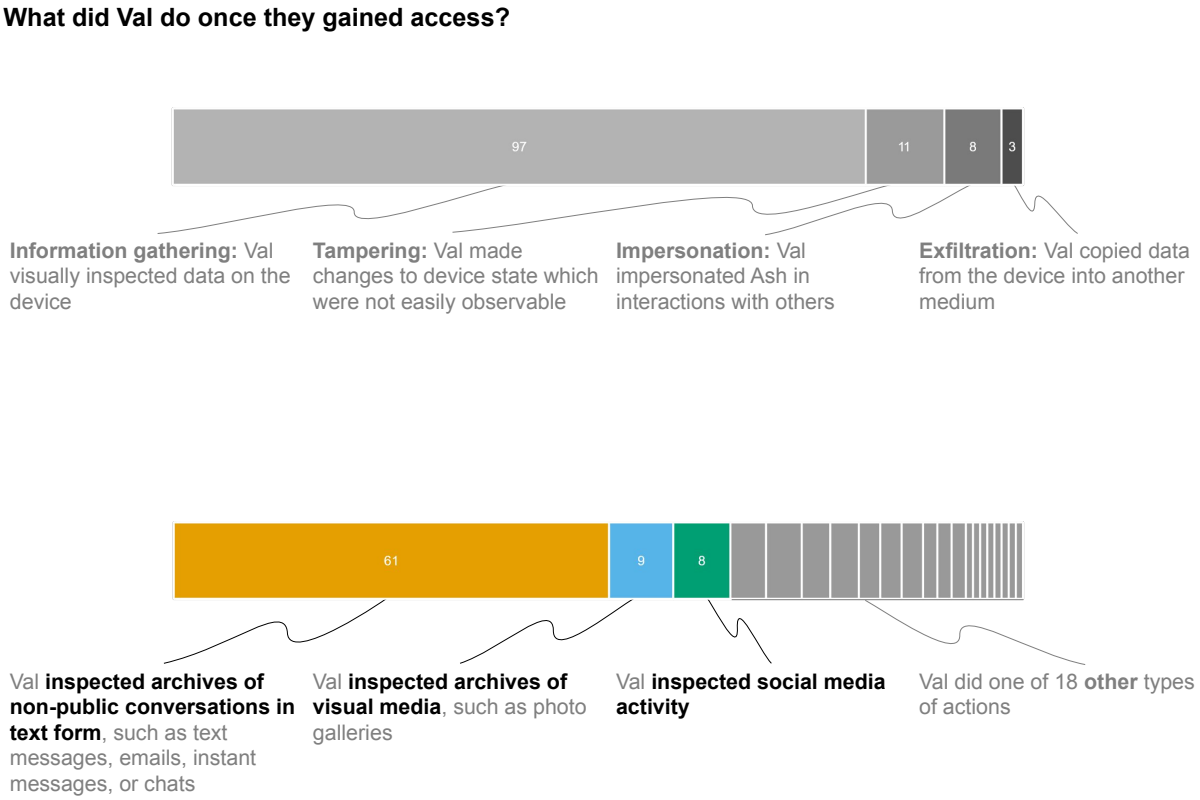

Figure 6.8: Distribution of actions executed by the person accessing the device, in 102 stories of unauthorized access to smartphones. At the top, distribution of four categories of actions; at the bottom, the most common actions.

access and, in all but 11 stories, we found direct evidence to attribute at least one code. Since some stories described more than one action, we attributed 119 codings to the remaining 91 stories.

Figure 6.8 shows the distribution of the actions we found. At the top, it shows the frequency of the four categories of actions; and, at the bottom, the most common combinations of actions with objects of action. Table 6.1 shows the frequencies of all action-object combinations, including those omitted in the figure.

Most commonly, we found stories to provide evidence of **information gathering** (77/102 stories). The most common object of information gathering we found was **text-based conversations**, such as text messages, instant messages, or emails. Inspection of text-based conversations was so prevalent that it appeared in the majority of stories (61/102), and the code was attributed about as many times as all the remaining 20 codes combined (61/119 code attributions). The only two other codes that we attributed more than five times also concerned information gathering. We found nine stories describing inspection of **media files** such as photos; and eight stories describing inspection **social media activity** other than conversations (e.g., posts). Occurring with less frequency, we found stories indicating the person who accessed the device inspected notifications, contacts, call logs, internet history, apps installed, and calendars. The diversity of objects of information gathering that we encountered largely coincides with types of data smartphones users have described as sensitive in prior research (e.g., Ben-Asher et al., 2011;

Felt et al., 2012; Hang et al., 2012; Hayashi et al., 2012; Karlson et al., 2009; Mazurek et al., 2010; Muslukhov et al., 2012, 2013). Previous research has also called attention to smartphones having a particular status as to their sensitivity (e.g., Chin et al., 2012; Dimond et al., 2011). Part of the reason may be a combination of smartphones being more heavily used for personal communication than other devices (see e.g., Müller et al., 2015), and users valuing personal communications more than other digital assets (see e.g., Muslukhov et al., 2013; Shay et al., 2014). With the caveat that our sample may not be representative, some of the data users deem as most sensitive, seems to coincide with the data most targeted for information gathering.

Although less frequently, we also found several instances of tampering, impersonation, and exfiltration of data. Stories described **tampering** with devices by changing settings, changing contact records, deleting contents, installing spyware, and capturing new photos; they described **impersonation** in social media, in text-based conversations, and in financial services; and they described **exfiltration** of photos, records of conversations, and contacts. Similar behaviors have been previously observed, for instance, in studies of the role of technology in intimate partner abuse (see e.g., Burke et al., 2011; Dimond et al., 2011; Woodlock, 2017). However, in the stories we collected, tampering, impersonation, and exfiltration were not always associated with control-motivated unauthorized access between intimate partners. We found instances of tampering in prank- or convenience-motivated incidents; instances of impersonation in prank- and exploit-motivated incidents; and instances of exfiltration in exploit-motivated incidents. This diversity is consistent with our earlier observation that behaviors associated with intimate partner abuse also occur in a wider spectrum of circumstances.

6.2.3 Consequences

6.2.3.1 Awareness

We suggested participants included detail on how, if at all, the person whose phone was accessed learned about it. As shown in Figure 6.9, in 22 stories, we found there was evidence indicating that people did not become aware of their phones being accessed; and in 61, that they did become aware. We further classified how people became aware, and found stories to describe three ways: by finding clues leading to a suspicion of unauthorized access, such as unusual device behaviors, or things said by the other person; by unprompted own admission, for instance, by confronting the device owner; or by encountering another in the act of accessing the device.

6.2.3.2 Emotional Aftermath

We also suggested that participants included details about “any consequences”. From the evidence provided in stories, we enumerated expressions of positive or negative sentiments resulting from incidents of unauthorized access. Positive sentiments included amusement, satisfaction, or relief; negative sentiments included annoyance, anger, guilt, humiliation, pain, regret, sadness, or shame. Figure 6.10 shows the distribution of these sentiments. We found negative sentiments to be expressed more often than positive sentiments. Negative sentiments were more prevalent than positive sentiments for either the person accessing the device, or the person whose device was accessed.

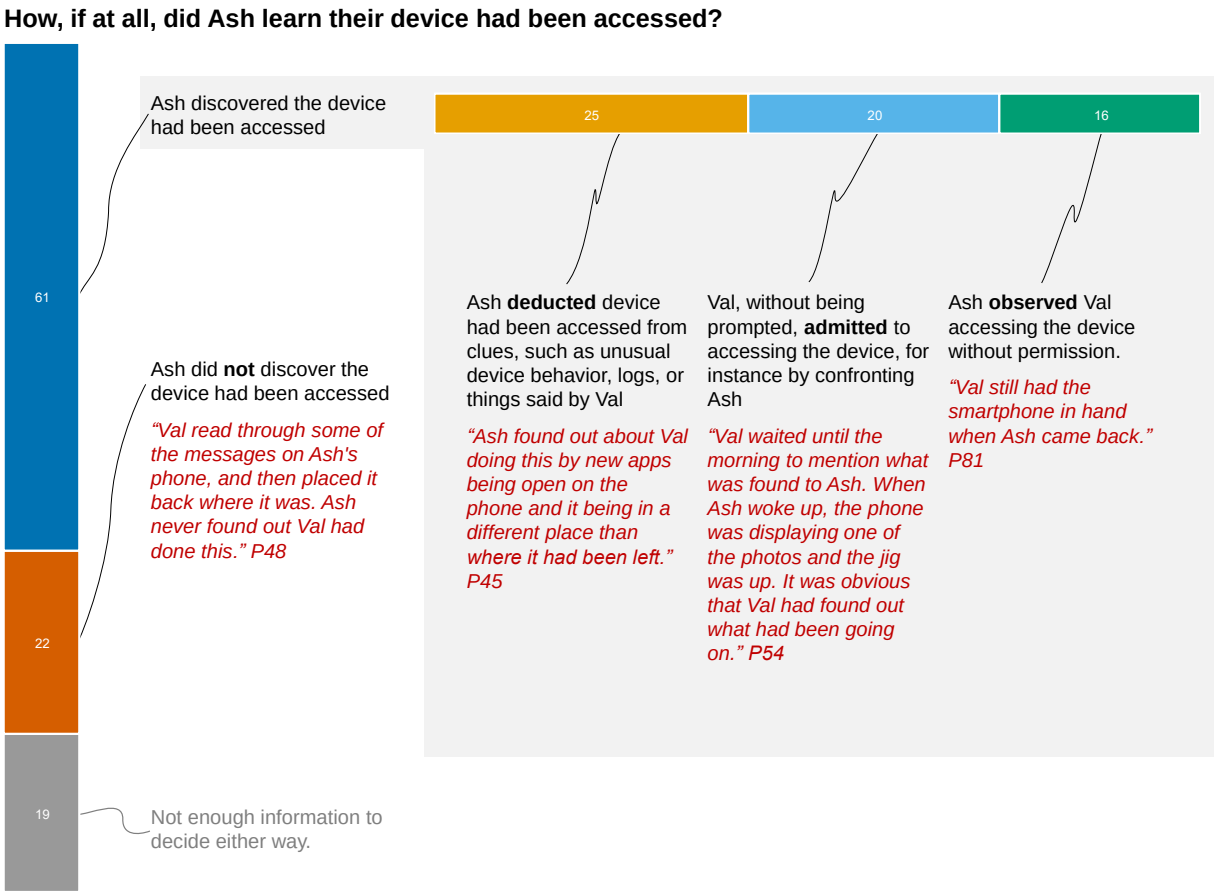

Figure 6.9: Distribution of the awareness status on the part of the person whose device was accessed, in 102 stories of unauthorized access to smartphones.

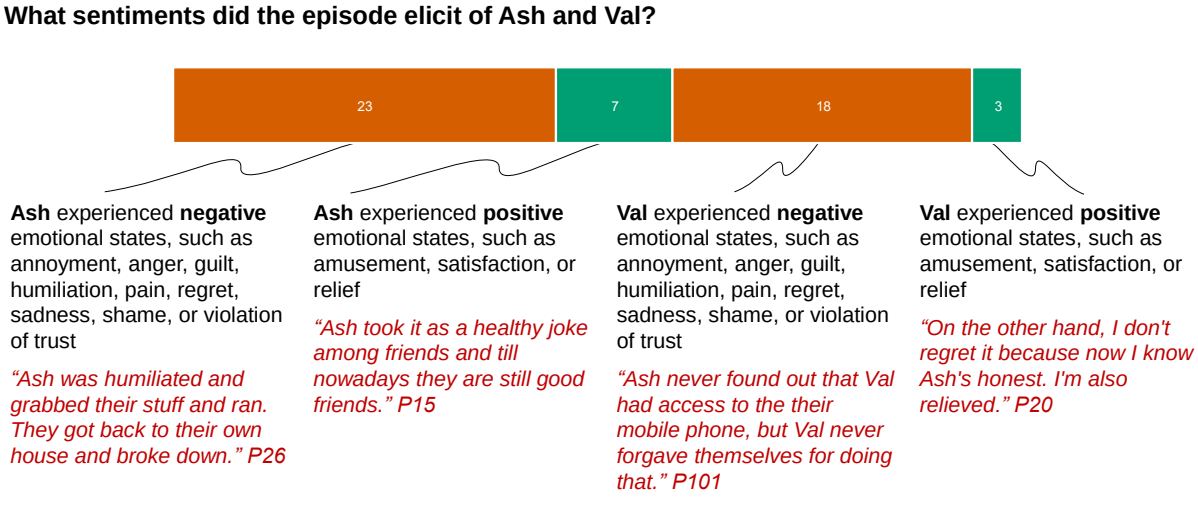

Figure 6.10: Distribution of explicitly-stated sentiments attributed to parties, in 102 stories of unauthorized access to smartphones.

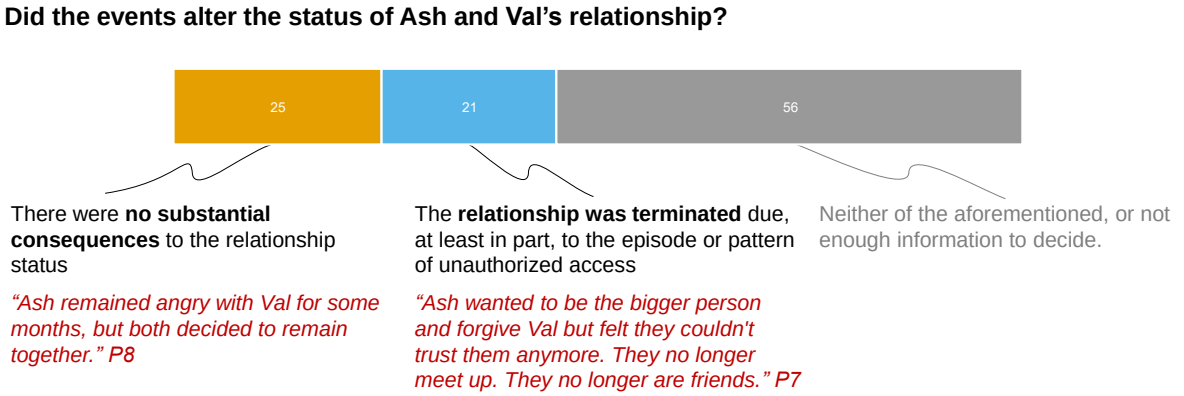

Figure 6.11: Distribution of relationship status outcomes, in 102 stories of unauthorized access to smartphones.

6.2.3.3 Relationship Termination

One consequence of incidents of unauthorized access that was referenced in some stories was the ending of relationships. Many stories did not provide enough direct evidence to classify relationship outcomes. As illustrated in Figure 6.11, in the stories that did provide enough detail, we found 21 stories indicated relationships had ended at least in part due to incidents of unauthorized access, and 25 stories indicating relationships had persisted.

In comparison to codes describing the context of incidents or the course of action, we found the codes for consequences to provide much less insight into participants' experiences. Participants often emphasized how consequential incidents of unauthorized access had been their lives. However, we could not capture that richness with a coding process that required direct and unambiguous evidence in the text of stories. Stories indicated an array of consequences that could not be captured by relationships having ended or not, nor by sentiments being explicitly positive or negative. There were relationships which did not end, but their persistence was painful. There were relationships which ended, but were eventually mended and made stronger. Participants sometimes described reactions to incidents which implied strong emotional states, but did not describe precise sentiments — the reactions spoke for themselves.

The qualitative analysis we started with, and described in this section, was informative in important aspects of participants ' experiences, but was insufficient to capture consequences. To address this limitation, we engaged in a second, more reflexive, type of analysis, which we discuss next.

6.3 Thematic Analysis

To offer a more rigorous account of participants' experiences with unauthorized access, we turned to thematic analysis. The codes we used in the previous section, based on direct assertions in the text, are *semantic codes*. Semantic codes are believed to be unsuitable for capturing latent meanings in qualitative data (Terry et al., 2017). Our data called for a more

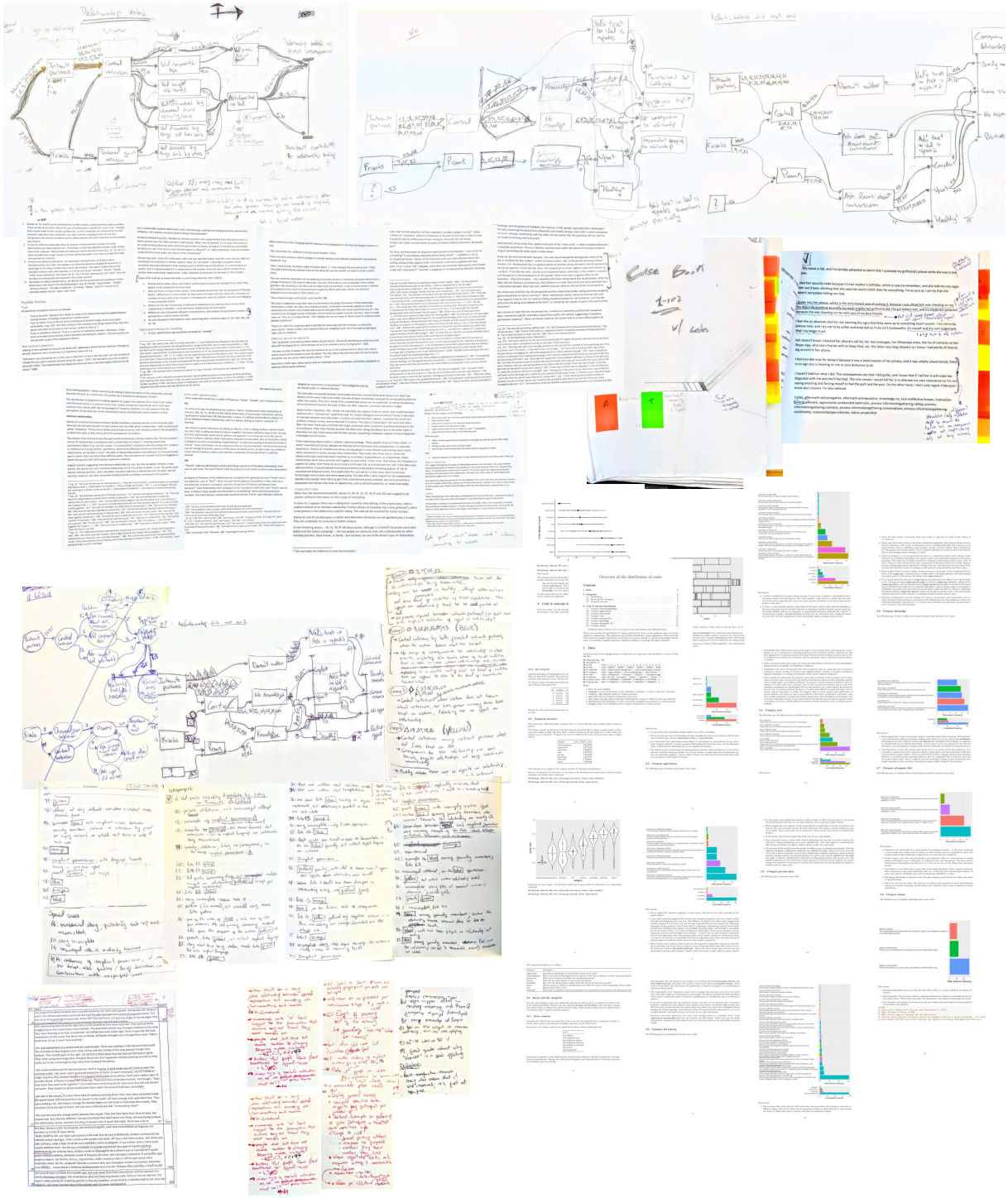

Figure 6.12: Collage of some of the media employed in the close reading of stories aimed at developing themes.

reflexive approach.

To understand how participants made sense of their experiences, we turned our attention to *how* they described them. We developed two themes to capture key aspects of participants' construction of their experiences. First, participants understood trust as performative vulnerability: trust was necessary to sustain relationships, but building trust required displaying vulnerability to breaches. Second, participants were self-serving in their sensemaking: they blamed the circumstances, or the other person's shortcomings, but rarely themselves.

Our process for developing these themes was inductive. We re-engaged with the data in multiple rounds of close reading. In each round, we used categories of semantic codes as lenses to look at the data. For instance, in the first round, we used the lenses from the *relationship type* code category, and closely read all stories with a focus on how relationships are represented, and how these relate to how incidents are experienced. In this process, we marked-up text, drafted thematic maps, collected quotes, and articulated patterns in written notes. Figure 6.12 depicts the variety of text engagement devices we employed, in the form of a collage. Gradually, we distilled our analysis into two organizing themes which, to our satisfaction, conveyed some of what was missing in our code-based analysis.

We next lay out these two themes, illustrating them with quotes from the stories. We lightly edited the quotes to make them easier to read, and to elide gender or other information that could de-anonymize the stories. The names of characters in quotes follow the convention we suggested to participants: **Ash** refers to the owner of the device, and **Val** refers to the person who accessed Ash's device without permission.

6.3.1 Trust as Performative Vulnerability

Central to participant's experiences of unauthorized access was seeing expectations of trust, which they believed were binding, being violated. Many stories conveyed a belief that mutual trust was not only desirable, but necessary to maintain relationships. However, to maintain trustworthiness, participants had to make themselves vulnerable to violations. This rationale is vividly illustrated in two of the stories of control-motivated unauthorized access among intimate partners, told from opposing perspectives:

> "Ash had nothing to hide but feared not being trusted if they kept their phone with them at all times" – S43

> "Val was suspicious. Ash would take their smartphone everywhere including when they were showering. Ash would turn their smartphone off if they had to leave it in a room with Val." – S75

In these stories, Ash not displaying vulnerability was detrimental to their trustworthiness, which was reciprocated by Val accessing Ash's smartphone without permission. Participants' representation of trust evoked other conceptions of trust rooted in vulnerability. In a review of trust development, Lewicki et al. (2006) distinguish a "psychological tradition", wherein trust is understood as one's willingness to accept vulnerability, conditioned on positive (or at least neutral) expectations of another's conduct. Trust as a marker of relationship health also frequently comes up in empirical work on privacy and security attitudes towards known people (e.g., Matthews et al., 2016; Mazurek et al., 2010; Park et al., 2018). For instance, a recent study of account sharing among intimate partners found that one common explanation for sharing was

a feeling that trust was necessary in relationships (Park et al., 2018). However, in the stories we collected, it was not enough to be vulnerable. People had to overtly display vulnerability, by very visibly taking on risks. Performatively taking on risks could mean to not visibly engage in risk-averting behaviors, such as in the case of S43. The alternative, of engaging in risk-averting behaviors, such as in the case of S75, could have been interpreted as meaning Ash was not trustworthy, which in turn revealed that the relationship was in peril.

The corollary to this conception of trust it that unauthorized access by someone close is not experienced as a security issue. Security issues could perhaps be fixed with stricter security regimens. Instead, the prevailing experience of unauthorized access was one of breach of trust, and hence existentially consequential to relationships. Participants' perceptions were that when the vulnerability they displayed was abused, changing a lock code was hardly a solution – instead, there *had* to be consequences for the relationship. This imperative is sometimes represented as a lack of rationale for the consequences, such as in these examples:

> "Ash discovered what had been done to their phone from unusual battery consumption. It was the end of their relationship." – S1

> "Ash found out about what Val did by new apps being open, and the phone being in a different place. Consequentially, Ash and Val are no longer roommates, and do no longer talk." – S45

In both stories, device owners terminated relationships immediately upon finding out that their devices had been accessed. Notably, the narrator does not find it necessary to articulate a rationale, such as how one party felt about the other's behavior, or what factors they weighted in making a decision regarding the future of the relationship. The causal link was so obvious to them that including it in the story would indicate a choice, when one was not understood to exist.

Through the same mechanics, unauthorized access could also benefit relationships. When displays of vulnerability were reciprocated with actions perceived by owners as not violating expectations, and instead being benign, relationships were strengthened. We saw that pattern in some episodes among intimate partners, in which the person accessing the phone used it for practical tasks: for instance, in story 12, where the phone is accessed while the owner is showering to facilitate planning a gathering with other people; or in story 44, where the phone is accessed to check the calendar for an open date for a surprise party. We also saw that pattern in some of the stories describing pranks. As long as an invisible line was not crossed, pranks served to build trustworthiness. Whether in stories of beneficial access or pranks, these episodes are portrayed as illustrations of well-functioning relationships.

In most of our data, displaying vulnerability, by taking risks with unauthorized access, seemed to be more of a choice than an obligation. That is not always the case. Research on technology-mediated intimate partner abuse has noted that taking such risks is often needed for personal safety (e.g., Matthews et al., 2017; Woodlock, 2017). Research on privacy-enhancing practices in non-Western geographies also indicates there are expectations of openness affecting women, which make taking risks more of an obligation (Sambasivan et al., 2018). Taking the patterns we saw in our data, and considering other accounts of risk-taking, the reasons for displaying vulnerability can be understood as existing in a spectrum. To what extent risk-taking is a choice or an obligation may be unclear, both to us and to those conveying their experiences.

Nonetheless, it appears that, ultimately, the ability to display vulnerability is understood by users, in their social and cultural contexts, to be a *requirement* which the technology must afford.

6.3.2 Self-Serving Sensemaking

Stories conveyed a stark pattern of attribution: when told from Ash's perspective, they blamed Val's intrinsic traits; when told from Val's perspective, they blamed the situation. With very few exceptions, stories were charitable to the narrator.

When told from Ash's perspective, strong statements assigning negative character traits to Val were common. A commonly assigned negative trait was being "jealous"; other related character flaws included:

"[being] the controlling type" – S2

"[being] quite possessive" – S5

"[being] a lunatic" – S69

"[having a] mind [which] works in a suspicious manner" – S40

When stories were told from Val's perspective, situational factors were invoked. Commonly, anomalous events, or a change in behavior, were portrayed as valid justifications for unauthorized access, such as in these examples:

"Ash's smartphone received a notification from a person Val did not like" – S51

"Val caught Ash in their bedroom talking on the phone at 3AM" – S53

"Val was worried because Ash received many texts in the last days" – S101

"Val started to think about how Ash had seemed distant lately" – S37

"They had been arguing more and more" – S47

The pattern of self-serving attribution, and the fact that it was so pronounced, indicates that incidents were experienced as significant episodes. Similar patterns of self-serving attribution have been found, for instance, when people describe experiences of being angered by someone else, versus them angering others (e.g., Baumeister et al., 1990; Kearns and Fincham, 2005; Zechmeister and Romero, 2002). In our data, the pattern of attribution is also consistent. Although it is most pronounced in stories of control-motivated intrusions, we saw it in many kinds of stories. For instance, in stories about pranks, expressions of negative emotional consequences were concentrated in stories told from the perspective of the targets of pranks. In stories told from the perspective of parents accessing their children's phones, the parent's actions are almost always represented as arising from an obligation to carry out protective responsibilities. Only in the one story told from the perspective of the child is that justification called into question: the parent is called out for meddling in the child's private affairs.

Participants also described forgiving transgressions, and mending their relationships. Previous research suggests that forgiveness is associated with a reduction in self-serving attributions (Zechmeister and Romero, 2002). We found an echo of that phenomenon in our participants' sensemaking. When stories were told from Ash's perspective, but relationships survived violations of trust, stories tended to not associate severely negative traits with Val. One common way to

minimize incidents was to note that the relationship was still nascent. Another strategy was to normalize access to devices as part of trust display, as in story 93:

> "Ash was a little hurt at the lack of trust but decided to forgive Val quickly. Ash now tries to let Val be more involved in Ash's smartphone activity so Val doesn't feel so anxious."

Similarly, when stories were told from Val's point of view and there were no long-term repercussions to the episode, situational explanations were muted. However, stories also avoided assigning strong negative traits to Val. To reconcile the lack of either situational or character explanations, stories typically expressed that unauthorized access had not been motivated by nefarious reasons, just "curiosity". Constructing Val as "nosey" (S6), "intrigued" (S35), or acting out of "boredom" (S64) avoided further blame attribution.

The few exceptions to self-serving attributions were also insightful. It was in these stories that we found most self-reflection on the narrator's own shortcomings, such as in the following stories:

> "I'm terribly ashamed. Ash didn't do anything to justify my mistrust. My last partner did and it has made me paranoid. I feel horrible now for doing it because it was a total invasion of Ash's privacy, and it was utterly unwarranted. The only reason I would now tell Ash would be to alleviate my own conscience. So I'm not saying anything, I'm forcing myself to feel the guilt and the pain." – S20

> "In reality, Val was experiencing some low self-esteem issues. Val wasn't aware of it until now. It was a hard journey to learn this fact." – S37

> "Some would try for fame and glory; others just like to watch the world." – DJ

When there was self-reflection, the significance of incidents of unauthorized access came into full display. For those who had accessed smartphones without permission, the emotional toll of dealing with their actions could be substantial. Recognizing that they had violated expectations of trust also meant that they had put the relationship at peril.

6.4 Conclusion

6.4.1 Summary

Our exploration of how people make sense of their experiences of unauthorized access portrays these incidents as personally significant, sometimes with severe consequences, and deeply entwined with interpersonal trust arrangements.

Through our exploration of stories of unauthorized access, we have provided finer-grained details on the diversity of circumstances involved in these kinds of incidents. Furthermore, we have advanced a framework to reason about how people's conceptions of interpersonal trust interact with security practices and user-facing security technologies. And, we observed how self-serving rationalizations from participants can offer a window into sensitive topics related to security.

6.4.2 Limitations

The methods we employed to address the question we set out to answer have some limitations, which we have pointed out throughout this chapter. We next highlight three significant limitations.

First, we asked participants to remember and write about past experiences. The experiences we collected are thus not a representative sample of experiences participants had, but of experiences which were salient to them. Furthermore, the set of participants who chose to take part in our study is also not a representative sample of a larger population.

Second, by approaching our analysis qualitatively, our findings are explicitly imbued with our frames of reference. Our combined backgrounds, previous knowledge, styles, and other factors, permeate every aspect of this research, from how we designed a data collection instrument, to how we built the codebook, to how we explored semantic codes, and to how we selected cross-cutting themes.

Third, the existence of a pattern of attribution of blame (see 6.3.2) suggests a possible fragility in our data collection method. We had asked participants to provide anonymous stories and, yet, we could often discern which story character participants identified with. Since participants experienced incidents from a particular perspective, the narrator's description could only provide insight from that perspective. With the benefit of hindsight, we cannot exclude that we could have collected richer accounts had we asked for direct first-person descriptions of incidents instead of stories. However, we expect that, having at least some plausible deniability, participants felt they could be more forthcoming in their writing. Furthermore, asking for stories, it seemed to us, encouraged participants not only to describe, but to also reflect on their experiences.

7 Conclusion

7.1 Summary

When we started this research, a shift was underway. It was clear that the new personal computing devices, and the smartphone in particular, were increasingly becoming extensions of ourselves, facilitating, but also recording, our interactions with the world. Then, as now, concerns over how this new normal could inflict damage on people's privacy abounded. Something else was also becoming clear: when users expressed what privacy issues afflicted them, it was not only thieves, hackers, governments, and corporations – the usual suspects – that populated their thoughts. People had started to realize that, like never before, their use of devices had created a situation in which those around them could steal from them, impersonate them, and surveil them; that these people could do so easily because of their physical and social proximity; and themselves did not know how to, or could not, do much about it. Researchers started framing a parallel between the worries of smartphone users and the most fraught class of adversary in computer security: the insider.

Since then, there has been a growing awareness of privacy concerns in relation to what we have called social insiders. Through our research, we have contributed to a growing body of knowledge, which speaks to emerging social insider risks. Our focus was on gathering an understanding of social insider intrusions that could substantiate technology design choices.

Our approach was rooted in understanding intrusions as security incidents. We took the intrusion, the moment in which user concerns were realized, as the observational unit. That unit, we reasoned, would enable a close inspection of how the emerging reality challenged the current design of systems on personal computing devices. We thus sought to characterize intrusions both quantitatively and qualitatively.

Our two main contributions to the understanding of social insider intrusions can be summarized as follows:

1. **Social insider intrusions are common occurrences.** For the two kinds of social insider intrusions we quantified, namely what we referred to as "snooping" (Chapter 4) and "facejacking" (Chapter 5), we estimated them to have been perpetrated by 31% of participants, and 24% of participants, respectively, in two large online list experiment studies (n = 1,381, and n = 1,308).

2. **Social insider intrusions are significant and multi-dimensional experiences.** Each experience of a social insider intrusion is unique, being both dependent on, and consequential to, the interpersonal relationship between the individuals involved. We

had set out to understand *what happens* in incidents of intrusion, and *how people experience them* (Chapter 6). We found social insider intrusions to be moments of personal significance to the parties involved. Despite their uniqueness, intrusions are not beyond systematization. We were able to distinguish patterns that describe important dimensions. Our analysis of *what happens* indicates, among others, that intrusions are often motivated by an impetus to control personal relationships with others; to occur in very limited time windows, often in homes and workplaces; and to commonly consist of inspecting communications in text form. Our analysis of *how people experience intrusions* revealed processes by which people make sense of incidents in the context of their relationships, in ways that resist simplification.

7.2 Limitations

We have, throughout this document, noted relevant limitations in the research methods we have selected. One more fundamental limitation, however, was our choice of observational unit.

Social insider intrusions, as we have defined them, are events, limited in time and scope. They are centered on a particular moment, in which a person, without permission, gains access to a device. Other observation units were possible. Notably, social insider intrusions have been framed as an element of intimate partner abuse. These analyses are rooted on a different observational unit, the identity. They ask: what is it like *being* a person who lives through intimate partner abuse? Intrusions to personal devices have been found to be increasingly part of those lives, but they are not the defining feature. Our analysis has greatly benefited from those accounts. However, in what refers to the intersection of intimate partner abuse and personal computing, our analysis is only complementary to the growing body of research that broaches the topic directly (e.g., Burke et al., 2011; Dimond et al., 2011; Freed et al., 2018, 2017; Leitão, 2019; Matthews et al., 2017; Woodlock, 2017).

7.3 Implications

Our research is grounded on the belief that *understanding* intrusions is necessary, to both demystify the conversation about social insiders, and to have actionable knowledge that can be put into action.

Social insider intrusions can be uncomfortable to discuss. Issues of security are, too often, predicated on simplified models of reality, in which there are only absolute rights and absolute wrongs. Under such models, design objectives are clear: we should, as best as we can, protect, mitigate, and recover from the consequences of wrong behaviors. This model, however useful, shows its limitations when confronted with issues of interpersonal privacy. Is it ever OK to inspect an intimate partner's device? Should parents have unrestricted access to their children's communications? Is it really a joke when someone defaces their friend's Facebook with innuendo about their sexual orientation? These are uncomfortable questions, and not everyone will agree on how to answer them. Yet we must not shy away from giving the best possible consideration to the consequences of different choices.

We strove to offer substance for reasoning through design choices. Our findings can be used to think through the space of possibilities, for instance by informing threat models, as

is practice in Security Engineering, or user journeys, as is practice in Interaction Design. We purposefully refrain from deriving strict prescriptions applicable to those or other disciplines of practice. The development of prescriptions, such as guidelines or design recommendations, is its own undertaking, and, done rigorously, arises from the interplay between the analysis of observational data, such as ours, and discipline-specific knowledge, which we lack. We can, however, offer a starting point to such work, in the form of a perspective grounded on our analysis: Computing systems are designed in ways that attempt to weigh different, sometimes opposing, dimensions. In systems operating on personal computing devices, a dimension that cannot be overlooked is the risk of intrusions by social insiders. The emerging reality of ubiquity and massification of personal computing devices has placed these technologies in a precarious space, where they can be mediators to building interpersonal relationships, but also of their erosion. Social insider intrusions violate the quest for agency in how others may know us. Agency, manifested in the exercise of meaningful consent, must guide the design of technologies that aim to be respectful of user's privacy.

Bibliography

Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514. https://doi.org/10.1126/science.aaa1465

Anne Adams and Martina Angela Sasse. 1999. Users are not the enemy. *Commun. ACM* 42, 12 (Dec. 1999), 40–46. https://doi.org/10.1145/322796.322806

Amazon Mechanical Turk. 2017. Best practices for managing Workers in follow-up surveys or longitudinal studies. Online, Last Accessed May 21, 2018. https://blog.mturk.com/tutorial-best-practices-for-managing-workers-in-follow-up-surveys-or-longitudinal-studies-4d0732a7319b

Judd Antin and Aaron Shaw. 2012. Social Desirability Bias and Self-reports of Motivation: A Study of Amazon Mechanical Turk in the US and India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2925–2934. https://doi.org/10.1145/2207676.2208699

Apple.com. 2017. Apple Support: Use Guided Access with iPhone, iPad, and iPod touch. Last accessed Feb. 21, 2017. https://support.apple.com/en-us/HT202612.

Jeffrey Bardzell and Shaowen Bardzell. 2016. Humanistic HCI. *interactions* 23, 2 (Feb. 2016), 20–29. https://doi.org/10.1145/2888576

Roy F. Baumeister, Arlene Stillwell, and Sara R. Wotman. 1990. Victim and perpetrator accounts of interpersonal conflict: Autobiographical narratives about anger. *Journal of Personality and Social Psychology* 59, 5 (1990), 994–1005. https://doi.org/10.1037/0022-3514.59.5.994

Adam Beautement, M. Angela Sasse, and Mike Wonham. 2008. The Compliance Budget: Managing Security Behaviour in Organisations. In *Proceedings of the 2008 New Security Paradigms Workshop (NSPW '08)*. ACM, New York, NY, USA, 47–58. https://doi.org/10.1145/1595676.1595684

Noam Ben-Asher, Niklas Kirschnick, Hanul Sieger, Joachim Meyer, Asaf Ben-Oved, and Sebastian Möller. 2011. On the Need for Different Security Methods on Mobile Phones. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)*. ACM, New York, NY, USA, 465–473. https://doi.org/10.1145/2037373.2037442

Chandrasekhar Bhagavatula, Blase Ur, Kevin Iacovino, Su Mon Kywe, Lorrie Faith Cranor, and Marios Savvides. 2015. Biometric authentication on iPhone and Android: Usability, perceptions, and influences on adoption. In *Proceedings of the NDSS Usable Security Workshop (USEC '15)*. Internet Society. https://doi.org/10.14722/usec.2015.23003

Graeme Blair and Kosuke Imai. 2010. list: Statistical Methods for the Item Count Technique and List Experiment. Available at The Comprehensive R Archive Network (CRAN). http://CRAN.R-project.org/package=list

Graeme Blair and Kosuke Imai. 2012. Statistical Analysis of List Experiments. *Political Analysis* 20, 1 (Jan. 2012), 47–77. https://doi.org/10.1093/pan/mpr048

Graeme Blair, Kosuke Imai, and Yang-Yang Zhou. 2015. Design and Analysis of the Randomized Response Technique. *J. Amer. Statist. Assoc.* 110, 511 (2015), 1304–1319. https://doi.org/10.1080/01621459.2015.1050028

Manuel Blum and Santosh Srinivas Vempala. 2015. Publishable Humanly Usable Secure Password Creation Schemas. In *Third AAAI Conference on Human Computation and Crowdsourcing*.

Virginia Braun, Victoria Clarke, Nikki Hayfield, Naomi Moller, and Irmgard Tischner. 2017. Qualitative Story Completion. In *Handbook of Research Methods in Health Social Sciences*. Springer, Singapore, 1–18. https://doi.org/10.1007/978-981-10-2779-6_14-1

Alex Braunstein, Laura Granka, and Jessica Staddon. 2011. Indirect Content Privacy Surveys: Measuring Privacy Without Asking About It. In *Proceedings of the Seventh Symposium on Usable Privacy and Security (SOUPS '11)*. ACM, New York, NY, USA, Article 15, 14 pages. https://doi.org/10.1145/2078827.2078847

Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling. 2011. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6, 1 (Feb. 2011), 3–5. https://doi.org/10.1177/1745691610393980

Brooke Bullek, Stephanie Garboski, Darakhshan J. Mir, and Evan M. Peck. 2017. Towards Understanding Differential Privacy: When Do People Trust Randomized Response Technique?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3833–3837. https://doi.org/10.1145/3025453.3025698

Sloane C. Burke, Michele Wallen, Karen Vail-Smith, and David Knox. 2011. Using technology to control intimate partners: An exploratory study of college undergraduates. *Computers in Human Behavior* 27, 3 (2011), 1162–1167. https://doi.org/10.1016/j.chb.2010.12.010

Daniel Buschek, Fabian Hartmann, Emanuel von Zezschwitz, Alexander De Luca, and Florian Alt. 2016. SnapApp: Reducing Authentication Overhead with a Time-Constrained Fast Unlock Option. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3736–3747. https://doi.org/10.1145/2858036.2858164

Business Insider. 2014. Facebook Rape — Frape — Is Now A Crime That Could Get You 10 Years In Prison, In Ireland. By Jim Edwards. Last accessed Nov. 27, 2018. http://www.businessinsider.com/frape-facebook-rape-now-a-crime-2014-7.

Business Insider. 2016. Project Abacus is Coming in 2016. By Rafi Letzter. Last accessed Feb. 21, 2017. http://www.businessinsider.com/project-abacus-is-coming-in-2016-2016-5.

Jesse Chandler, Gabriele Paolacci, Eyal Peer, Pam Mueller, and Kate A. Ratliff. 2015. Using Nonnaive Participants Can Reduce Effect Sizes. *Psychological Science* 26, 7 (2015), 1131–1139. https://doi.org/10.1177/0956797615585115

Ivan Cherapau, Ildar Muslukhov, Nalin Asanka, and Konstantin Beznosov. 2015. On the Impact of Touch ID on iPhone Passcodes. In *Proceedings of the Eleventh Symposium On Usable Privacy and Security (SOUPS '15)*. USENIX Association, Berkeley, CA, USA, 257–276. https://www.usenix.org/conference/soups2015/proceedings/presentation/cherapau

Sonia Chiasson, Alain Forget, Robert Biddle, and P C van Oorschot. 2009. User interface design affects security: patterns in click-based graphical passwords. *International Journal of Information Security* 8, 6 (2009), 387. https://doi.org/10.1007/s10207-009-0080-7

Erika Chin, Adrienne Porter Felt, Vyas Sekar, and David Wagner. 2012. Measuring User Confidence in Smartphone Security and Privacy. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12)*. ACM, New York, NY, USA, Article 1, 16 pages. https://doi.org/10.1145/2335356.2335358

Richard Chow, Philippe J. P. Golle, and Jessica N. Staddon. U.S. Patent 8,095,112, 2012. Adjusting security level of mobile device based on presence or absence of other mobile devices nearby.

N. L. Clarke and S. M. Furnell. 2005. Authentication of users on mobile telephones - A survey of attitudes and practices. *Computers and Security* 24, 7 (Oct. 2005), 519–527. https://doi.org/10.1016/j.cose.2005.08.003

Fernando J. Corbató, Marjorie Merwin-Daggett, and Robert C. Daley. 1962. An Experimental Time-sharing System. In *Proceedings of the May 1-3, 1962, Spring Joint Computer Conference (AIEE-IRE '62 (Spring))*. ACM, New York, NY, USA, 335–344. https://doi.org/10.1145/1460833.1460871

Daniel Corstange. 2008. Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT. *Political Analysis* 17, 1 (feb 2008), 45–63. https://doi.org/10.1093/pan/mpn013

Elisabeth Coutts and Ben Jann. 2008. Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods & Research* 40, 1 (Feb. 2008), 169–193. https://doi.org/10.1177/0049124110390768

Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann. 2012. Touch Me Once and I Know It's You!: Implicit Authentication Based on Touch Screen Patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 987–996. https://doi.org/10.1145/2207676.2208544

Alexander De Luca, Alina Hang, Emanuel von Zezschwitz, and Heinrich Hussmann. 2015. I Feel Like I'm Taking Selfies All Day!: Towards Understanding Biometric Authentication on Smartphones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1411–1414. https://doi.org/10.1145/2702123.2702141

Alexander De Luca, Marian Harbach, Emanuel von Zezschwitz, Max-Emanuel Maurer, Bernhard Ewald Slawik, Heinrich Hussmann, and Matthew Smith. 2014. Now You See Me, Now

You Don'T: Protecting Smartphone Authentication from Shoulder Surfers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2937–2946. https://doi.org/10.1145/2556288.2557097

Jill P. Dimond, Casey Fiesler, and Amy S. Bruckman. 2011. Domestic violence and information communication technologies. *Interacting with Computers* 23, 5 (2011), 413–421. https://doi.org/10.1016/j.intcom.2011.04.006

Serge Egelman, Sakshi Jain, Rebecca S Portnoff, Kerwell Liao, Sunny Consolvo, and David Wagner. 2014. Are You Ready to Lock? Understanding User Motivations for Smartphone Locking Behaviors. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS '14*. 750–761. https://doi.org/10.1145/2660267.2660273

Malin Eiband, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4254–4265. https://doi.org/10.1145/3025453.3025636

Adrienne Porter Felt, Serge Egelman, and David Wagner. 2012. I've got 99 problems, but vibration ain't one: A survey of smartphone users' concerns. In *Proceedings of the 2nd ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*. ACM Press, New York, NY, USA, 33–44. https://doi.org/10.1145/2381934.2381943

John C Flanagan. 1954. The critical incident technique. *Psychological bulletin* 51, 4 (1954), 327.

Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. 2018. "A Stalker's Paradise": How Intimate Partner Abusers Exploit Technology. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 667–667. https://doi.org/10.1145/3173574.3174241

Diana Freed, Jackeline Palmer, Diana Elizabeth Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. 2017. Digital Technologies and Intimate Partner Violence: A Qualitative Analysis with Multiple Stakeholders. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW, Article 46 (Dec. 2017), 22 pages. https://doi.org/10.1145/3134681

Ujwal Gadiraju, Ricardo Kawase, Stefan Dietze, and Gianluca Demartini. 2015. Understanding Malicious Behavior in Crowdsourcing Platforms: The Case of Online Surveys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1631–1640. https://doi.org/10.1145/2702123.2702443

Adam N. Glynn. 2013. What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment. *Public Opinion Quarterly* 77, S1 (Feb. 2013), 159–172. https://doi.org/10.1093/poq/nfs070

Google. 2018. Set your Android device to automatically unlock. Online, Last accessed Sep. 1, 2018. https://support.google.com/android/answer/9075927.

Google.com. 2017a. Accounts Help: Remotely ring, lock, or erase a lost device. Last accessed Feb. 21, 2017. https://support.google.com/accounts/answer/6160500.

Google.com. 2017b. Nexus Help: Add, switch, or delete users. Last accessed Feb. 21, 2017. https://support.google.com/nexus/answer/2865483.

Google.com. 2017c. Nexus Help: Pin and unpin screens. Last accessed Feb. 21, 2017. https://support.google.com/nexus/answer/6118421.

Google.com. 2017d. Nexus Help: Set up your device for automatic unlock. Last accessed Feb. 21, 2017. https://support.google.com/nexus/answer/6093922.

Google.com. 2017e. Nexus Help: Use restricted profiles on tablets. Last accessed Feb. 21, 2017. https://support.google.com/nexus/answer/3175031.

Alina Hang, Emanuel von Zezschwitz, Alexander De Luca, and Heinrich Hussmann. 2012. Too much information! User Attitudes towards Smartphone Sharing. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction Making Sense Through Design - NordiCHI '12*. ACM Press, New York, NY, USA, 284. https://doi.org/10.1145/2399016.2399061

Marian Harbach, Alexander De Luca, and Serge Egelman. 2016. The Anatomy of Smartphone Unlocking: A Field Study of Android Lock Screens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4806–4817. https://doi.org/10.1145/2858036.2858267

Marian Harbach, Emanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. 2014. It's a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception. In *Proceedings of the Tenth Symposium On Usable Privacy and Security*. USENIX Association, Berkeley, CA, USA, 213–230. https://www.usenix.org/conference/soups2014/proceedings/presentation/harbach

David J. Hauser and Norbert Schwarz. 2016. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods* 48, 1 (2016), 400–407. https://doi.org/10.3758/s13428-015-0578-z

Eiji Hayashi and Jason I. Hong. 2015. Knock x Knock: The Design and Evaluation of a Unified Authentication Management System. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 379—-389. https://doi.org/10.1145/2750858.2804279

Eiji Hayashi, Oriana Riva, Karin Strauss, A. J. Bernheim Brush, and Stuart Schechter. 2012. Goldilocks and the Two Mobile Devices: Going Beyond All-or-nothing Access to a Device's Applications. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12)*. ACM, New York, NY, USA, Article 2, 11 pages. https://doi.org/10.1145/2335356.2335359

Cormac Herley. 2009. So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. In *Proceedings of the 2009 New Security Paradigms Workshop (NSPW '09)*. ACM, New York, NY, USA, 133–144. https://doi.org/10.1145/1719030.1719050

Cormac Herley. 2014. More Is Not the Answer. *IEEE Security & Privacy magazine* 12, 1 (Jan. 2014), 14–19. https://doi.org/10.1109/MSP.2013.134

Kosuke Imai. 2011. Multivariate Regression Analysis for the Item Count Technique. *J. Amer. Statist. Assoc.* 106, 494 (June 2011), 407–416. https://doi.org/10.1198/jasa.2011.ap10415

Lilly C. Irani and M. Six Silberman. 2013. Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 611–620. https://doi.org/10.1145/2470654.2470742

Bjorn Markus Jakobsson, Richard Chow, and Runting Shi. U.S. Patent 12/955,825, 2012a. Implicit authentication.

Bjorn Markus Jakobsson, Mark J Grandcolas, Philippe JP Golle, Richard Chow, and Runting Shi. U.S. Patent 8,312,157, 2012b. Implicit authentication.

Markus Jakobsson, Elaine Shi, Philippe Golle, and Richard Chow. 2009. Implicit Authentication for Mobile Devices. In *Proceedings of the 4th USENIX Conference on Hot Topics in Security (HotSec'09)*. USENIX Association, Berkeley, CA, USA, 9.

Maritza Johnson, Serge Egelman, and Steven M. Bellovin. 2012. Facebook and Privacy: It's Complicated. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12)*. ACM, New York, NY, USA, Article 9, 15 pages. https://doi.org/10.1145/2335356.2335369

Adam Kapelner and Dana Chandler. 2010. Preventing Satisficing in Online Surveys: A "Kapcha" to Ensure Higher Quality Data. In *Proceedings of the 2010 CrowdConf*. CrowdConf, San Francisco, CA, USA, 10 pages.

Amy K. Karlson, A.J. Bernheim Brush, and Stuart Schechter. 2009. Can I Borrow Your Phone?: Understanding Concerns when Sharing Mobile Phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1647–1650. https://doi.org/10.1145/1518701.1518953

Jill N. Kearns and Frank D. Fincham. 2005. Victim and Perpetrator Accounts of Interpersonal Transgressions: Self-Serving or Relationship-Serving Biases? *Personality and Social Psychology Bulletin* 31, 3 (2005), 321–333. https://doi.org/10.1177/0146167204271594

Larry Koved, Pau-Chen Cheng, Diogo Marques, Nalini Ratha, Kapil Singh, Cal Swart, and Shari Trewin. 2016. *Usable Multi-Factor Authentication and Risk-Based Authorization*. Technical Report RC25619. IBM Research.

Balachander Krishnamurthy and Craig E. Wills. 2008. Characterizing Privacy in Online Social Networks. In *Proceedings of the First Workshop on Online Social Networks (WOSN '08)*. ACM, New York, NY, USA, 37–42. https://doi.org/10.1145/1397735.1397744

Stan Kurkovsky and Ewa Syta. 2010. Digital natives and mobile phones: A survey of practices and attitudes about privacy and security. In *Proceedings of the 2010 IEEE International Symposium on Technology and Society (ISTAS '10)*. IEEE, Piscataway, NJ, USA, 441–449. https://doi.org/10.1109/ISTAS.2010.5514610

Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2010. *Research methods in Human-Computer Interaction*. John Wiley & Sons.

Roxanne Leitão. 2019. Anticipating Smart Home Security and Privacy Threats with Survivors of Intimate Partner Abuse. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. ACM, New York, NY, USA, 527–539. https://doi.org/10.1145/3322276.3322366

Roy J. Lewicki, Edward C. Tomlinson, and Nicole Gillespie. 2006. Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions. *Journal of Management* 32, 6 (2006), 991–1022. https://doi.org/10.1177/0149206306294405

Thomas Lumley. 2004. Analysis of complex survey samples. *Journal of Statistical Software* 9, 8 (2004), 1–19. http://www.jstatsoft.org/article/view/v009i08

Joseph Maguire and Karen Renaud. 2012. You Only Live Twice or "the Years We Wasted Caring About Shoulder-surfing". In *Proceedings of the 26th International BCS Human Computer Interaction Conference (BCS-HCI '12)*. British Computer Society, Swinton, UK, 404–409. http://doi.acm.org/10.1145/2377916.2377975

Ahmed Mahfouz, Ildar Muslukhov, and Konstantin Beznosov. 2016. Android users in the wild: Their authentication and usage behavior. *Pervasive and Mobile Computing* 32 (2016), 50 – 61. https://doi.org/10.1016/j.pmcj.2016.06.017 Mobile Security, Privacy and Forensics.

Diogo Marques, Tiago Guerreiro, Luís Duarte, and Luís Carriço. 2013. Under the Table: Tap Authentication for Smartphones. In *Proceedings of the 27th International BCS Human Computer Interaction Conference (BCS-HCI '13)*. British Computer Society, Swinton, UK, Article 33, 6 pages. http://doi.acm.org/10.1145/2578048.2578090

Tara Matthews, Kerwell Liao, Anna Turner, Marianne Berkovich, Robert Reeder, and Sunny Consolvo. 2016. "She'll Just Grab Any Device That's Closer": A Study of Everyday Device & Account Sharing in Households. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5921–5932. https://doi.org/10.1145/2858036.2858051

Tara Matthews, Kathleen O'Leary, Anna Turner, Manya Sleeper, Jill Palzkill Woelfer, Martin Shelton, Cori Manthorne, Elizabeth F. Churchill, and Sunny Consolvo. 2017. Stories from Survivors: Privacy & Security Practices when Coping with Intimate Partner Abuse. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2189–2201. https://doi.org/10.1145/3025453.3025875

Michelle L. Mazurek, J. P. Arsenault, Joanna Bresee, Nitin Gupta, Iulia Ion, Christina Johns, Daniel Lee, Yuan Liang, Jenny Olsen, Brandon Salmon, Richard Shay, Kami Vaniea, Lujo Bauer, Lorrie Faith Cranor, Gregory R. Ganger, and Michael K. Reiter. 2010. Access Control for Home Data Sharing: Attitudes, Needs and Practices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 645–654. https://doi.org/10.1145/1753326.1753421

Paul McDonald, Matt Mohebbi, and Brett Slatkin. 2012. Comparing Google Consumer Surveys to existing probability and non-probability based internet surveys. Google Whitepaper,

Retrieved Jan 19, 2016. https://www.google.com/insights/consumersurveys/static/consumer_surveys_whitepaper_v2.pdf.

Susan McNeeley. 2012. Sensitive Issues in Surveys: Reducing Refusals While Increasing Reliability and Quality of Responses to Sensitive Survey Items. *Handbook of Survey Methodology for the Social Sciences* (2012), 377–396. https://doi.org/10.1007/978-1-4614-3876-2_22

James Mickens. 2014. This World Of Ours. *;login: logout* January 2014 (jan 2014).

Hendrik Müller, Jennifer L. Gove, John S. Webb, and Aaron Cheang. 2015. Understanding and Comparing Smartphone and Tablet Use: Insights from a Large-Scale Diary Study. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction (OzCHI '15)*. ACM, New York, NY, USA, 427–436. https://doi.org/10.1145/2838739.2838748

Ildar Muslukhov, Yazan Boshmaf, Cynthia Kuo, Jonathan Lester, and Konstantin Beznosov. 2012. Understanding users' requirements for data protection in smartphones. In *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering Workshops (ICDEW '12)*. IEEE, Piscataway, NJ, USA, 228–235. https://doi.org/10.1109/ICDEW.2012.83

Ildar Muslukhov, Yazan Boshmaf, Cynthia Kuo, Jonathan Lester, and Konstantin Beznosov. 2013. Know Your Enemy: The Risk of Unauthorized Access in Smartphones by Insiders. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. ACM, New York, NY, USA, 271–280. https://doi.org/10.1145/2493190.2493223

Diana C. Mutz. 2011. *Population-Based Survey Experiments*. Princeton University Press. http://www.jstor.org/stable/j.ctt7sf3s

Hendrik Müller, Aaron Sedley, and Elizabeth Ferrall-Nunge. 2014. Survey Research in HCI. In *Ways of Knowing in HCI*, Judith S. Olson and Wendy A. Kellogg (Eds.). Springer New York, 229–266. https://doi.org/10.1007/978-1-4939-0378-8_10

Daniel M. Oppenheimer, Tom Meyvis, and Nicolas Davidenko. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology* 45, 4 (2009), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Leysia Palen and Paul Dourish. 2003. Unpacking "Privacy" for a Networked World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 129–136. https://doi.org/10.1145/642611.642635

Gabriele Paolacci and Jesse Chandler. 2014. Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science* 23, 3 (2014), 184–188. https://doi.org/10.1177/0963721414531598

Cheul Young Park, Cori Faklaris, Siyan Zhao, Alex Sciuto, Laura Dabbish, and Jason Hong. 2018. Share and Share Alike? An Exploration of Secure Behaviors in Romantic Relationships. In *Proceedings of the Fourteenth Symposium on Usable Privacy and Security (SOUPS '18)*. USENIX Association, Berkeley, CA, USA, 83–102. https://www.usenix.org/conference/soups2018/presentation/park

Christopher Parsons, Adam Molnar, Jakub Dalek, Jeffrey Knockel, Miles Kenyon, Bennett Haselton, Cynthia Khoo, and Ron Deibert. 2019. The Predator in Your Pocket: A Multidisciplinary Assessment of the Stalkerware Application Industry. Citizen Lab Research Report No. 119, University of Toronto, June 2019.

Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70 (may 2017), 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Eyal Peer, Joachim Vosgerau, and Alessandro Acquisti. 2014. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods* 46, 4 (Dec. 2014), 1023–1031. https://doi.org/10.3758/s13428-013-0434-y

Pew Research Center. 2012a. A Comparison of Results from Surveys by the Pew Research Center and Google Consumer Surveys. Report. Last accessed Feb. 21, 2017.. https://www.people-press.org/2012/11/07/a-comparison-of-results-from-surveys-by-the-pew-research-center-and-google-consumer-surveys/

Pew Research Center. 2012b. Privacy and Data Management on Mobile Devices. Report. Last accessed Feb. 21, 2017.. http://www.pewinternet.org/2012/09/05/privacy-and-data-management-on-mobile-devices/

Pew Research Center. 2014. The Future of Privacy. Report. Last accessed Feb. 21, 2017.. http://www.pewinternet.org/2014/12/18/future-of-privacy/

Pew Research Center. 2015a. The Demographics of Social Media Users. Report. Last accessed Feb. 21, 2017.. http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/

Pew Research Center. 2015b. The Smartphone Difference. Report. Retrieved Jan 19, 2016. http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/

Salil Prabhakar, Sharath Pankanti, and Anil K. Jain. 2003. Biometric Recognition: Security and Privacy Concerns. *IEEE Security and Privacy* 1, 2 (March 2003), 33–42. https://doi.org/10.1109/MSECP.2003.1193209

D. Raghavarao and W. T. Federer. 1979. Block Total Response as an Alternative to the Randomized Response Method in Surveys. *Journal of the Royal Statistical Society. Series B (Methodological)* 41, 1 (1979), 40–45.

Oriana Riva, Chuan Qin, Karin Strauss, and Dimitrios Lymberopoulos. 2012. Progressive Authentication: Deciding When to Authenticate on Mobile Phones. In *Proceedings of the 21st USENIX Security Symposium (USENIX Security 12)*. USENIX Association, Berkeley, CA, 301–316. https://www.usenix.org/conference/usenixsecurity12/technical-sessions/presentation/riva

Bryn Rosenfeld, Kosuke Imai, and Jacob N. Shapiro. 2015. An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions. *American Journal of Political Science* (2015). https://doi.org/10.1111/ajps.12205

Nithya Sambasivan, Garen Checkley, Amna Batool, Nova Ahmed, David Nemer, Laura Sanely Gaytán-Lugo, Tara Matthews, Sunny Consolvo, and Elizabeth Churchill. 2018. "Privacy is not for me, it's for those rich women": Performative Privacy Practices on Mobile Phones by Women in South Asia. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. USENIX Association, Berkeley, CA, USA, 127–142. https://www.usenix.org/conference/soups2018/presentation/sambasivan

Julian Seifert, Alexander De Luca, Bettina Conradi, and Heinrich Hussmann. 2010. TreasurePhone: Context-Sensitive User Data Protection on Mobile Phones. In *Proceedings of the 8th International Conference on Pervasive Computing (Pervasive 2010)*. Springer Berlin Heidelberg, Berlin, Heidelberg, Germany, 130–137. https://doi.org/10.1007/978-3-642-12654-3_8

Richard Shay, Iulia Ion, Robert W. Reeder, and Sunny Consolvo. 2014. "My Religious Aunt Asked Why I Was Trying to Sell Her Viagra": Experiences with Account Hijacking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2657–2666. https://doi.org/10.1145/2556288.2557330

Elaine Shi, Yuan Niu, Markus Jakobsson, and Richard Chow. 2010. Implicit Authentication through Learning User Behavior. In *Proceedings, Information Security - 13th International Conference (ISC 2010)*. Springer Berlin Heidelberg, Berlin, Heidelberg, Germany, 99–113. https://doi.org/10.1007/978-3-642-18178-8_9

Tasos Spiliotopoulos and Ian Oakley. 2013. Understanding Motivations for Facebook Use: Usage Metrics, Network Structure, and Privacy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3287–3296. https://doi.org/10.1145/2470654.2466449

Evan Stark. 2007. *Coercive control: The entrapment of women in personal life*. Oxford University Press, New York, NY, USA.

Statista. 2018. Distribution of Facebook users in the United States as of January 2018, by age group and gender. Last accessed Nov. 21, 2018, from https://www.statista.com/statistics/187041/us-user-age-distribution-on-facebook/.

Gareth Terry, Nikki Hayfield, Victoria Clarke, and Virginia Braun. 2017. Thematic Analysis. In *The SAGE Handbook of Qualitative Research in Psychology*. SAGE Publications Ltd, London, UK, 17–36. https://doi.org/10.4135/9781526405555.n2

Roger Tourangeau and Ting Yan. 2007. Sensitive questions in surveys. *Psychological Bulletin* 133, 5 (Sept. 2007), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

Dirk Van Bruggen, Shu Liu, Mitch Kajzer, Aaron Striegel, Charles R. Crowell, and John D'Arcy. 2013. Modifying Smartphone User Locking Behavior. In *Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUPS '13)*. ACM, New York, NY, USA, Article 10, 14 pages. https://doi.org/10.1145/2501604.2501614

Emanuel von Zezschwitz, Alexander De Luca, Bruno Brunkow, and Heinrich Hussmann. 2015a. SwiPIN: Fast and Secure PIN-Entry on Smartphones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1403–1406. https://doi.org/10.1145/2702123.2702212

Emanuel von Zezschwitz, Alexander De Luca, Philipp Janssen, and Heinrich Hussmann. 2015b. Easy to Draw, but Hard to Trace?: On the Observability of Grid-based (Un)Lock Patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2339–2342. https://doi.org/10.1145/2702123.2702202

Emanuel von Zezschwitz, Paul Dunphy, and Alexander De Luca. 2013. Patterns in the Wild: A Field Study of the Usability of Pattern and Pin-based Authentication on Mobile Devices. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. ACM, New York, NY, USA, 261–270. https://doi.org/10.1145/2493190.2493231

Stanley L. Warner. 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *J. Amer. Statist. Assoc.* 60, 309 (1965), 63–69. https://doi.org/10.1080/01621459.1965.10480775

Washington Post. 2016. FBI paid professional hackers one-time fee to crack San Bernardino iPhone. By Ellen Nakashima. Last accessed Feb. 21, 2017. https://washingtonpost.com/5397814a-00de-11e6-9d36-33d198ea26c5_story.html.

Alma Whitten and J. D. Tygar. 1999. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *Proceedings of the 8th USENIX Security Symposium (USENIX Security 99))*. USENIX Association, Berkeley, CA, USA, 169–184. https://www.usenix.org/legacy/events/sec99/full_papers/whitten/whitten_html/index.html

Michael W. Wiederman. 1997. The Truth Must Be in Here Somewhere: Examining the Gender Discrepancy in Self-Reported Lifetime Number of Sex Partners. *The Journal of Sex Research* 34, 4 (1997), 375–386. http://www.jstor.org/stable/3813479

Oliver Wiese and Volker Roth. 2016. See You Next Time: A Model for Modern Shoulder Surfers. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)*. ACM, New York, NY, USA, 453–464. https://doi.org/10.1145/2935334.2935388

Wired. 2012. The World's First Computer Password? It Was Useless Too. By Robert McMillan. Last accessed Feb. 21, 2017. https://www.wired.com/2012/01/computer-password/.

Delanie Woodlock. 2017. The Abuse of Technology in Domestic Violence and Stalking. *Violence Against Women* 23, 5 (2017), 584–602. https://doi.org/10.1177/1077801216646277

Jeanne S. Zechmeister and Catherine Romero. 2002. Victim and offender accounts of interpersonal conflict: Autobiographical narratives of forgiveness and unforgiveness. *Journal of Personality and Social Psychology* 82, 4 (2002), 675–686. https://doi.org/10.1037/0022-3514.82.4.675

Mary Ellen Zurko and Richard T. Simon. 1996. User-centered Security. In *Proceedings of the 1996 New Security Paradigms Workshop (NSPW '96)*. ACM, New York, NY, USA, 27–33. https://doi.org/10.1145/304851.304859