User Privacy Preservation on Mobile Devices: Investigating the Role of Contextual Integrity

by

Ming Di Leom

MSc (Cyber Security and Forensic Computing)

Bachelor in Computing (Hons)

A thesis submitted for the degree of

Doctor of Philosophy

Computer and Information Science

School of Information Technology & Mathematical Sciences
Division of Information Technology, Engineering and the Environment

University of
South Australia

June 2020

Table of Contents

List of Figures vi

List of Tables vii

List of Abbreviations viii

Abstract ix

Declaration x

Acknowledgments xi

1 Introduction 1

2 Literature Review 5

2.1 Privacy as a concept 5

2.2 Privacy paradox 7

2.3 Trust 10

2.4 Self-Disclosure 12

2.5 Privacy profiling 14

2.6 Dichotomy of privacy 18

2.7 Framework of Contextual Integrity 19

2.7.1 Sender/ Recipient 24

2.7.2 Subject of information 24

2.7.3 Type of information 25

2.7.4 Purpose of information collection 25

2.8 Discussion 26

2.9 Research Design 29

2.10 Conclusion 31

3 A Practical Exploration of the Privacy Paradox: The Role of Contextual Integrity 32

3.1 Introduction 32

3.2 Study 1A 35

3.2.1 Methodology 35

3.2.2 Results 36

3.3 Study 1B 41

3.3.1 Measures 41

3.3.2 Validation 43

3.3.3 Results 44

3.3.4 Moderation and mediation 60

3.4 Discussion 62

3.5 Conclusion 65

4 Information Disclosure in Mobile Device: Examining the Influence of Information Relevance and Recipient 66
4.1 Introduction 66
4.2 Study 2A 67
4.2.1 Methodology 67
4.2.2 Results 69
4.3 Study 2B 71
4.3.1 Measures 71
4.3.2 Methodology 72
4.3.3 Demographics 74
4.4 Discussion 79
4.5 Conclusion 81
5 Discussion and Conclusion 82
Appendix A Questionnaires for Studies 1A and 1B 88
Appendix B Questionnaires for Studies 2A and 2B 90
Appendix C Participant Information Sheet 91
References 93

List of Figures

Figure 1 Contextual Privacy Framework 30
Figure 2 Influence of trust 33
Figure 3 Power-law distribution (Study 1A) 37
Figure 4 Boxplots of top 15 groups 39
Figure 5 Demographical differences in Simple Trust 47
Figure 6 Demographical differences in ITS 49
Figure 7 Demographical differences in privacy concern 52
Figure 8 Demographical differences in self-disclosure 55
Figure 9 Q-Q plot of Frequent group 59
Figure 10 Q-Q plot of Infrequent group 59
Figure 11 Q-Q plot of combined group 60
Figure 12 Mediation effects 62
Figure 13 Influence of information relevance 67
Figure 14 Power-law distribution (Study 2A) 70
Figure 15 Willingness to disclose across demographics 75
Figure 16 Contextual Privacy Framework 84

List of Tables

Table 1 Comparison of current literature that adapted the CI framework. 24
Table 2 Demographics of Study 1A 37
Table 3 Compilation of groups 38
Table 4 Statistical information of groups 39
Table 5 Conover Test with t statistic values 40
Table 6 Statistical information of each scale 43
Table 7 Categorisation of top 15 groups into Frequent and Infrequent 43
Table 8 Demographics of Study 1B 45
Table 9 Differences in Simple Trust (ST) among demographics 48
Table 10 Differences in ITS between demographics 50
Table 11 Differences in privacy concern among demographics 53
Table 12 Self-disclosure differences between demographics 56
Table 13 Summary of all regression models. 57
Table 14 Correlation between demographics, trust, privacy concern and self-disclosure 58
Table 15 Results of regression model 1 and 2 60
Table 16 Moderation effect comparison between frequent and infrequent groups 61
Table 17 Moderation effect with or without separation of ST & ITS 61
Table 18 Demographics of Study 2A 69
Table 19 Compilation of types 71
Table 20 15 most popular types 71
Table 21 Demographics of Study 2B 73
Table 22 Demographics differences in willingness to disclose 76
Table 23 Average indexes in different groups 77
Table 24 Average indexes of each information type 77
Table 25 Average disclosure index 78
Table 26 Average relevance index 78
Table 27 Differences in disclosure and relevance indexes 79
Table 28 Regression effect of relevance on willingness to disclose 79
Table 29 Empirically-derived classifications 85

List of Abbreviations

AI	Artificial Intelligence
CI	Contextual Integrity
DTS	Dyadic Trust Scale
ITS	Individualized Trust Scale, alias of WITS
GIPC	Global Information Privacy Concern
MTurk	Mechanical Turk
OS	Operating System
PET	Privacy Enhancing Technologies
RITS	Rotter's Interpersonal Trust Scale
SD	Self-disclosure
SITS	Specific Interpersonal Trust Scale
SNS	Social networking sites
ST	Simple Trust
WITS	Wheeless' Individualized Trust Scale

Abstract

Privacy issues emerge when mobile devices can not only collect user information but also share it automatically in the background, opportunities for which were previously limited. The general approach of privacy preservation in mobile devices through permissions management alone is not optimal due to the gap between flexibility and usability. The framework of contextual integrity (CI) has been proposed to accommodate diversity in contexts and also users' privacy preferences. Accommodating the diversity in contexts and also users' privacy preferences is complicated by privacy paradox—a discrepancy between expressed concern and the actual behaviour. We examined how the framework can address privacy paradox and its implications on mobile devices in two user studies.

In the first study, we examined the prevalence of privacy paradox through the lens of the CI framework. The framework emphasises on the influence of contextual factors in our every day's mobile usage. We examined one such contextual factor is the recipients—user's attitude towards them. The results suggest trust having a significant influence on the user's disclosure behaviour, particularly on the relationship between privacy concern and self-disclosure. The mediation effect of trust in our results suggest its significant role in determining users' self-disclosure despite the existence of privacy concern. The findings offer a meaningful explanation behind privacy paradox; where a user is more likely to disclose to a trusted recipient, despite having privacy concern.

In the second study, we examined the impact of two contextual factors—recipient and information type—on the relationship between information relevance and self-disclosure. While there is evidence of a significant relationship between information relevance and disclosure, several discrepancies showed the relationship is not always clear-cut. The results highlight users' attitude on disclosure within the mobile ecosystem is often fraught with nuances and the use of generic information relevance in predicting the tendency to disclose may not be as effective as expected. Our results from the second study also cast doubt over the established effects of "sensitivity" and its usefulness in privacy enhancing technologies (PET). We observed inconsistent response in willingness to disclose a type of information across recipients. This further demonstrates that sensitivity can vary according to the intended recipient.

Overall, this thesis demonstrates the relevance of the CI framework in the mobile space and its potential to improve the current approach in PET, particularly the privacy recommendation system. The privacy recommendation system is a promising answer to the dilemma of having too little or too much privacy control. We believe by incorporating a crucial metric, "recipient", in addition to other contextual factors, the privacy recommendation system can advance its effectiveness. By taking into account of users' interactions with their recipients, the metric enables the ability to accommodate the ever-changing contexts and the diversity of users' privacy preferences.

Declaration

This thesis presents work carried out by myself and does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; to the best of my knowledge it does not contain any materials previously published or written by another person except where due reference is made in the text; and all substantive contributions by others to the work presented, including jointly authored publications, are clearly acknowledged.

Ming Di Leom 11 June 2020

Acknowledgments

I would like to thank my supervisor, Dr. Gaye Deegan, for her support and guidance. I want to thank my co-supervisors, Prof. John Boland, Dr. Raymond Choo and Dr. Ben Martini for their insights. I also would like to thank my colleagues. Dr. Alireza Heravi, for sharing his expertise. Baber and Haneen for making the lab a lively space. I would like to express my gratitude to my parents and my partner Claire for their undying companionship.

I received funding from the UniSA STEM (formerly School of Information Technology & Mathematical Sciences) to conduct the user studies.

1 Introduction

Today's mobile devices are equipped with a myriad of features introduced by sensors and mobile applications (apps). Mobile devices serve a wide variety of purposes including productivity, entertainment, and socialising. There are known privacy implications in the use of mobile devices and apps since they not only collect user information (e.g. location, health data and other sensitive data) but also share user information automatically in the background.

Several studies have revealed the gap between experience and expectation. Many users may not be aware of the extent of information collected in the background (Mehrnezhad et al. 2017) that could potentially be traded for developer's profit in exchange for "free" apps (Alvarez et al. 2019; Isaac 2017). Users might be surprised and feel uneasy when confronted with such possibilities (Jung, J, Han & Wetherall 2012; Shih, Liccardi & Weitzner 2015; Thompson, C et al. 2013). The discrepancy between experience and expectation is perceived by the user as a privacy violation.

Excessive tracking and data collection (Cyphers & Gebhart 2019) on the web (Howe & Nissenbaum 2017; Mathur et al. 2019), to mobile device (FTC 2013; Kaldestad & Myrstad 2018; Razaghpanah et al. 2018; Thompson, SA & Warzel 2019), even to television (FTC 2016; Huang, DY et al. 2019; Moghaddam et al. 2019; Ren et al. 2019), user data is often collected under flawed notice and consent (more commonly known as privacy policy) (Acquisti & Grossklags 2005) and most likely *surrendered* by the users rather unwillingly (Walker 2016). Studies estimated that as much as 10% of the permissions were granted reluctantly (Bonné et al. 2017); and at least 80% participants wished they could have denied the permission request, once they knew its purpose (Wijesekera et al. 2015). Consequently, fuelled by the recent high-profile scandals (Yahoo (Lord 2016), Cambridge Analytica (Rosenberg 2018), Equifax (FTC 2017),

Google+ (Smith, B 2018)), there is a diminishing public trust in public and private institutions (Chan et al. 2008; Olmstead & Smith 2017; Shipman & Marshall 2020).

To mitigate the gap between experience and expectation, there is significant research on Privacy Enhancing Technology (PET). PET is designed to protect user privacy in a system by preventing or minimising the unnecessary or unwanted collection, processing, and storage of data without loss of functionality (Van Blarkom, Borking & Olk 2003). PET is more relevant than ever, in a world where mass *datafication* (Mejias & Couldry 2019) is the new norm; it is a tool for check and balance, to tame the unwieldly datafication thas has lead to the power disparity that we can observe today between private institution and consumers (King & Katsanevas 2019). The growing interest in PET research is a sign that privacy issue is anything but gone. The proliferation of dark patterns (Nouwens et al. 2020) designed to obscure unrestrained data extraction and sidestep existing regulations further exacerbate privacy issues and threaten to erode the effectiveness of PET.

In addition to extensive research, PET has been recognised by several international organisations. European Commission of Organisation for Economic Co-operation and Development (OECD) held several symposiums to discuss PET (OECD 2013). The Office of the Australian Information (OAIC 2014) recommends minimising privacy risks when handling user data in an organisation. The Madrid Declaration includes the proclamation to "Reaffirm support for genuine Privacy Enhancing Techniques that minimise or eliminate the collection of personally identifiable information and for meaningful Privacy Impact Assessments that require compliance with privacy standards" (Civil Society 2009, p.2). These affirm the importance of PETs in protecting user privacy.

Existing PET is inflexible in accommodating the diversity of users' privacy preferences and the context that can varies even throughout a single day. This limitation hinders its full potential in privacy preservation. "Almost everything—things that we do, events that occur, transactions that take place—happens in a *context* [emphasis added]" (Nissenbaum 2004); any data collection is always within a particular context and to serve a specific purpose(s). An immediate response is to provide more fine-grained control; however, introducing more control through plethora of options merely kicks the can down the road. Without careful treading, excessive choices can be a significant cognitive burden

that subsequently causes decision fatigue to the user (Svirsky 2019; Utz et al. 2019), thereby not exerting control over technology. Thus, it is futile having to balance between flexibility (or control) and user-friendliness, losing control tends to lead to reduction of PET's effectiveness.

An ideal approach should not affect PET's effectiveness, regardless where a scale tips toward. Researchers have recently ventured into integrating artificial intelligence (AI) technique such as machine learning into PET to automate its decisions (Gao et al. 2020; Liu, B 2019). Combining with the emerging of AI accelerator in mobile devices (Grob 2017), this is a promising approach in which users still can exert control by adjusting the levels of automation (Colnago et al. 2020), without adversely affect the PET. Prior to implementing AI, an important consideration is deciding what input to feed into its decision engine.

This thesis aims to gain more understanding on mobile device usage to conform to the current enormous diversity in contexts and users' privacy preferences. Specifically, we aim to identify the factors that are cardinal to the mobile users' privacy needs. The identified factors have a potential utility in incorporating AI into PET—as input to its decision engine. The insights that we gain from this thesis can also inform private institutions in tailoring their products or services to individual's unique privacy needs.

The framework of contextual integrity (CI) (Nissenbaum 2004) has been proposed to accommodate diversity in contexts and also users' privacy preferences. The framework evaluates whether the flow of information is appropriate in a given context. It suggests that privacy is violated whenever the information flow is deemed as inappropriate due to social outrage that it could cause. This property is especially relevant to the highly volatile mobile ecosystem and has been gaining a foothold in the mobile-related research field (Jia et al. 2017; Martin & Shilton 2016a; Wijesekera et al. 2015; Wu, Vitak & Zimmer 2019).

Accommodating the diversity in contexts and users' privacy preferences is complicated by privacy paradox—a discrepancy between expressed concern and the actual behaviour. Improving platform protection required understanding of the occurrences of privacy paradox, understanding the factors that influence users' privacy decisions and put forward recommendations that different stakeholder can implement to protect mobile user's privacy.

The rest of this thesis is structured as follows: Chapter 2 provides background on the CI framework and a survey of related work on the mobile ecosystem's PET. Two user studies are conducted to investigate the influence of contextual factors: *recipient* in Chapter 3 and *type of information* in Chapter 4, to examine the phenomenon of privacy paradox. Chapter 3 investigates the influence of trust and privacy concern on self-disclosure, in relation to the typical groups of recipient found in mobile devices. Chapter 4 investigates the influence of relevance of information types on the willingness of disclosure towards typical groups of recipient. Chapter 5 concludes this thesis with a summary of contributions and possibilities for future work.

2 Literature Review

To conduct a review of related work, we adopted the three-stage approach (Webster and Watson, cited in Smith, HJ, Dinev & Xu 2011) to identify the relevant publications up until 07/2017 that will be discussed in this chapter. As privacy theories have been discussed throughout history, some articles we discussed can date back to the 19th century, but mostly from 1973 onwards. In the three-stage approach, we first search for articles using a combination of keywords on academic databases. Second, then reviewed the citations of those publications to identify additional potential articles. Finally, we used Google Scholar to identify additional potential publications that cited the relevant articles identified previously. For privacy theory topic, we initially identified publications in English using Google Scholar using search terms "privacy (theory OR framework) review". Since most of the research in developing PET for mobile devices is focused on the Android operating system (OS) (Fang, Han & Li 2014; Neisse et al. 2016), we used the search terms "Android (security OR access control OR privacy)" on Google Scholar and academic databases such as ScienceDirect, ACM Digital Library, IEEE Xplore, and Springer. We then choose articles that adapted privacy theories.

The rest of this chapter is organised as follows: Section 1 discusses various definitions of privacy. Section 2 introduces the phenomenon of privacy paradox. Section 3 discusses the use of privacy profiling in Privacy Enhancing Technology (PET). Section 4 cogitates the potential issues of treating privacy as dichotomies. Section 5 details the framework of contextual integrity (CI). Section 6 discusses the literature gaps. Section 7 concludes this chapter.

2.1 Privacy as a concept

Historical accounts of privacy, date as far back as the Ancient Greeks: Hippocrates and Aristotle. Hippocrates asserts the importance of patient's privacy in the Hippocratic Oath, "And about whatever I may see or hear in treatment, or even without treatment, in the life of human beings—things that should not ever be blurted out outside—I will remain silent, holding such things to be unutterable [sacred, not to be divulged]" (von Staden 1996). Aristotle proposed a distinction between the public sphere of political activity, the *polis*,

and the private sphere of the family, the *oikos* (DeCew 2015). Instances of privacy are also arguably present in Roman law and Biblical literature (Moore, cited in Wacks 2010).

Even with the long history of privacy, the concept is still under active debate in the 21st century (Allmer 2011; Tavani 2008; Westin 2003), and each researcher or philosopher has their definition of the concept. This is because the concept of privacy is ever-evolving and often driven by the development of new technologies (Westin 2003). The notion of "modern" technology stepping on privacy is hardly new; in response to eavesdropping devices employed by the press, Warren and Brandeis (1890) argued: "…the existing law affords a principle which may be invoked to protect the privacy of the individual from invasion either by…the possessor of any other modern device for recording or reproducing scenes or sounds". The statement also sets forth their argument of privacy as a right to be left alone.

Despite people's desire for seclusion, Altman (1975) argued that people also seek interaction. *Desired* privacy, in his view, is when people seek or restrict an ideal level of interaction depending on the circumstances. Similarly, Westin (2003) proposed four states of individual privacy (i.e. solitude, intimacy, anonymity, and reserve) that a person may experience at different times. At one moment, a person may want to be left alone (solitude) while another moment seeks companionship (intimacy). These definitions share a similar property, whereby they view privacy as the *physical* interaction between people.

Unprecedented computational power and sophistication in the 21st century enable most of the data that exist today to be digitalised. It is estimated that there are over 2.7 zettabytes of digital data exist today (Vesset et al. 2012). Digital data has the advantages of storage, transmission and can be accessed instantly. However, instant information retrieval introduced new sets of privacy challenges that did not exist before (Moor 1997). This can cause a person to be susceptible to harm in the future, dubbed an "architectural" issue (Solove 2006), in contrast to direct reputational harm, for example by the press (Warren & Brandeis 1890). Solove (2006) proposed two common forms of architectural issues: (1) it could increase the likelihood of harm. Companies often conduct consumer profiling—aggregating relevant information about a consumer from different sources—usually for targeted advertising. However, inappropriate uses could distort the profile of a consumer and subsequently cause reputational harm; (2) it can create a power

asymmetry or "informational inequality" between the public or private institution and the consumers (Flaherty 2014; King 2018; King & Katsanevas 2019; Nissenbaum & Patterson 2016; Solove 2006; White House 2017). Thus, although physical privacy is still as vital as ever, much of the concern today is in the informational sense.

Warren and Brandeis (1890) proposed the influential "right to be left alone" (White House 2012, 2017). They argue that a person should be able to enjoy solitude without intrusion. However, Tavani (2008) argues that the statement conflates the notion of "having privacy" and "being let alone". They are not mutually exclusive considering our modern lives revolve around the digital world in the 21st century. For example, in cyberstalking, the stalker follows the victim's digital trails initially to identify the victim's movement pattern, without the victim noticing. The stalker is leaving the victim alone, but the victim's privacy has been violated.

2.2 Privacy paradox

One approach to studying users' privacy preferences is to survey their expressed preferences, where they express their preferences explicitly. Survey participants are given a set of questions to recall their privacy experiences and then express them as their privacy preferences. It is possible that they could inflate their privacy concerns, thus causing a disparity between expressed concern and the actual behaviour. This phenomenon is known as the privacy paradox (Barnes 2006; Norberg, Horne & Horne 2007). As memory is not always reliable, this can cause cognitive bias in the recall-based survey approach. Shih (2015) argued that the privacy paradox might be due to such bias. These factors can cast doubts on the representation of users' privacy concern. Several studies have also challenged the privacy paradox hypothesis (Kokolakis 2017).

Ginosar and Ariel (2017) outlined that privacy paradox is explained by three variables in the research literature: (a) knowledge (b) benefits (c) trust. For example in *knowledge*, Martin and Nissenbaum (2016) argued that those surveys (that claim privacy paradox) are flawed because people might not be aware of the extent of data collected, how the collected data can be used later, or how they can be affected by those practices. Consumers often lack the understanding of the business practices of online firms; thus are unable to make informed decisions on the available privacy options (Acquisti, Taylor & Wagman 2015; Hoofnagle & Urban 2014; Nehf 2011). Privacy-friendly options are often deliberately obscured in favour of privacy-intrusive options (Kaldestad & Myrstad

2018). Even with privacy-friendly options, cognitive bias could lead users to overconfidence which lower privacy concern, and consequently lead to oversharing (Brandimarte, Acquisti & Loewenstein 2012; Buchanan et al. 2007).

In mobile app, people discloses information in return for *benefit* such as an app's functionality (e.g. personalisation) or financial incentives (Grossklags & Acquisti 2007; Hann et al. 2007; Hanson et al. 2020; Mohammad 2019) or due to app's positive presentation (Kehr et al. 2015) or just for the sheer enjoyment (Lin, K-Y & Lu 2011; Tamir & Mitchell 2012) and possibly many others (Muhammad, Dey & Weerakkody 2017) (cf. Bylund, Peterson & Cameron 2012; Li, Y 2012, for theoretical review on motivations behind self-disclosure). Moreover, people participate in those seemingly data-intensive activities such as social networking sites (SNS) for its social gratifications while avoiding social seclusion (Heravi, Mubarak & Choo 2018; Knausenberger, Hellmann & Echterhoff 2015; Raynes-Goldie 2010; Sheldon 2008; Taddicken 2012) or exclusion from certain services that such trade-off is too costly (Martin & Nissenbaum 2016). Some users also reported having disclosed more information in a bid to have more control of their online presence (Alaqra & Wästlund 2019). Worryingly, increasing reliance on technology has empowered companies to have "free reign in collecting and using information" leading to an "informational inequality" condition (Nissenbaum 2004) where consumer *surrenders* information without adequate understanding (Walker 2016). In light of this asymmetry, McDonald and Forte (2020) argued that privacy paradox is a reflection of modern consumers' powerlessness. As such, concern and disclosure are not necessarily dichotomous (Preibusch 2013), nor participating in self-disclosure implies a lack of concern (Raynes-Goldie 2012).

Masur and Scharkow (2016) suggested that SNS users tend to engage in disclosure management by sharing only perceived non-sensitive information appropriate for their audience, indicating different levels of *trust*. Other studies suggest, given a choice, users tend to disclose minimal personal information (King 2018) or at least minimum required for an app to function (King 2012). Sheehan (2002) and Nissenbaum (2010) argued that if the informational flow is context-appropriate, there is no paradox in having privacy concerns while participating in information sharing.

Previous studies suggested that privacy concern may not necessarily inhibit self-disclosure (Heravi, Mubarak & Choo 2018; Taddicken 2014) nor purchase decision

(Voicebot & Voicify 2019). Chen, HT and Chen (2015) suggested that users' self-efficacy or confidence in mitigating privacy risk (e.g. limiting their profile visibility) can encourage self-disclosure. Their result was suggested to be in line with privacy calculus model (Culnan & Armstrong 1999; Laufer, Prohansky & Wolfe 1973) which proposed consumer would assess the future consequences if they were to engage in certain behaviour, where one of the consequences can be a privacy invasion. This is consistent with the argument by Nissenbaum (2010, p.186) that consumers engage in self-disclosure as long it conforms to the *norms*, where the norm—in the context of SNS—is not sharing information with strangers. For instance, teenagers who spend a significant portion of their daily life on SNS might give off the impression that they are engaging in careless self-exposure (Barnes 2006). Previous studies (Gogus & Saygın 2019; Lin, J et al. 2014) observed that younger people tend to be less concerned about privacy.

Other studies, however, suggested that teenagers are also concern about privacy (Raynes-Goldie 2010). For example, some teenage users proactively leverage existing privacy controls offered by social networking sites to limit who could view their information (Ahn 2011). They may also be more aware of strategies to manage their privacy compared to adults (Brandtzæg, Lüders & Skjetne 2010), and employ privacy strategies such as using obscure wording that is only understood by peers and creating multiple profiles (Vickery 2014). Such behaviours show that people still engage in self-disclosure, despite holding privacy concern. We could also argue that a secret is not private when it is shared (in confidence) among close friends, but private to the unknowing outsiders.

Martin and Nissenbaum (2016, p.14) criticised that the notion of privacy paradox conflates "giving up privacy and giving up information". Information sharing does not necessarily mean losing privacy if the flow is appropriate. Laufer and Wolfe (1977) argued that the act of disclosing does not imply that the situation is invasive. Wacks (2010) argued that a person does not totally waive the claim over its personal information, even when the person voluntarily discloses it in the first place. King (2018) argued that due to lack of better alternatives, consumers are often left with companies with deficient privacy protections. This trend is also observed in the current proliferation of smart speaker. An industrial report (Voicebot & Voicify 2019) suggested that despite the majority of consumers exhibit some privacy concerns, the concerns do not significantly affect their adoption of a smart speaker. Certain disclosures are also increasingly

necessary to navigate in the modern networked world. Since it is no longer voluntary, some consumers would feel it is unreasonable for their data to be used outside of the original transaction (Sheehan 2002; Votipka et al. 2018; Waldman 2018). This sentiment is also echoed by Marshall (*Smith v. Maryland* 1979, 442 Supreme Court of United States 735 at 749), "[t]hose who disclose certain facts to a…company for a limited business purpose need not assume that this information will be released to other persons for other purposes".

2.3 Trust

Studies in social science and psychology explored how one's self-disclosure behaviour could be affected by different category of the target person (the person being disclosed) (Greene & Faulkner 2002; Serovich & Greene 1993; Serovich, Greene & Parrott 1992; Wheeless & Grotz 1977). Since the advent of social networking sites (SNS), researchers have also studied how SNS users categorise their "friends" or recipients (Johnson 2012; Kelley, Patrick Gage et al. 2011b; Norouzizadeh Dezfouli et al. 2015; Zhang et al. 2013). In mobile space, Shih, Liccardi and Weitzner (2015) examined with whom (e.g. family, friends, colleagues) users are comfortable in sharing their activity at a particular location. The study assumed the sender is the user and the recipient are the app. The research also concerns with users' willingness to disclose to certain apps for specific purposes.

Previous studies (Larzelere & Huston 1980; Wheeless & Grotz 1977) proposed trust as an influencing factor to self-disclosure behaviour. As trust has been a subject of interest throughout history regardless of disciplines, there have been numerous attempts to measure it, some articles we discussed can date back to 1967. Using the three-stage approach as described in this chapter's introduction, we initially identified publications in English using Google Scholar using search terms "trust scale (questionnaire OR measurement)". Second, we then reviewed the citations of those publications to identify additional potential articles. Finally, we used Google Scholar to identify other papers that cited the relevant articles identified previously. We then choose articles that are related to trust scale. We gathered publications up until 04/2018.

As part of our search for appropriate trust scales, we identified the following highly cited publications:

1 Interpersonal Trust Scale (Rotter 1967) (RITS)

2 Individualized Trust Scale (Wheeless & Grotz 1977) (WITS)

3 Dyadic Trust Scale (Larzelere & Huston 1980) (DTS)

4 Specific Interpersonal Trust Scale (Johnson-George & Swap 1982) (SITS)

Trust is not just limited to interpersonal communication or social purposes. We use mobile devices to perform everyday tasks like conducting business, banking and education. These tasks often involve communicating with an organisation which would involve some form of trust. Our questionnaire also asks participants to list the organisations they interact with, in addition to persons. Trust on people is arguably different from the trust on organisations as they constitute a different level of trust. We could use two levels of scale each for its subject (people and organisation), but the results from two separate scales may be too distinct to be interpreted together. We found WITS to be sufficiently generic to measure trust on people and organisations, whereas DTS and SITS are better suited for personal trust only.

Wheeless and Grotz (1977), Larzelere and Huston (1980) argued that previous studies did not find a significant relationship between trust and self-disclosure because those studies measured generalised trust (RITS) instead of trust to a person. RITS (Rotter 1967) aimed to assess the *generalised* trust of an individual towards others. It asks participants how much they trust others such as friends, parents and the world, and how optimistic they are towards the society. It is arguably one of the motivations behind WITS, DTS and SITS as they discuss RITS. As this study aims to measure how the trust level differs across different groups of recipient, we find generalised trust measured in RITS not suitable for our main purpose.

Several previous studies are closely related, but dissimilar in a way their assumption on trust. In the present study, we measure how much sender trusts recipient, instead of trust on commercial entities (Ginosar & Ariel 2017) or the survey conductor (Joinson et al. 2010). This is more in line with the CI framework by having the trust between the sender and the recipient as the focal point.

The measurement of the trust level is somewhat similar to *tie-strength*. Tie-strength is used to quantify how *close* a person to another, especially in social media research (Fogues et al. 2018). A previous study (Wiese et al. 2011) suggests tie-strength plays a role in disclosure behaviour. While tie-strength could imply a level of trust, we note the distinct methods for each measurement.

2.4 Self-Disclosure

Privacy studies often focus on the relationship between privacy concern and willingness to disclose (information). There have been numerous studies (Kokolakis 2017) conducted to examine the relationship between privacy concern and willingness to disclose information, in which some results suggest the phenomenon of privacy paradox. However, Nissenbaum (2010) argued that if the informational flow is context-appropriate, there is no paradox in having privacy concerns while participating in information sharing.

While contextual factors have been considered in privacy studies, their relationship to privacy concern and information disclosure behaviour are often studied independently. Malhotra, Kim and Agarwal (2004) examined how privacy concern can be affected by the contexts, especially by the type of information. Shih, Liccardi and Weitzner (2015) examined the influence of various factors (i.e. app, type of location and purpose) on the willingness to disclose. In this work, we investigate how willingness to disclose can be affected by an individual's privacy concern across different contextual factors. We also expand on Shih, Liccardi and Weitzner (2015) by measuring the full range of contextual factors (i.e. sender, recipient, subject, attributes and transmission principle) as proposed in the framework of contextual integrity (Nissenbaum 2010).

Few studies have examined the influence of contextual factors on privacy expectations. Martin and Nissenbaum (2016) surveyed the effect of type of information used in different situations or contexts on privacy expectation. In each context, they also studied the difference in privacy expectation when a specific type of information is used in relation to the context or commercial use. The result showed the respondents respond positively to privacy expectation even when the information is considered 'sensitive', as long it is related to the context. Most of the respondents, including those considered as 'privacy unconcerned', reacted negatively when the information is used solely for commercial purpose. This result suggests the limitation of the sensitivity of the information. Another study (Nicholas et al. 2019) in the healthcare field also exhibits a similar result. In that study, the sensitivity of the health sensor data did not significantly affect the willingness to disclose. The participants instead considered the contextual factors—data type and recipient—as essential factors. This result also suggests the relevance of those contextual factors, which are the focus of this thesis.

There are countless privacy norms that exist within a context, and it can be challenging to enumerate them. Shvartzshnaider et al. (2016) proposed a context discovery method using crowdsourcing. Focusing on educational context, the study examined how different combinations of contextual factors can affect privacy expectation. The study did consider all the essential contextual factors, despite being limited to just a context. During the study, each participant was shown various hypothetical situations or 'vignettes' formed by a different subset of contextual factors and evaluate whether they meet privacy expectation. Crowdsourcing is particularly useful because it can scale to many combinations of contextual factors and available to a broad audience. Privacy norms are discovered depending on how positively (or negatively) participants respond to each combination. A similar study is conducted to discover norms in the context of a smart home (Apthorpe et al. 2018).

Martin and Shilton (2016a, 2018) examined the difference in mobile device user's privacy expectations when data is used for its intended purposes and when it is used for commercial purposes (i.e. tracking and targeted ads). They also utilised crowdsourcing to distribute a variety of vignettes similarly with Shvartzshnaider et al. (2016). The study is conducted mainly by manipulating the purpose or transmission principle of each vignette, along with other contextual factors. Unlike other studies, the sender and recipient are the same organisation that collect user data using the apps, often in the background without the user's knowledge (FTC 2013). The results suggested nuances of privacy expectation in different use scenarios, especially in commercial usage. Users expect certain data types to be used for their intended purposes, such as location data for navigation apps. The study also observed that, while users are generally uncomfortable with the commercialisation of their data, they do recognise the benefit of keyword harvesting in improving targeted ads. Such nuance suggests that consumers are not necessarily against commercial usage. Another study (King 2018) observed similar behaviour, whereby most participants responded they could accept aggregate data collection as long as the purpose is not for targeted advertising.

The study conducted by Martin and Shilton (2016a) is expanded in a later study (Martin & Shilton 2016b) to investigate the effect of mobile app usage on the relationship between contextual factors and privacy expectation. The result suggests that contextual factors have more influence in judging privacy expectation on users with more experience (in using mobile apps).

2.5 Privacy profiling

Existing mobile devices have privacy issues when an app requests or collects more data than necessary to perform any given function. Permission control is a common approach (as found on Android and iOS platforms) utilised by PET to minimise the data collection. To ease the user's burden in permission management, privacy profiling and recommendation systems have been proposed by researchers. In privacy profiling, a person is typecast into a particular category based on certain characteristics which resemble Westin's categories (in Kumaraguru & Cranor 2005). In this section, we discuss the Westin's categories and the drawbacks.

In a series of surveys conducted by Westin (in Kumaraguru & Cranor 2005), he proposed that consumers can be categorised into three categories in their privacy concerns: fundamentalists, pragmatists and unconcerned. Fundamentalists are like high-privacy oriented proponents, concerned with all forms of privacy violations, regardless of benefits. "Unconcerned" have no privacy expectations and readily share their personal information. Pragmatists would weigh the trade-offs between the benefits of available services and the amount of required personal information disclosure. In the survey, pragmatists account for 50% of respondents, 25% for fundamentalists, and the other 25% are "unconcerned". Westin (in Committee on Energy and Commerce 2001) noted that fundamentalists are outliers and policy should be catered for the majority pragmatists.

Despite the influence of Westin's surveys, Martin and Nissenbaum (2016) argued that pragmatist and unconcerned had been misinterpreted to support the notion that they are willing to sacrifice privacy in return for benefits (e.g. free or discounted service or app). They also claimed that the misinterpretation goes as far as to assume disclosed information is no longer private and does not warrant privacy consideration on how the information is subsequently used.

The segmentation of consumers (as proposed by Westin) is arguably influential in the development of "notice and choice" approach to privacy regulation. In this approach, the consumer is expected to read privacy policies and choose whichever services that are consistent with the consumer's privacy expectations. Westin (in Kumaraguru & Cranor 2005) argued that most consumers are a pragmatist, constantly weighing choices based on the cost and benefit and make rational decisions. However, Martin and Nissenbaum (2016), based on their survey, offered an alternative view in which consumer across those

categories could share a similar view on what constitutes a privacy violation, and that even 'unconcerned' can identify its occurrence. This suggests the limited use of Westin's scale outside of its initial setting in understanding consumer's privacy expectation (King 2014), especially when facing advancing technology that increasingly challenges the boundary of social norms.

Terpstra et al. (2019) outlined three issues with the "notice and choice" approach. First, consumers often receive a "notice" via privacy policies, but privacy policies are notorious for having an excessive length and legal jargons that are intimidating to the average users. Yet, despite the details, privacy policies do not necessarily inform the users—including the 'pragmatist'—sufficiently to make better decisions.

Second, the 'pragmatist' ability to make decisions might have been overestimated. Hoofnagle and Urban (2014) suggested that there is a substantial deficit in consumers' awareness of the actual business practices. They also tend to underestimate the scope of data collection, while overestimating the legal regulation of the marketplace. Opaque and unfair business practices can prevent consumers from making rational, informed decisions as 'pragmatists' (Wu, Vitak & Zimmer 2019).

Third, consumers often left without any meaningful choice. A "choice" of a privacy-friendly option could induce more costs than benefits (Martin & Shilton 2018). Westin (in Committee on Energy and Commerce 2001) argued that the pragmatists have a significant influence in the marketplace, predicting the companies would eventually adapt their business practice to cater for pragmatists' preference. However, businesses are incentivised not to ask consumers whether they would prefer more privacy options, or outright do not provide them. Some companies attempt to nudge users away from privacy options or threaten them with a loss of functionality if those options are enabled (Kaldestad & Myrstad 2018; Mathur et al. 2019). This is because businesses assume to have less profitable consumer data (to be sold to third-party) if privacy options are in place. Even when the options are available, a majority of users struggle to locate them (Boyd 2019; Habib et al. 2020; Habib et al. 2019). Difficulty in privacy management could also lead to more information disclosure (Mourey & Waldman 2020), thereby perpetuating the vicious growth of deceptive options. The choice is also seemingly vanishing to growing dependence on technology that functions on information disclosure (Waldman 2018). While this information disclosure—consciously or not—is

"theoretically voluntary…but the costs of refusal are high and getting higher" leading consumers into "a prison that, although it has no walls, bars, or wardens, is difficult to escape" (Brunton & Nissenbaum 2019). Hence, this puts most of the consumer in a disadvantaged position to negotiate privacy in the marketplace, particularly in the mobile device ecosystem.

Ironically, having too many options is not optimal either. Excessive choices can be a significant cognitive burden that subsequently causes decision fatigue to the user (Svirsky 2019; Utz et al. 2019), thereby not exerting control over technology. This deficiency increases the risk of developing "self-censorship" where they withhold something due of the fear of a privacy breach (Misra & Such 2015) or even "learned helplessness" behaviour whereby they "stop responding to invasions even when presented with ways to defend themselves" (Shklovski et al. 2014). Further increasing the risk is the tendency by the media to cover privacy solutions offered by tech companies, to shift the focus towards maximization of the benefits (behind disclosure) and ignore the underlying privacy issue (Popiel 2019). Worse, privacy discourse is increasingly shaped by advocacy groups that are heavily funded by tech companies (Stoller 2019).

Despite the demand of consumers for more control over their data and privacy, "the (digital marketing) industry has largely failed to regulate itself by consistently implementing privacy controls in broad strokes, overlooking the intricacies of consumer needs and information sensitivities. Some of this has been naiveté, while some has been the deliberate and explicit stretching of the boundaries of individual expectations and the implicit social contracts with consumers" (Doorey, Wilcox & Easin 2017).

Sheehan (2002) suggested (along with other factors) that consumers become more concern whenever the data collected is used for a secondary purpose. Hoofnagle and Urban (2014) proposed that the Westin's three categories of consumers can be regarded as two groups instead; the fundamentalists who are more knowledgeable on the actual marketplace practice, while pragmatists and unconcerned who are less informed.

Existing mobile devices have privacy issues when an app requests or collects more data than necessary to perform any given function. Despite the implementation of run-time permission control in Android 6.0 (Amadeo 2015) and app-approval process for its app marketplace (Kim 2015), the information leakage is still widespread (Bosu et al. 2017; FTC 2013; Grace et al. 2012; Ren et al. 2018; Seo et al. 2016; Snoopwall 2014;

Stevens et al. 2012). Previous studies suggested that more fine-grained or flexible permission management can better reflect the users' privacy preferences (Ahern et al. 2007; Benisch et al. 2010; Fang, Han & Li 2014; Kelley, Patrick Gage et al. 2011a). However, such flexibility comes at the cost of user-friendliness (Felt et al. 2012b; Liu, B et al. 2016).

To ease the user's burden in permission management, researchers have proposed privacy profiling and recommendation systems. Despite recognising the diversity of individual privacy preferences, crowdsourced preferences that show percentage of users approved (Bokhorst 2016; Liu, R et al. 2015), confidence level (Rashidi, Fung & Vu 2015) or simply recommend the settings (Agarwal & Hall 2013; Bokhorst 2016) for each permission is still a form of persuasion to the users that those are the "correct" privacy settings. Later study also cast doubt on the practicality of privacy recommendation systems as they tend to require a large sample size for bootstrapping (Warberg, Acquisti & Sicker 2019). This leads to a catch-22 whereby the sample size requirement could only be afforded by the largest institutions, the very targets that those systems are supposed to protect consumers from.

In profiling, a person is typecast into a particular category based on certain characteristics (Vaidya & Atluri 2008). The user categories resulted from privacy profiling studies in mobile space (Knijnenburg 2014; Lin, J et al. 2014; Liu, B et al. 2016) resemble Westin's categories (in Kumaraguru & Cranor 2005): conservative, unconcerned, pragmatist. Privacy profiling attempts to be flexible to the variety of user's privacy preferences. When privacy profiling is applied to permission management, the 'unconcerned' profile often results in highly permissive settings, rendering the permission manager ineffective. As mentioned earlier, the user exhibits 'unconcerned' behaviour due to lack of awareness of the scope of data collection by apps and websites, and the actual business practices. More importantly, a study (Martin & Nissenbaum 2016) has also suggested that consumer across those categories could share a similar view on what constitutes a privacy violation and that even 'unconcerned' can identify its occurrence. When user privacy is at stake, suggesting a permissive setting seems unwise.

Despite recognising the diversity of individual privacy preferences, crowdsourced preferences that show the percentage of users approved (Bokhorst 2016; Liu, R et al. 2015) or confidence level (Rashidi, Fung & Vu 2015) for each permission is still a form

of persuasion to the users that those are the "correct" privacy settings. Complicating the matter is the fact that user also acquiesces to discloses private information, in exchange for various benefits (Grossklags & Acquisti 2007; Hann et al. 2007; Kehr et al. 2015; Lin, K-Y & Lu 2011; Muhammad, Dey & Weerakkody 2017).

2.6 Dichotomy of privacy

The bulk of the privacy studies in mobile space regard 'privacy' as dichotomies—sensitive and non-sensitive, risky and non-risk, private (personal) and not-private—where only one halves warrant privacy consideration. In the mobile platforms, users are usually prompted with consent dialogue whenever an app request for 'sensitive' data for the first time (ask-on-first-use).

Classifying the sensitivity or riskiness of information introduces a new issue. Sensitive information is often predefined in the discussion of PET. However, what information constitute as *sensitive* is subject to the users' varying privacy preferences and may also vary according to circumstances. This suggests that relying on predefined sensitive information may be impractical in serving a wide user base. Sensitive information is also often predefined by the respective mobile platform. For instance, Android 6 saw the separation of permission into *dangerous* and *normal*. Despite the implementation of ask-on-first-use permission system in that version—an improvement to the previous ask-on-install approach—it only applies to dangerous permissions while normal permissions are automatically granted. Later studies uncovered hidden gaps that pose severe privacy risks to the users (Alepis & Patsakis 2018; Kywe et al. 2016).

Sensitive information may be defined as information that "may result in harm to its subjects" (Nissenbaum 2010, p.124) once disclosed. However, predicting which type of information can inflict harm is subjective and may not always consistent (Fazlioglu 2019; Martin & Nissenbaum 2016). Similarly, The OECD Privacy Framework (OECD 2013) also clarified that certain data could become sensitive depending on the context and use, despite not being so at first glance. Even classification of *private* information is also problematic, whereby "the same information may be regarded as very private in one context and not so private or not private at all in another" (Wacks 2010, p.45).

As Solove (2006, p.486) cautioned "using the general term 'privacy' can result in the conflation of different kinds of problems…". To avoid this issue, the discussion of

'context' should be included in any privacy study. Previous studies (Hoofnagle & Urban 2014; Martin & Nissenbaum 2016; Martin & Shilton 2016a, 2016b; Shih 2015) suggest that participants react differently whenever the context changes, e.g. privacy 'unconcerned' express concern whenever they are informed the subsequent use of their data. "Almost everything—things that we do, events that occur, transactions that take place—happens in a context" (Nissenbaum 2004, p.119). Thus, any data collection is always within a particular context and to serve a specific purpose(s). Participants express more concern than they were before, after being given explanations on the *potential* uses of information (Liu, B et al. 2016; Shih, Liccardi & Weitzner 2015; Toch 2012).

2.7 Framework of Contextual Integrity

In a draft Consumer Privacy Bill of Rights proposed by the Obama administration, one of the principles is:

> Respect for Context: Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data. (White House 2012)

This principle is consistent with the argument outlined in the CI framework (Nissenbaum 2010) that consumers feel unease whenever they believe to be operating in one context, only to discover it has been used in another context. The resulting uneasiness is the result of the conflict of contexts.

Our everyday lives are governed by a plethora of norms. Nissenbaum (2010, p.137) refers to the appropriateness of the flow of personal information as "informational norms". The CI framework emphasises that "there are no arenas of life *not* governed by *norms of information flow,* no information or spheres of life for which *anything goes*" (Nissenbaum 2004, p.119). Contextual integrity "is preserved when informational norms are respected and violated when informational norms are breached" (Nissenbaum 2010, p.137).

Contexts, actors, attributes and transmission principles are the key factors in shaping the informational norms. The framework evaluates, in a given context, which *sender* (*actor*) can share what type of information (*attribute*) with which *recipient* (*actor*) regarding whose information (*subject*) under certain conditions (*transmission principles*). The following list summarises the contextual factors.

5 Sender (e.g. app, app category)
6 Recipient (e.g. friend, family, company)
7 Subject of information (e.g. user itself, friend, family)
8 Type of information (e.g. contacts, location, multimedia files, documents)
9 Purpose of information collection (e.g. public transport, friend finder, health monitoring)

For example, in a financial context, a financial institution might be required to submit a client's credit history to a credit bureau but not to the client's employer, unless with client's explicit consent (a transmission principle). The healthcare and employment contexts are merely a few notable examples. There could be countless contexts and enumerating them is outside of the scope of this research.

Contextual integrity can be useful in evaluating a novel practice. We could perceive a novel situation as "creepy" whenever it does not line up with our social norms (Tene & Polonetsky 2014). This is despite, on the surface, that the novel practice does not seem to be violating existing laws. For instance, Facebook was discovered to engage in inappropriate social listening, "the analysis of social media content to understand user sentiments", (Sponder, cited in Tene & Polonetsky 2014) to help advertisers to estimate users' (i.e. teenagers) emotional states. While users could be aware of targeted or personalised advertising, this sort of behaviour is probably distasteful.

Nissenbaum (2010, p.144) proposed the following decision heuristic to evaluate a novel system or practice and help understand the source of privacy issues (of the system, if any):

1 Determine the social context(s) to help identify the norms. Examples given in the original text include "a grade school in an educational context; a hospital in a healthcare context; a department store in a commercial marketplace".
2 Determine the key actors (i.e. recipient, subject and sender). The new system could have more recipients.
3 Determine the attributes. The new system could have more types of information.
4 Determine the principles of transmission. New practices may entail a revision in the principles governing the transmission of information from one party to another. The new system could mandate information sharing that was optional previously.

5 Red flag. Changes in the parameters listed above "flags the need for further examination and evaluation" (Nissenbaum & Patterson 2016).

Nissenbaum (2010, p.161) noted that the decision heuristic seems to favour entrenched norm by "flagging as problematic any departure from entrenched practice" and such issue is apparent when trying to apply the framework to newer technologies. Nissenbaum (2010, p.161) expanded the decision heuristic with the following elements:

1 Prima facie assessment… A breach of information norms yields a prima facie judgment that contextual integrity has been violated because presumption favours the entrenched practice.
2 Evaluation I: Consider moral and political factors affected by the practice in question…
3 Evaluation II: Ask how the system or practices directly impinge on values, goals, and ends of the context…
4 By these findings, contextual integrity recommends in favour of or against systems or practices under study…

Thus, when considering the above elements, it is possible that novel practice might be preferred, even when there is a red flag that signifies that the novel practice violates the entrenched norm. This is because the new practice might improve upon existing practice on attaining ends, values, and purpose, or generally promote the greater good, while remaining morale.

Regarding the application of the decision heuristic, Nissenbaum (2010) briefly reviewed a few technologies (e.g. library management and caller ID) to illustrate how can decision heuristic be applied to evaluate new technologies. Since then, scholars have given more reviews in the computing field. Barkhuus (2012) argued the relevance of the framework in the age of apps that can share a large amount of information, opportunities for which were previously limited. Drawing on the example of mobile social media where the user is encouraged and motivated to share information, the author suggests that privacy study should focus more on why people disclose, rather than on what they disclose. Raynes-Goldie (2012) also argued for the relevance of the framework on social media. Zimmer, M (2008) reviewed on Google's practices based on the framework to illustrate the possible privacy threats. Grodzinsky and Tavani (2011) focused on the possible privacy implications of cloud technology specifically on Google Docs. They

have also weighed in on the users' privacy expectations when blogging and using peer-to-peer (P2P) file-sharing (Grodzinsky & Tavani 2008, 2010). These examples illustrate the practicality of the CI framework on the latest technologies (e.g. mobile device and cloud computing). Krupa and Vercouter (2012) proposed a framework to protect user privacy in a decentralised virtual community in a P2P network, which lacks a central authority to enforce privacy like existing social networking sites (e.g. Facebook and Instagram). A more in-depth discussion on literature related to the framework can be found in Benthall, Gürses and Nissenbaum (2017).

The CI framework is also relevant in the mobile space. Barth et al. (2006) proposed a logical framework to formalise the expression of the norms of information flow using logical notations. This is useful in translating the norms (expressed in natural language) into a programming language. The logical framework expresses an information flow either in positive or negative norms. These two norms are akin to the allow and deny rules in conventional access control schemes; a positive norm allows the information flow while a negative norm denies it unless certain conditions are met, depending on the context. In mobile computing field, Shih, Liccardi and Weitzner (2015) proposed ContextProbe, a framework that utilises quantitative and qualitative research methods to investigate the effects of contextual factors on users' privacy preferences when using apps. In Android platform, Liu, R et al. (2016) and Liu, R et al. (2015) proposed PriMe and PriWe that measure and mitigate privacy risk respectively. PriMe considered the inherent sensitivity of specific data and individual sensitivity to that data with a particular focus in mobile participatory sensing (e.g. Mass and Madaus (2014)). PriWe is a crowd-powered privacy recommendation system to provide decision support in using the permission manager. Baokar (2016), Wijesekera et al. (2015) and Tsai et al. (2017) proposed permission management that utilises machine learning to predict users' privacy preferences.

Wijesekera et al. (2015) conducted a study to investigate how often apps request resources unexpectedly. Users were interviewed whether the requests were expected after being shown with screenshots taken whenever an app requests sensitive resources. The sensitive resources predefined in the study were previously determined by Felt et al. (2012a). There is a proposal (Jia et al. 2017) to provide contextual integrity in the Internet-of-Things (IoT) platforms, which is relevant to the mobile space as IoT platforms are also developed by the companies that develop mobile platforms. However, it focuses on

improving the security of IoT platforms that it is more akin to context-aware access control than the CI framework. Security should not be confused with privacy since each has different goals (Camp 1999) and could interpret the term differently (Benthall, Gürses & Nissenbaum 2017). We will briefly discuss context-aware access control later in Section 2.8.

The studies all similarly argued the necessity of identifying the context surrounding data flows to meet privacy expectations. They have a slightly different approach to the CI framework, specifically on how they determine the context(s). Shih (2015, p.33) refers to context as the "representation of people's physical surroundings with the subjective interpretations they attach to it". Wijesekera et al. (2015) and Tsai et al. (2017) refer to context as "visibility (foreground or background) of the requesting application and the frequency at which (data) requests". Shih (2015) refers to it as whether the user is using it actively or not. This interpretation is arguably similar to Wijesekera et al. (2015) and Tsai et al. (2017), as an app is 'visible' to the user during active usage and switch to the background when the device is idle (switched on while the display is off).

Despite their similarity, those studies considered contextual factors differently. We summarised our findings in Table 1.

Table 1 Comparison of current literature that adapted the CI framework.

The symbol ● denotes the contextual factor is considered by the literature, ○ denotes the contextual factor is considered but with limited attributes, and × denotes the contextual factor is not considered.

	Sender	**Recipient**	**Subject of information**	**Type of information**	**Purpose of information collection**
(Martin & Shilton 2016b, 2016a)	●	○	○	●	●
(Tsai et al. 2017; Wijesekera et al. 2015; Wijesekera et al. 2017)	●	×	×	●	×
(Shih, Liccardi & Weitzner 2015)	×	●	●	●	●
(Shklovski et al. 2014)	○	○	×	●	○

2.7.1 Sender/ Recipient

Sender and recipient are assumed in the contextual integrity framework (Nissenbaum 2004) to possibly consists of single, multiple individuals or organisations. This research involves studies of the relationship between mobile users and their contacts—which can be a group of relatable people or organisations. While other related works (Martin & Shilton 2016b; Shklovski et al. 2014; Tsai et al. 2017; Wijesekera et al. 2015; Wijesekera et al. 2017) tend to assume a mobile app as the sender, such assumption is irrelevant to our purpose. Another study (Shih, Liccardi & Weitzner 2015), meanwhile, assumed the sender is the user and the recipient is the apps.

2.7.2 Subject of information

The myriads of information stored in our mobile device, especially through chat and social media apps, would inadvertently contain other subjects. Not to mention the prevalence of multi-user functionality in modern mobile platforms (Cunningham 2016; Google 2017). However, for the purpose of this research project, we currently focus on the single-user personal mobile device; hence, the sender (user) and the subject of information would be the same individual most of the time.

None of the studies we have discussed explicitly mention the "subject of information". We can assume that most studies also adopted similar assumption of the sender and subject being the same person., There is another aspect that could lead to information that has subject other than the user as shown by Martin and Shilton (2016b), in addition to multi-user. While not explicitly mention the subject, the study mentioned "friend's activity" as a type of information. "Friend's activity" can be found in social sharing function that is increasingly common in most categories of apps, in addition to social media apps. For example, we can share our current playlist with one another on a music app. This illustrates that the subject of the information is not always just the user.

2.7.3 Type of information

Equally important to the subject (*who* the information is about) is the attribute (*what* the information is about). It is more commonly known as the *data type* in computer science. In the mobile field, mobile platforms typically offer permission management which allows user grant or deny access to specific data types by an app. This is the most common approach in PET in tackling user privacy.

2.7.4 Purpose of information collection

The most crucial element of the framework is the transmission principle (Nissenbaum 2010) or the purpose of information collection. It puts constraints on the information flow and largely explains why information transfers ought (or ought not) to occur. Examples including a person could be legally compelled to disclose specific information, information could be essential to provide certain services (e.g. healthcare), or in mobile platforms, information that is necessary to provide an app's functionality (e.g. current location when using a mapping app). Despite its significance, Nissenbaum (2010) cautioned that "the list (of transmission principles) is probably indefinite, particular if we allow for nuanced and complicating variations". The effect of such nuanced nature is that it is not as straightforward to implement compared to other parameters. We interpret it as "purpose of information collection" to offer better clarity to our discussions. We do note the work of Barth et al. (2006) that offered a novel interpretation of the transmission principle by considering past decisions and mandating additional actions in the future. An example given is a past requirement of confidentiality. This is also considered by Shvartzshnaider et al. (2016), in addition to the subject's knowledge and consent. These could be considered as part of our future works.

Most of the studies we have examined considered the purpose of information collection. Shklovski et al. (2014) partially considered it as the study only focused the excessive data collection by flashlight or gaming apps in which the collected data is later sold to third-party companies. Despite the importance of transmission principle, implementation in the mobile platforms is still lacking. Mobile platforms often recommend app developers to display the purpose of a permission request to the user, but such practice is not compulsory (Apple 2013; Google 2018). Even when it is shown, its authenticity could not be verified. Such verification could one day be introduced via the advancement of natural language processing—driven by the growth of artificial intelligence-powered virtual assistant (e.g. Google Assistant, Apple's Siri, Amazon Alexa and Microsoft Cortana)—and empower the user in making privacy decisions.

2.8 Discussion

The concept of privacy is ever-evolving. There is yet a consensus on the definition of privacy (Tavani 2008). Nissenbaum (2010) argued that attempting to define would risk of having "vagueness and internal inconsistency". As interpreted by Raynes-Goldie (2012), the framework of contextual integrity essentially recognises "the variedness of privacy definitions and understandings are part of the very nature of privacy". Thus, contextual integrity does not intend to define privacy; instead, the purpose is to serve as a benchmark for evaluating novel practices against the entrenched norms through the aforementioned decision heuristic (Nissenbaum 2010; Zimmer, M 2005).

We discussed the importance of 'context' in the privacy-related discussion. The exclusion may result in a conflation of different kinds of problems (Solove 2006). In physical privacy, according to the Merriam-Webster dictionary, privacy is defined as "the quality or state of being apart from company or observation". However, if privacy is simply being secluded, this definition would conflate the notion of "having privacy" and "being let alone" (Tavani 2008). As discussed previously, a person can have a sense of solitude, yet its privacy is violated in some way. In informational privacy, some argued that a person's privacy is proportional to its personal information that others possess. Such an argument would conflate "giving up privacy" and "giving up information" (Martin & Nissenbaum 2016). Our everyday lives often involve giving a certain amount of information in exchange for service. In some cases, we are even culturally *expected* to do

so. However, this does not mean we waive the claim over that information (Wacks 2010) and expect them not being used for secondary purposes.

Some research on PETs is based on the result of the previous user surveys. Those surveys investigated user's privacy concern on information collection and sharing. The results from those surveys enable the researcher to evaluate the effectiveness of PETs in addressing those concerns. For example, the PET designed by Liu, B et al. (2016) is partially based on the results of Felt, Egelman and Wagner (2012) and Lin, J et al. (2012). However, much of inquiry of those surveys are based on hypothetical scenarios or personas and often lacks reference to people's real experience (Shih 2015), which could inflate an individual's concern. Later studies (Martin & Nissenbaum 2016; Martin & Shilton 2016a) suggested that privacy concern is often context-specific and continuously changing according to an individual's particular needs and desires. We also discussed the limitations of existing privacy profiling (due to variable context) that could subsequently affect the effectiveness of the recommendation system.

Most studies on PET lack a pre-design user study or prototype user evaluation. This limitation can impede on meeting the actual user's privacy expectation and subsequently limited in describing the *effectiveness* of a PET. Notable exceptions include Felt (2012) that surveyed which permissions that users perceive as risky and design a PET based on the result; other studies (Liu, R et al. 2015; Tsai et al. 2017) had the users evaluated the prototype PET. We argue it is imperative to gather user requirement for a more relevant system.

Despite the ubiquity of permission management system, we argued that privacy preservation through the system alone not optimal due to the gap between flexibility and usability. Fine-grained permission managers have been proposed to enable more flexibility in configuring permissions; however, the plethora of options can overwhelm the average mobile user. To simplify configuration for end users, crowd-powered privacy profiling and recommendation systems have been proposed. These systems advise the user of the percentage of other users who granted arbitrary permission, at the time that they are requested. However, merely showing this percentage on each permission is still a form of persuasion to the users that those are the "correct" privacy settings, despite the diversity of users' privacy preferences.

Numerous research in Android platform has adopted the CI framework (Liu, R 2015; Shih 2015; Tsai et al. 2017; Wijesekera et al. 2015) to justify the necessity of identifying the context surrounding data flows to meet privacy expectation. Despite having a finer-grained of data access control (i.e. more permissions), those studies lack the contextual factors described in the CI framework, as shown in Table 1. This is possibly due to the induced cost of complexity causing the system to be less intuitive to use. While having the system automatically detect the context is a natural answer in reducing complexity, Wijesekera et al. (2017) conceded that such an approach could be very challenging. Barth et al. (2006) adapted the framework into logical expressions that could assist in its integration into a mobile platform; their practical applications can be explored further in the future. While the studies we have examined adapted most of the contextual factors, we posit that adopting the decision heuristic would have made a more thorough application of the CI framework.

The term context has been used in the literature on Android security without any connection to the CI framework. Context-aware access control (Abdella, Özuysal & Tomur 2016; Bai et al. 2010; Chakraborty et al. 2014; Conti, Nguyen & Crispo 2011; Jung, K & Park 2013; Miettinen et al. 2014; Rohrer et al. 2013; Zhauniarovich et al. 2014) is designed to trigger a particular security policy whenever the user is at a specific physical surrounding. The concept of context-awareness in those studies is most likely adapted from ubiquitous computing field (Osbankk 2007).

We noticed certain privacy concepts or enhancements proposed by the scholars are eventually implemented into real-world products (whether by purpose or coincident):

- As part of Google's announcement on Project Strobe (Smith, B 2018), it explained the new restrictions on the developer API access to the users' Gmail data is due to the observation, "When users grant apps access to their Gmail, they do so with certain use cases in mind". This further underscores the importance of accommodating the changes in contexts to avoid privacy issues, as posited in the CI framework.
- Google also observed in the Project Strobe that, "people want fine-grained controls over the data they share with apps", which echo with the researchers' call for more fine-grained permission (Jeon et al. 2012).

- The proposal by Felt et al. (2012a) to replace ask-on-install with ask-on-first-use privacy management mechanism is eventually implemented in Android 6 (Amadeo 2015).

Researchers observed that a majority of users are uncomfortable with background data collection (Jung, J, Han & Wetherall 2012; Tsai et al. 2017; Wijesekera et al. 2015). Thompson, C et al. (2013) proposed an enhancement to the user interface by having a persistent notification when a background app is collecting data. This enhancement is later implemented in Android 8 (Amadeo 2017). iOS 11 (Selleck 2017) and Android 10 (Burke 2019) further enhanced it by having the ability to grant permission only to foreground (visibly running) apps. Android app that requires background location access need to get an approval from Google prior to publishing at the Google Play (Android app market), from November 2020 onwards (Vitaldevara 2020).

2.9 Research Design

There are active discussions surrounding the phenomenon of privacy paradox across different fields. Scholars theorise the cause through the lens of the CI framework but do not examine the framework's practicality. In this thesis, we seek to elucidate privacy paradox based on the framework through two quantitative studies. In the first study, we examine privacy paradox by investigating the influence of recipient and type of information on mobile device users. Privacy literature mostly focuses on the effect of privacy concern and trust on self-disclosure. These three factors are the main variables of this user study. Although their effects are well-established, they are often studied independently (Martin & Shilton 2016b). To address this gap, in Study 1, we investigate the effects of trust between different groups of recipient on the relationship between privacy concern and self-disclosure.

Existing studies have shown users often assess an information flow based on diverse contextual factors. A series of studies (Lin, J et al. 2014; Lin, J et al. 2012) showed a significant influence of *purpose* on users' subjective judgement. This is also in line with Zimmer, JC et al. (2010) that showed users are more willing to disclose information when it is perceived to be *relevant* to the function provided by the receiving service provider. These studies, in a way, also suggest users are increasingly demanding mobile apps to be more upfront about information request. This is evident in a study (Wijesekera et al. 2017) where the results suggest users consider app visibility as an essential factor in deciding

on permission request, as users are usually not comfortable with an app collecting data in the background. A study on personal health data (Nicholas et al. 2019) showed participants considered not only the recipient but also the data type before disclosure. The result is also in line with Martin and Nissenbaum (2016) which showed the influence of the type of information, contextual actor (recipient) and purpose of information; the study also showed 'sensitivity' is subjectively influenced by contextual factors.

In Study 2, we examine additional factors that can affect an information flow between sender and recipient; we investigate the relationship of data type and its relevance on the willingness to disclose to specific groups of recipients. Distinct from another similar studies (Marmion et al. 2019; Martin & Nissenbaum 2016) which utilize generic data types, our study is more specific to mobile device usage where we derive data types from mobile users.

This thesis aims to gain insights of mobile users' privacy attitudes by examining the factors that are involved in the information flow between sender and recipient. Figure 1 illustrates how we will approach our user studies.

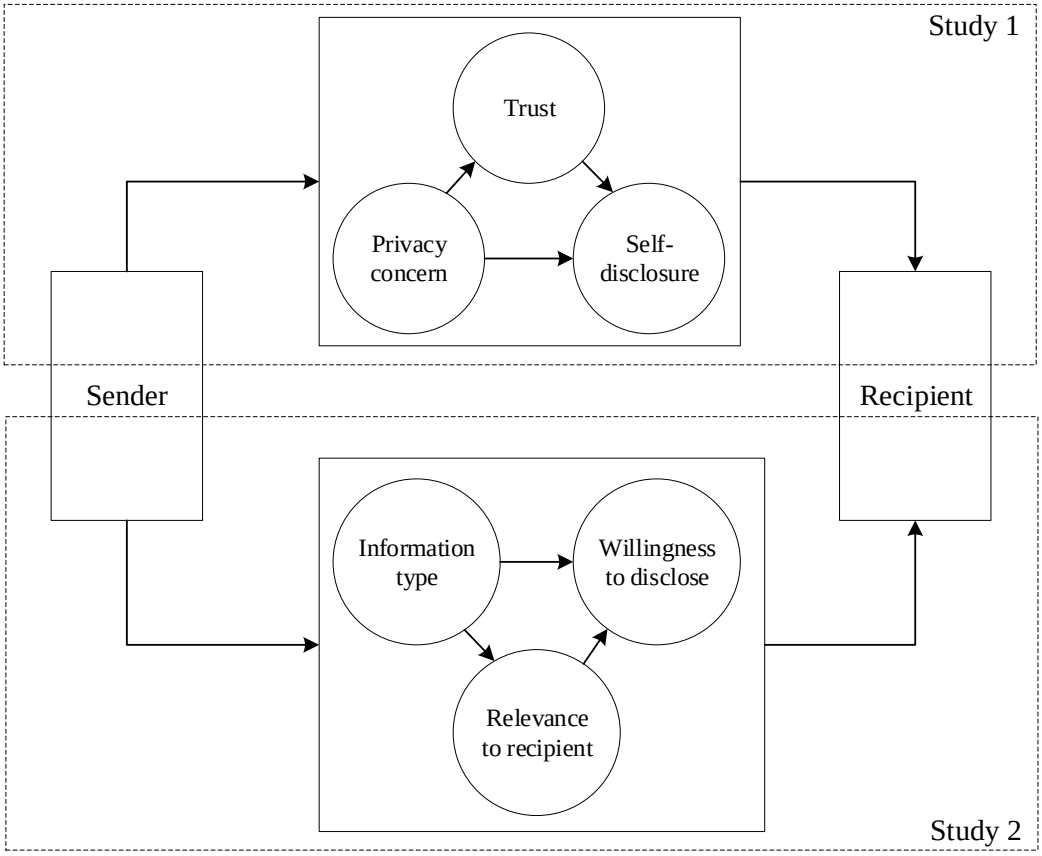

Figure 1 Contextual Privacy Framework

2.10 Conclusion

This work is part of a project in a bid to address the shortcomings of the existing studies, particularly in adapting the contextual factors. The overall aim of this dissertation is to evaluate whether current privacy enhancing techniques are effective in addressing user privacy and identify the factors that are essential to address deficiencies of PET in mobile platforms. We examine how contextual factors are considered by mobile users in determining the effectiveness of privacy preservation in two quantitative studies. In the first study, we investigate the influence of a critical contextual factor—recipient—on mobile device users' attitude. In the second study, we examine the relationship between two contextual factors, recipient and information type, and their influence on users' perception.

3 A Practical Exploration of the Privacy Paradox: The Role of Contextual Integrity

Abstract

The prevailing approach of privacy preservation in mobile devices through permissions management alone is not optimal due to the gap between flexibility and usability. Accommodating the diversity in contexts and also users' privacy preferences is complicated by privacy paradox—a discrepancy between expressed concern and the actual behaviour. Based on Nissenbaum's framework of contextual integrity (CI), we investigate the influence of contextual factors in users' mobile usage to examine the phenomenon of privacy paradox. We conducted two studies as part of this research. In Study 1A, we identified 15 most common groups of recipient from a sample (n = 282) of mobile users. Study 1B investigates the influence of trust and privacy concern on self-disclosure, in relation to the typical groups of recipient identified in Study 1A. Our results (n = 301) suggest trust has a significant influence on the user's disclosure behaviour, particularly on the relationship between privacy concern and self-disclosure. The mediation effect of trust in our findings suggest its significant role in determining users' self-disclosure regardless of the existence of privacy concerns. Our results also show significant demographical differences in those three factors (trust, privacy concern and self-disclosure). Overall, we believe the existence of privacy paradox can be attributed to the gap in understanding the interactions between users and their recipients. The findings suggest the potential of incorporating users' behavioural attitude on recipients in privacy management.

3.1 Introduction

Examining user's privacy attitude is often complicated by the phenomenon of privacy paradox. Scholars theorise its cause through the lens of the CI framework but do not examine the framework's practicality. In this chapter, we attempt to uncover this phenomenon based on the framework, by investigating the influence of recipient and type of information on mobile device users. Privacy literature mostly focuses on the effect of privacy concern and trust on self-disclosure. These three factors are the main variables of this user study. Although their effects are well-established, they are often studied

independently (Martin & Shilton 2016b). To address this gap, in this chapter, we investigate the effects of trust between different groups of recipient on the relationship between privacy concern and self-disclosure.

Before we could initiate the user study, since the similar studies usually derived the groups of recipients based on assumptions of previous studies and not empirically derived, it is necessary for us to first inquire about how mobile users group their contacts. We address it in Study 1A through the research question:

RQ1: What are the most common recipients identified by mobile users?

From our literature review, we find that the privacy literature mostly focuses on the effect of privacy concern and trust on self-disclosure. These three factors are the main variables of this user study. Although their effects are well-established, they are often studied independently (Martin & Shilton 2016b). Little is known of the effect of trust on the relationship between privacy concern and self-disclosure. In Study 1B, we ask the following research question:

RQ2: What are the effects of trust between different groups of recipient on the relationship between privacy concern and self-disclosure? (Figure 2)

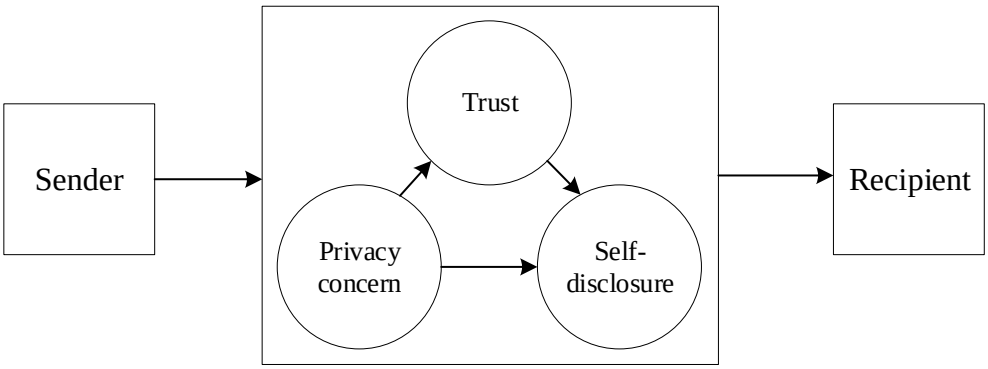

Figure 2 Influence of trust

Along with those three variables, our questionnaire also asked participants of their demographical backgrounds.

RQ3: How do mobile users' background affect their privacy concern, trust and self-disclosure?

The rest of this chapter is organised as follows. Section 2 introduces Mechanical Turk that will be utilised in the user studies. Section 3 introduces the relevant concepts. Section

4 reports on Study 1A. Section 5 contains Study 1B. Section 6 encapsulates the significance of results in both user studies. Section 6 concludes this chapter.

3.2 Mechanical Turk

Amazon Mechanical Turk (MTurk) is a crowdsourcing Internet marketplace enabling "Requesters" to recruit and pay subjects ("Turkers") to perform Human Intelligence Tasks. Although it is not specifically designed for academic research, it has since been utilised by researchers to deploy questionnaires. Although it includes survey design tools, Turkers can be instructed to answer a questionnaire on other websites (including Requesters'). It has the fastest collection times among six major US survey providers (Schnorf et al. 2014).

Several studies have been conducted to assess the representativeness and reliability of MTurk for survey deployment. The data quality is generally good (Behrend et al. 2011; Buhrmester, Kwang & Gosling 2011; Paolacci, Chandler & Ipeirotis 2010). The sample pool is more diverse than the university sample that is often used as a convenience sample (Behrend et al. 2011; Casler, Bickel & Hackett 2013; Paolacci & Chandler 2014). Despite the diversity of the sample pool, Paolacci and Chandler (2014) cautioned the reliance on Internet survey limits the representativeness of the general population, particularly in countries with low Internet penetration rate. The sample tends to be a population of heavy users, early adopters and technology optimists (Schnorf et al. 2014).

To reduce junk data, researchers have incorporated several measures on MTurk. To excludes bots or spammers, the requester can specify Turkers with certain task approval rate and have completed a number of tasks (Kuziemko et al. 2015; Shay et al. 2014). Trap question—a question with obviously wrong responses—or instructional manipulation check also can be incorporated to test whether Turkers are paying attention (to the questionnaire) (Oppenheimer, Meyvis & Davidenko 2009; Shay et al. 2014). Completion time has been suggested to detect spammer or inattentive respondent who try to complete a survey as quickly as possible (Crowston 2012). However, a study by Downs et al. (2010) suggests completion time is not a reliable indicator. They argued that valid respondent could complete quicker than usual (possibly due to good computer "reflex"), while inattentive spammer could be distracted by other tasks that would increase completion time.

3.3 Study 1A

3.3.1 Methodology

This preliminary study targeted to mobile users is created to inquire how users group their contacts. We consider this study somewhat analogous to a survey of social categorisation (van Knippenberg 1984; Zhang et al. 2013), with this study focus on general mobile device usage instead. Social categorisation is defined as "the ordering of the social environment in terms of social categories, that is, in terms of groupings of persons in a manner which is meaningful to the individual concerned" (van Knippenberg 1984, p.561). The categories used in those studies, however, were usually based on assumptions of previous studies. This entails the necessity of RQ1—to enumerate a list of groups commonly found in users' phonebook—so that RQ2 and the rest can be addressed based on empirical results.

The questionnaire has five parts; Part 1 is demographic questions, Part 2 asks about the recent three apps participants used to communicate with others, Part 3 asks participants to list three groups of contacts they *frequently* contact, Part 4 to list *infrequent* groups while Part 5 to list *organisations*. Participants were asked to rate their trust in each group they list in Part 3 to 5. See Appendix A for the questionnaire. The questionnaire was approved by the Human Research Ethics Committee of our institution (equivalent to IRB approval in the US) before the recruitment of participants. Participants were presented with a participant information sheet (Appendix C) before responding to the questionnaire.

We measured the trust level of each participant on their groups of contact with a single question "How much do you trust *x*" adapted from Molm, Takahashi and Peterson (2000) with five intervals (Not at All to Very Strongly) adapted from Butler and Cantrell (2016). Trust score was measured on a 5-point Likert-type scale, ranging from 1 (not at all) to 5 (very strongly). We found the scale to be statistically reliable with acceptable level (Nunnally, cited in Dinev & Hart 2004; Kline 2000) of internal consistency (Cronbach's $\alpha = 0.73$). This trust scale is later validated through a comparison with Individualized Trust Scale (Wheeless & Grotz 1977) (WITS) in Study 1B.

We opted for an anonymous survey, i.e. name and email address were not collected—which allows participants to be more open (Wang et al. 2011). This is also in

line with the Mechanical Turk's (MTurk) policy, which prohibits the collection of personally identifiable information (Amazon 2018). We advertised the survey on MTurk for four days in May 2018. Participants were asked to respond to our survey that we implemented on LimeSurvey. Participants were paid USD 0.10 for completing the survey. Participants spent 3 min and 38 seconds on average (median = 3 minutes 11 seconds) to complete the survey.

We took several measures suggested previously (Kuziemko et al. 2015; Page, Kobsa & Knijnenburg 2012; Shay et al. 2014) to minimise junk data. These measures are:

1 The survey is only shown to Turkers from the US and Australia locations. Location is also part of the demographic questions, and only responses with those two locations were considered valid.
2 During our pilot tests, we found completion rate and "Masters" qualification requirements are not only ineffective in reducing junk data but detrimentally reduce the response rate.
3 Respondents were required to input a password that was only shown at completion to get paid. We cross-checked responses from MTurk and LimeSurvey to identify invalid responses with a blank or incorrect password. Respondents were not able to leave any blank answer.
4 We identified incomplete or out of topic responses.
5 We identified responses with no variance in Likert scales (e.g., selecting all "Moderately").
6 We identified responses with unrealistic completion times or from the same IP address. They are not entirely invalid since those with good computer "reflex" could finish faster (Downs et al. 2010). Respondents could share a public IP address when behind a Network Address Translation (NAT) gateway. They are further inspected using measure 1-5 to verify they are invalid.

3.3.2 Results

We had a total of 423 responses from LimeSurvey. With all the measures above, we removed 141 responses and remained with 282 usable responses. The following Table 2 shows participants demographics.

Table 2 Demographics of Study 1A

Attribute	Distribution
Gender	Male (35.82%, n = 101), Female (63.83%, n = 180), Others (0.35%, n = 1)
Age	18-25 (23.4%, n = 66), 26-35 (45.39%, n = 128), 36-45 (15.25%, n = 43), 46-55 (9.57%, n = 27), 56 or above (6.38%, n = 18)
Education	Less than high school (1.42%, n = 4), High school (34.04%, n = 96), Bachelor's (48.23%, n = 136), Honours/Master's (14.18%, n = 40), Doctorate (2.13%, n = 6)
Location	Australia (1.42%, n = 4), United States (98.58%, n = 278)

RQ1: What are the most common contact groups on mobile devices?

We asked the respondents to list the names of each group of their contacts. The responses were given in free text form, resulting in a wide variety of names. We validated the response; the word frequencies of all groups fits a power-law distribution with $\alpha = 2.01$, $p = 0.59$ (Figure 3) is similar to observed distributions for English word frequencies (i.e. Moby Dick ($\alpha = 1.95$, $p = 0.69$) (Clauset, Shalizi & Newman 2009, p.684)). When counting the names, capitalisation and punctuation differences were ignored, but no stemming was performed.

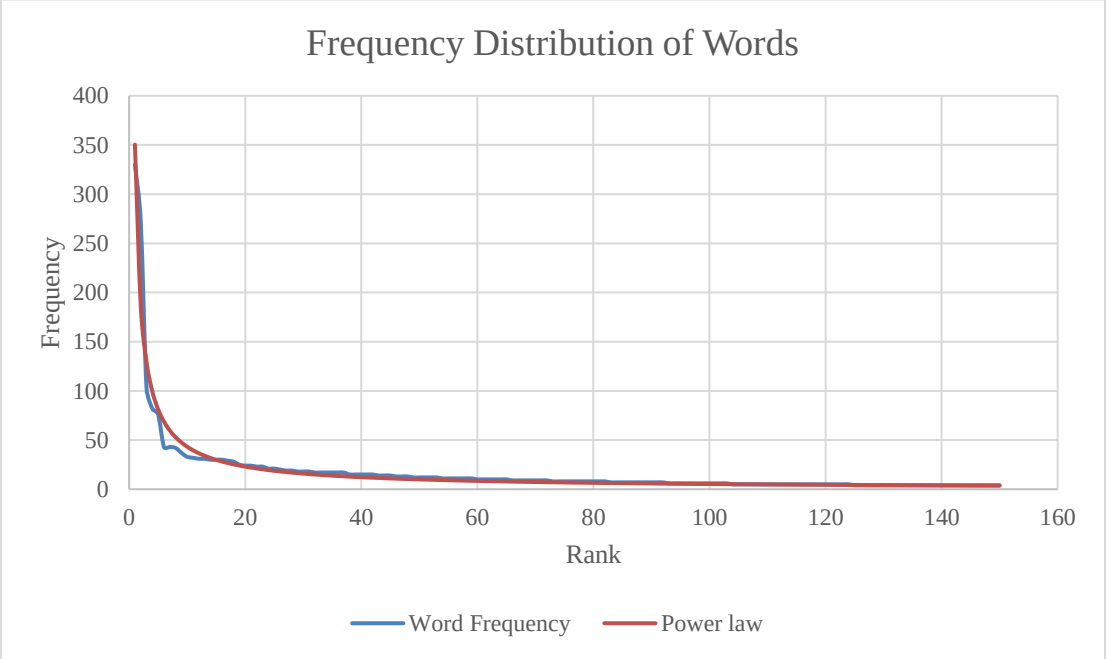

Figure 3 Power-law distribution (Study 1A)

Next, related groups were identified and combined for a smaller and more practical list. We coded gender-specific nouns into gender-neutral and companies into their relevant industry. Some groups are further aggregated together by similar industry

or synonyms to reduce the number of groups. Table 3 illustrates some examples. This combination resulted in 33 groups of people and 24 groups of organisations where each category has a frequency of at least 5. Table 4 shows the 15 most common groups and the average trust score of each group.

While not part of the main research questions, we also compared the trust score among the 15 most common groups (Table 4). We conducted a Kruskal-Wallis test (one-way ANOVA on ranks) to detect any difference in the trust score. The Kruskal-Wallis test was significant ($H = 580.41$, d.f. = 14, $p < 0.001$). Subsequent post-hoc tests are performed using Conover test with Bonferroni adjustment. The result suggests there are significant differences in trust score between most of the groups. Table 5 and Figure 4 summarise the results of the pairwise comparison.

The trust score (n = 282) were compared over three categories—*Frequent*, *Infrequent* and *Organisation*. The Friedman test was significant ($\chi^2 = 436.94$, d.f. = 2, $p < 0.01$). Subsequent tests between categories using Conover test with Bonferroni adjustment and signed-rank tests were significant to suggest Frequent is the most trusted, followed by Infrequent and Organisation respectively.

Table 3 Compilation of groups

Groups	**New groups**	**Final groups**
Husband		Spouse
Wife		
Father		Parents
Mother		
Gardening club		Hobby
Book club		
High school mates		Classmates
College friends		
Chase Bank	Bank	Financial
Insurance		
Restaurant		F&B
Pizza Hut	Fast food restaurant	
Starbucks	Café	
Doctor	Medical	Healthcare
Hospital		
Pharmacy	Pharmaceutical	

Table 4 Statistical information of groups

Groups	Frequency	Mean	Std.Dev
friends	247	4.040	0.825
colleague	241	3.154	1.011
family	235	4.502	0.736
financial	95	3.411	1.225
healthcare	86	3.779	0.975
relatives	79	3.139	1.308
social media	78	2.654	1.079
business	69	2.826	0.923
retail	68	2.868	1.078
employment	66	3.576	0.912
school	66	3.227	0.957
npo	64	3.656	1.011
classmates	62	2.468	1.020
acquaintances	59	2.220	0.789
strangers	50	1.280	0.607

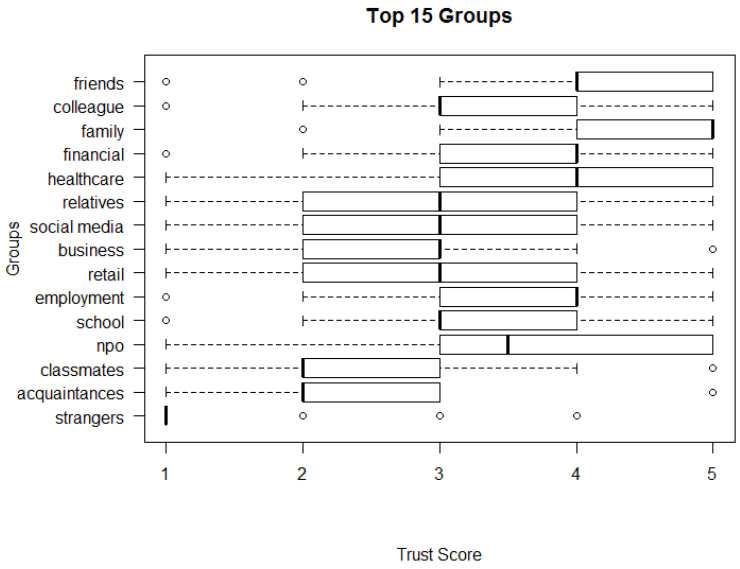

Figure 4 Boxplots of top 15 groups

Table 5 Conover Test with t statistic values

** < 0.05, ** < 0.01, *** < 0.001*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1-friends															
2-colleague	-10.896***														
3-family	5.606***	-16.334***													
4-financial	-5.318***	-2.844	9.483***												
5-healthcare	2.374	-5.488***	6.412***	-2.317											
6-relatives	7.262***	-0.369	11.146***	1.948	4.116**										
7-social media	11.04***	3.434	14.882***	5.182***	7.269***	3.102									
8-business	-9.751***	-2.499	-13.428***	-4.335***	-6.377***	-2.361	0.642								
9-retail	9.016***	1.808	12.677***	3.731*	5.777***	1.79	1.2	-0.544							
10-employment	-3.827*	-3.284	-7.474***	0.697	-1.425	2.449	5.402***	-4.632***	4.077**						
11-school	6.583***	-0.535	10.216***	1.686	3.758*	-0.159	3.119	-2.414	-1.867	2.194					
12-npo	3.172	-3.852*	6.779***	-1.219	0.895	2.936	5.863***	-5.087***	4.535***	-0.487	2.664				
13-classmates	-11.638***	-4.682***	-15.158***	-6.193***	-8.139***	-4.211***	-1.289	1.86	-2.383	-6.349***	-4.19**	-6.781***			
14-acquaintances	-13.27***	-6.446***	-16.713***	-7.727***	-9.616***	-5.719***	-2.834	-3.356	-3.867**	-7.772***	-5.641***	-8.189***	-1.482		
15-strangers	16.43***	10.048***	19.641***	10.908***	12.656***	8.905***	6.15***	6.57***	7.049***	10.762***	8.724***	11.142***	4.707***	3.252	

We performed independent-sample rank-sum test and there was no evidence (α = 0.3) to suggest *people*'s groups (*Frequent* and *Infrequent* combined) is significantly different from the *Organisation*'s trust.

3.4 Study 1B

3.4.1 Measures

Among the trust scales we discussed in Chapter 2.3, we find WITS to be suitable for our purpose, which is to measure how much a person trusts a certain group. However, we noted the scale has up to 15 items, which is only suitable for a minimal number of targets. In Study 1A, participants were asked to rate nine groups of contacts, whereas Study 1B involved two groups. Thus, to minimise survey fatigue, we deployed WITS in Study 1B only.

This study investigates the relationship between privacy concern, trust and self-disclosure. To measure them, we adapted existing scales that have been rigorously developed and tested (in terms of validity and reliability) (cf. Preibusch 2013) with slight modifications for clarity on the purpose of the study. This way, we improve on past studies by adapting from well-established scales for better validity. In measuring privacy concern, we adapted the Global Information Privacy Concern (GIPC) scale (Malhotra, Kim & Agarwal 2004). To measure trust, we adapted from the Wheeless' Individualized Trust Scale (ITS, abbreviated as WITS in the literature review section) (Wheeless & Grotz 1977). We also added a single-item scale to measure trust adapted from Molm, Takahashi and Peterson (2000) which we used in Study 1A (identified as Simple Trust, ST), separate from ITS. For self-disclosure, we adapted from Wheeless and Grotz (1976).

We initially adopted the self-disclosure scale (Wheeless & Grotz 1976) word-by-word but later found it to be insufficiently reliable. The scale had low internal consistency with Cronbach's alphas between 0.64 and 0.7 when tested with 320 participants from the MTurk. We suspected the generic items might be ambiguous to the participants. We later added clarity by mentioning "mobile devices" in each item to better align with the focus of our study. This resulted in a significant increase of alphas to between 0.716 and 0.822 (Table 6), as such can be assumed to be sufficiently reliable (> 0.7) along with the rest of the scales. Single-item trust scale is not relevant to the internal consistency measure. See Appendix A for the questionnaire.

We asked participants to rate their trust and self-disclosure towards each contact group, while privacy concern is rated generally. We compiled a list of most popular 15 groups (Table 3 and Table 7) from Study 1A. We found the frequency of the top three groups is highly disproportionate compared to the rest—the top three have 46% of the frequency of the whole 15 groups in the pilot study. To account for this proportion, we asked each participant to rate their trust and self-disclosure towards two separate groups; one is randomly from the top three groups (denoted as *Frequent*), and the other is drawn from the rest of the 12 groups (denoted as *Infrequent*). The selected groups are constant throughout each participation, so each participant rated two groups only. This arrangement made the top three groups have half of the frequency of the whole groups in this study (Table 7), which is close to the proportion in Study 1A (Table 3).

Note that *Frequent* and *Infrequent* categories used in Study 1B are distinct from similarly named in Study 1A. In Study 1A, we asked participants to list three groups of contacts (in free text) that they *frequently* and *infrequently* contact, then rate their trust on those groups. This is in contrast with Study 1B, where we asked participants to rate the groups (Table 7) coded from the results of Study 1A.

We included six demographic questions to study their possible effects on the dependent variables. The flow of the questionnaire is described as follow:

1 MTurk members (commonly known as "Turkers") are offered to participate in this questionnaire with a brief introduction.
2 Interested Turkers are redirected to the questionnaire hosted at LimeSurvey.
3 Details of the study—including its purpose, information to be collected, data anonymity, data storage and ethics information—were shown on the cover page. (Appendix C)
4 Turkers that have consented proceed as participants.
5 The survey asked the participants of their demographic background.
6 ST, ITS and self-disclosure are measured twice; each with a different group, as mentioned in the previous paragraph.
7 GIPC is shown between ITS' and self-disclosure' items.
8 At the completion of the questionnaire, each participant is presented with a random completion code to be submitted to us via MTurk.

9 Once we verified the completion code, each participant is compensated with US$0.10.

Table 6 Statistical information of each scale

***Average inter-item Spearman's correlation*

Scales	**Items**	**Mean**	**Std.Dev**	**Homogeneity***	**α**
Simple Trust (Frequent) (ST1)	1	5.279	1.522	NA	
Simple Trust (Infrequent) (ST2)	1	3.880	1.724		
Wheeless Trust (Frequent) (ITS1)	15	5.247	1.544	0.56	0.947
Wheeless Trust (Infrequent) (ITS2)	15	4.104	1.721	0.57	0.955
Privacy Concern (GIPC)	3	5.755	1.267	0.58	0.779
Self-Disclosure (Infrequent) (SD1)	5	4.837	1.617	0.34	0.716
Self-Disclosure (Infrequent) (SD2)	5	4.380	1.707	0.48	0.822

Table 7 Categorisation of top 15 groups into Frequent and Infrequent

Groups	**Frequency**	**Category**
Colleague	104	Frequent
Friends	101	
Family	96	
Education Institutions	33	Infrequent
Commercial Organisations	29	
Strangers	29	
Non-profit organizations	26	
Retail	26	
Social media (e.g. online friends)	26	
Classmates	25	
Acquaintances	24	
Employers	24	
Healthcare Organisations	21	
Relatives	21	
Financial Institutions	17	

3.4.2 Validation

We took several measures suggested previously (Kuziemko et al. 2015; Page, Kobsa & Knijnenburg 2012; Shay et al. 2014) to minimise junk data. These measures are:

1 The survey is only advertised to Turkers located in the US. Location is also part of the demographic questions, and only responses with this location are considered valid.
2 During our pilot tests, we found completion rate and "Masters" qualification requirements are not only ineffective in reducing junk data but detrimentally reduce the response rate.
3 Respondents were required to input a password that was only shown at completion to get paid. We cross-checked responses from MTurk and LimeSurvey to identify invalid responses with a blank or incorrect password. Respondents were not able to leave any blank answer.
4 We eliminated incomplete or out of topic responses.
5 We located responses with no variance in Likert scales (e.g., selecting all "Moderately") and verified with measure 8.
6 The Likert scales used in measuring trust are reversed alternately.
7 We identified responses with unrealistic completion times. They are not entirely invalid since those with good computer "reflex" could finish faster (Downs et al. 2010). They are further inspected using measure 1-6 to verify they are invalid.
8 We identified responses from the same IP address and further verified using measure 1-6.

We had a total of 358 responses from LimeSurvey. With all the measures above, we removed 57 responses and remained with 301 usable responses.

We performed several regression diagnostics to validate the regression analyses. The Variance Inflation Factor (VIF) values ranged from 1.025 to 2.069, suggesting no sign of multicollinearity between the independent variables. The Durbin-Watson values ranged from 1.825 to 2.088, suggesting no significant presence of autocorrelation. The Cook's distance values ranged from 0.048 to 0.098, thus no evidence to suggest there were highly influential outliers.

3.4.3 Results

We opted for an anonymous survey—similarly to Study 1A. We advertised the survey to MTurk's Turkers located in the US for two days in January 2019. Participants were asked to respond to our survey that we implemented on LimeSurvey. Participants were paid

USD 0.10 for completing the survey. Participants spent 4 min and 6 seconds on average (median = 3 minutes 33 seconds) to complete the survey. Table 8 shows the demographic.

Table 8 Demographics of Study 1B

Attribute	Distribution
Gender	Male (37.87%, n = 114), Female (62.13%, n = 187)
Age	18-25 (23.92%, n = 72), 26-35 (44.85%, n = 135), 36-45 (18.94%, n = 57), 46-55 (6.31%, n = 19), 56 or above (5.98%, n = 18)
Education	High school (30.9%, n = 93), Bachelor's (50.5%, n = 152), Honours/Master's (17.28%, n = 52), Doctorate (1.33%, n = 4)
Employment	Student (6.64%, n = 20), Employed (62.13%, n = 187), Employed student (6.64%, n = 20), Unemployed (7.31%, n = 22), Retired (1.99%, n = 6), Others (2.66%, n = 8)
Mobile	Android (51.83%, n = 156), iOS (42.86%, n = 129), Others (5.32%, n = 16)
Experience	0-1 year (2.33%, n = 7), 2-4 years (20.6%, n = 62), 5-7 years (35.88%, n = 108), 8 years or more (41.2%, n = 124)

3.4.3.1 Frequent / Infrequent

RQ3: How do mobile users' background affect their privacy concern, trust and self-disclosure?

To answer this research question, we compared their differences in terms of privacy concern, trust and self-disclosure.

The signed-rank test showed Frequent groups are significantly different to Infrequent groups across all three scales (i.e. ST, ITS, SD) ($p < .001$, two-tailed).

3.4.3.2 Demographics

We compare trust, privacy concern and self-disclosure among demographics. Frequent and infrequent groups are combined. We excluded "Doctorate degree" and "0-1 year experience" groups from the subsequent analyses due to low counts.

3.4.3.3 Trust

Trust is compared among the demographics, separately using a single-item trust scale (Molm, Takahashi & Peterson 2000) (identified as Simple Trust, ST) which we used to measure trust in Study 1A and Wheeless' Individualized Trust Scale (ITS) (Wheeless & Grotz 1977).

Age

Trust is compared among all five age groups. The Kruskal-Wallis test was significant and subsequent post-hoc tests suggested “18-25” is significantly different to “26-35” and “36-45”, while “56 or above” is significantly different to “36-45”.

Gender

Rank sum test is significant in ITS to suggest female has significantly higher trust than male.

Education

Trust is compared between three education levels, excluding doctorate due to low count. The Kruskal-Wallis test was significant and subsequent post-hoc tests showed “Bachelor’s” is significantly different to “Honours/Master’s.

Employment

Trust is compared between employment types, excluding retired and “others” due to low count. The Kruskal-Wallis test was significant and subsequent post-hoc tests showed “employed student” is significantly different to “employed”, “self-employed” and “student”, while “unemployed” is significantly different to “employed” and “self-employed”.

Mobile

Trust is compared between Android and iOS as other categories have insufficient count. Rank sum test is significant in ITS to suggest that iOS has significantly higher trust than Android ($p < 0.001$).

Figure 5 and Table 9 summarise the demographical differences in ST; Figure 6 and Table 10 summarise the demographical differences in ITS.

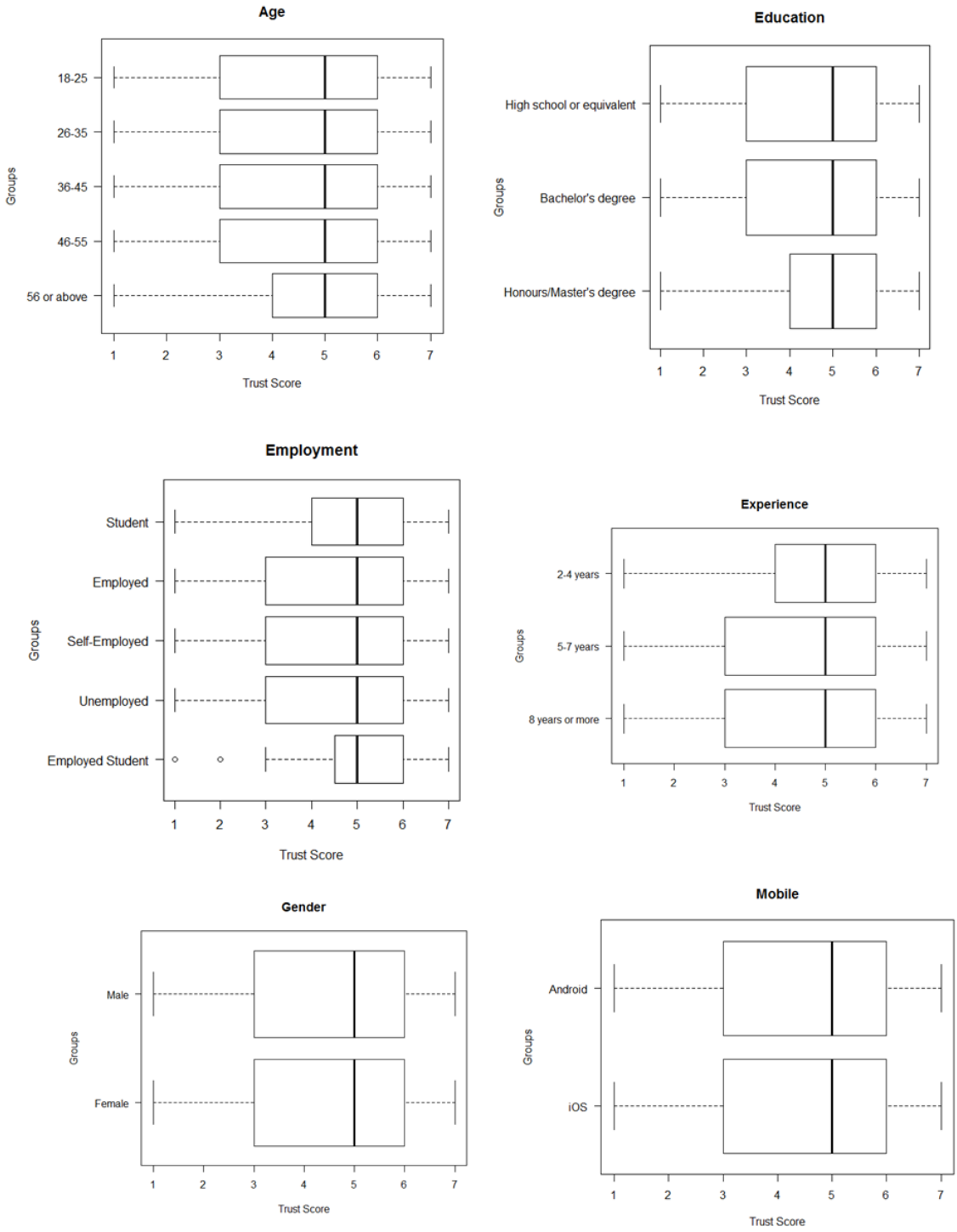

Figure 5 Demographical differences in Simple Trust

Table 9 Differences in Simple Trust (ST) among demographics

ST	Mean	Std.Dev	Significance
Age:			H = 0.508, df = 4, p = .973
18-25	4.63	1.80	
26-35	4.59	1.72	
36-45	4.46	1.81	
46-55	4.66	1.68	
56 or above	4.56	2.01	
Education:			H = 1.505, df = 2, p = .471
High school or equivalent	4.44	1.85	
Bachelor's degree	4.62	1.75	
Honours/Master's degree	4.69	1.70	
Employment:			H = 4.203, df = 4, p = .379
Student	4.95	1.60	
Employed	4.52	1.81	
Self-Employed	4.61	1.60	
Unemployed	4.52	1.81	
Employed Student	5.03	1.72	
Experience:			H = 2.256, df = 2, p = .324
2-4 years	4.73	1.67	
5-7 years	4.63	1.81	
8 years or more	4.45	1.80	
Gender:			p = .259
Male	4.47	1.80	
Female	4.64	1.75	
Mobile:			p = .288
Android	4.46	1.81	
iOS	4.64	1.74	

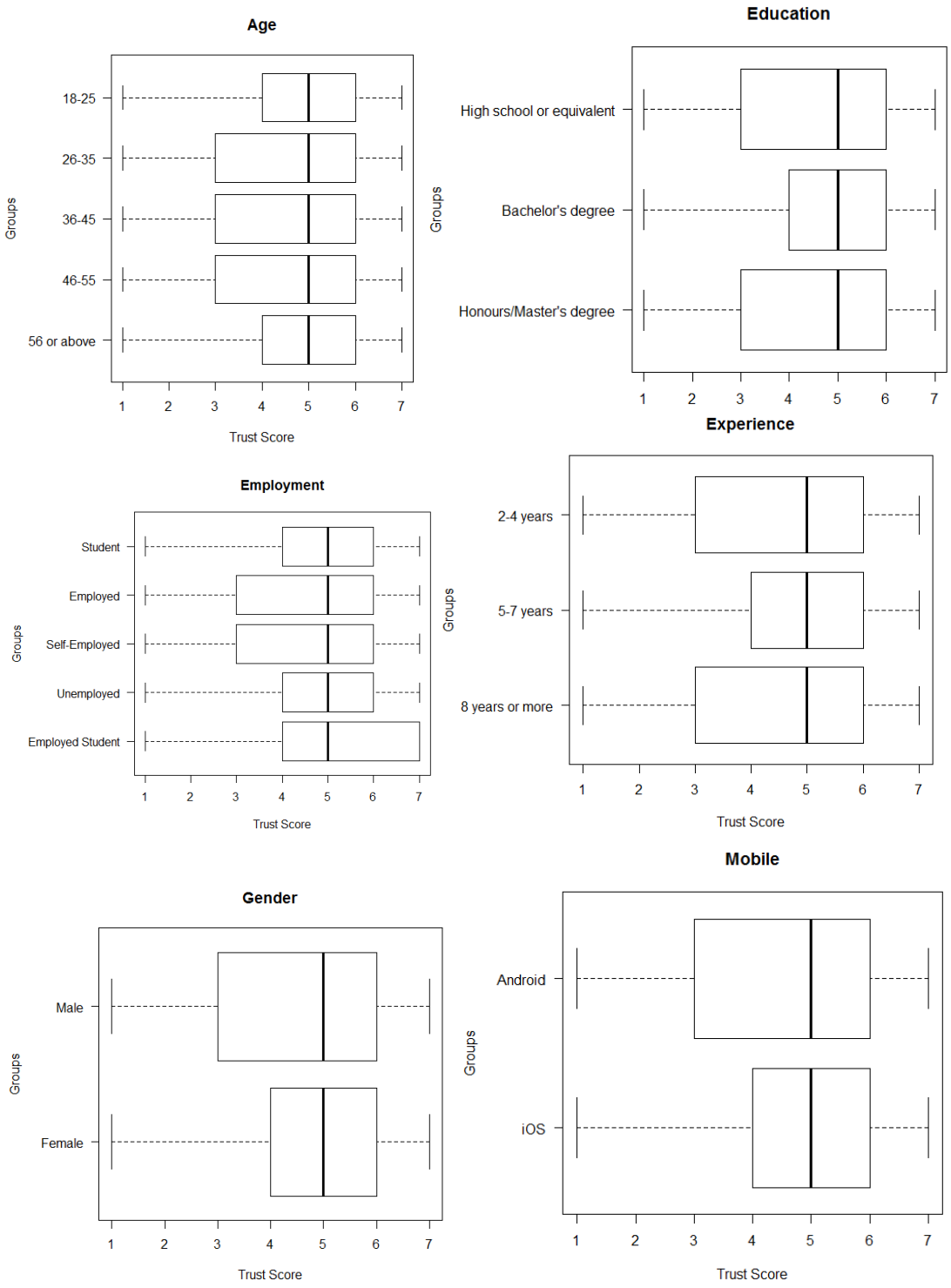

Figure 6 Demographical differences in ITS

Table 10 Differences in ITS between demographics

ITS	Mean	Std.Dev	Significance
Age:			H = 38.686, df = 4, $p < 0.001$
18-25	4.82	1.78	
26-35	4.65	1.68	
36-45	4.53	1.74	
46-55	4.60	1.70	
56 or above	4.80	1.86	
Education:			H = 15.290, df = 2, p < .001
High school or equivalent	4.65	1.81	
Bachelor's degree	4.73	1.72	
Honours/Master's degree	4.57	1.62	
Employment:			H = 58.997, df = 4, p < .001
Student	4.79	1.70	
Employed	4.61	1.75	
Self-Employed	4.58	1.72	
Unemployed	4.99	1.47	
Employed Student	5.07	1.72	
Experience:			H = 3.504, df = 2, p = .174
2-4 years	4.65	1.71	
5-7 years	4.73	1.70	
8 years or more	4.65	1.79	
Gender:			p < .001
Male	4.52	1.81	
Female	4.77	1.68	
Mobile:			p < .001
Android	4.59	1.80	
iOS	4.77	1.66	

3.4.3.4 Privacy concern

Age

The Kruskal-Wallis test was significant and subsequent post-hoc tests showed "56 or above" age group is significantly different to "18-25", "26-35" and "36-45" groups.

Experience

The Kruskal-Wallis test was significant and subsequent post-hoc tests showed the participants with at least 8 years of smartphone experience is significantly different from those with 5 to 7 years.

Gender

The rank-sum test was significant ($p < 0.01$) to suggest female has significantly deeper privacy concern than male.

Figure 7 and Table 11 summarise the demographical differences in privacy concern.

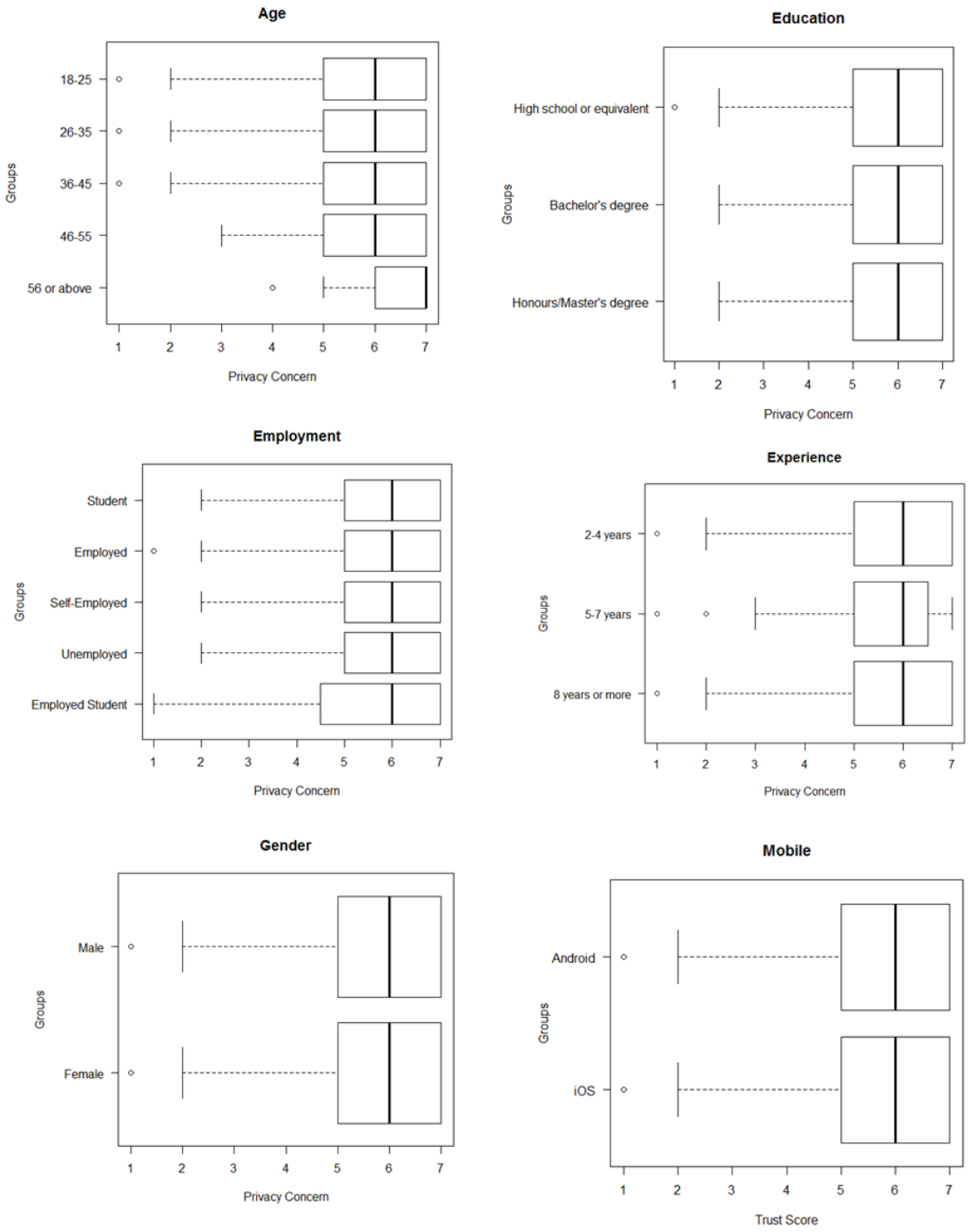

Figure 7 Demographical differences in privacy concern

Table 11 Differences in privacy concern among demographics

GIPC	Mean	Std.Dev	Significance
Age:			H = 31.558, df = 4, p < .001
18-25	5.49	1.42	
26-35	5.70	1.27	
36-45	5.91	1.15	
46-55	6.00	1.04	
56 or above	6.46	0.75	
Education:			H = 1.472, df = 2, p = .479
High school or equivalent	5.63	1.49	
Bachelor's degree	5.78	1.17	
Honours/Master's degree	5.89	1.13	
Employment:			H = 13.414, df = 6, p = .037
Student	5.78	1.25	
Employed	5.75	1.27	
Self-Employed	5.87	1.18	
Unemployed	5.65	1.21	
Employed Student	5.38	1.54	
Experience:			H = 13.433, df = 2, p = .004
2-4 years	5.74	1.29	
5-7 years	5.61	1.25	
8 years or more	5.89	1.27	
Gender:			p = .004
Male	5.62	1.29	
Female	5.84	1.24	
Mobile:			p = .288
Android	5.83	1.27	
iOS	5.66	1.28	

3.4.3.5 Self-Disclosure

Education

The Kruskal-Wallis test was significant and subsequent post-hoc tests showed participants who hold a bachelor's degree is significantly different to high school and Master's degree, while participants with high school education are significantly different to those with Master's degree.

Mobile

The rank-sum test was significant ($p < 0.001$) to suggest that Android has significantly higher self-disclosure than iOS.

Experience

The Kruskal-Wallis test was significant and subsequent post-hoc tests showed the participants with 2 to 4 years of smartphone experience is significantly different to those with 5 to 7 years and those with at least eight years of experience.

Gender

The rank-sum test was significant ($p = 0.015$) to suggest female has significantly higher self-disclosure than male. Figure 8 and Table 12 summarise the demographical differences in self-disclosure.

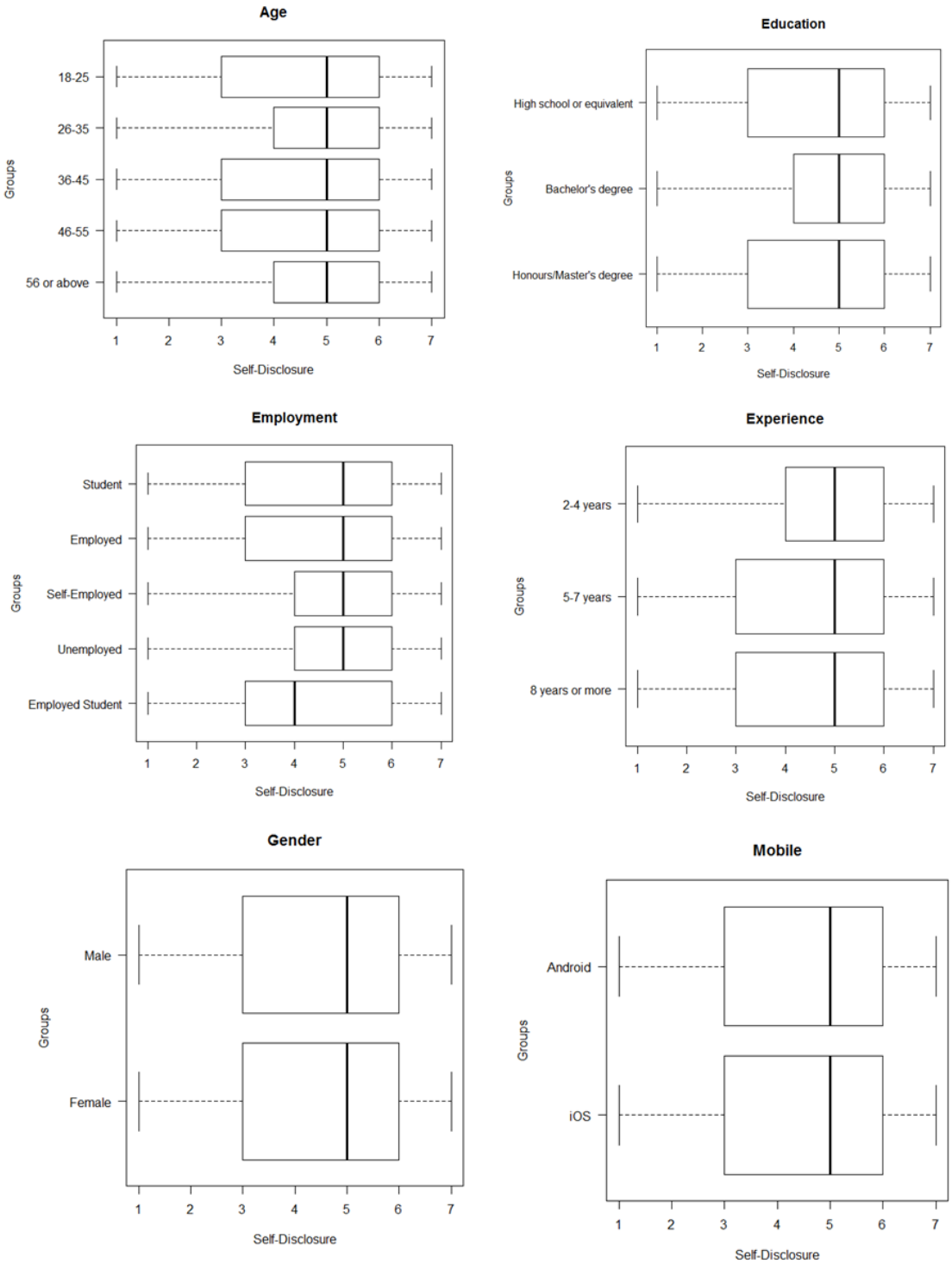

Figure 8 Demographical differences in self-disclosure

Table 12 Self-disclosure differences between demographics

SD	Mean	Std.Dev	Test statistics
Age:			H = 5.916, df = 4, p = 0.206
18-25	4.50	1.73	
26-35	4.70	1.61	
36-45	4.53	1.72	
46-55	4.60	1.74	
56 or above	4.63	1.77	
Education:			H = 34.846, df = 2, p < .001
High school or equivalent	4.54	1.68	
Bachelor's degree	4.75	1.68	
Honours/Master's degree	4.29	1.63	
Employment:			H = 13.012, df = 4, p = .011
Student	4.61	1.61	
Employed	4.63	1.69	
Self-Employed	4.74	1.67	
Unemployed	4.56	1.59	
Employed Student	4.25	1.69	
Experience:			H = 12.975, df = 2, p = .002
2-4 years	4.81	1.63	
5-7 years	4.53	1.68	
8 years or more	4.56	1.71	
Gender:			p = .015
Male	4.51	1.70	
Female	4.67	1.66	
Mobile:			p < .001
Android	4.69	1.78	
iOS	4.45	1.55	

3.4.3.6 Correlations and regressions

RQ2: What are the effects of trust between different groups of recipient on the relationship between privacy concern and self-disclosure?

Correlation analysis showed that privacy concern and trust are significantly correlated with self-disclosure in both frequent and infrequent groups (Table 14). Figure 9, Figure 10 and Figure 11 illustrate the relationship between trust, privacy concern and self-disclosure through quantile-quantile (Q-Q) plots. Table 14 summarises the comparison between regression models. The regression models (1 and 2) with privacy concern and trust explained 12% and 22% of the variances in self-disclosure in both frequent and infrequent groups respectively. Correlation analysis results indicated that the demographics are weakly and not significantly correlated with self-disclosure. Adding them as predictors in Model 1 and 2 and resulted in Model 3 and 4 did not significantly

increase the R^2 (p = .571). The insignificant results of regression models of demographics predicting self-disclosures further suggest they were impractical in predicting self-disclosure and therefore excluded in subsequent analyses. Model 1 and 2 (Table 15) formed the basis for the moderation and mediation tests detailed in the next section.

Table 13 Summary of all regression models.

Model	R^2	Adjusted R^2	Significance	Standard Error of Estimate	F-statistic
1	.12	.11	< .001	.96	(3,297) = 13.44
2	.22	.21	< .001	1.13	(3, 297) = 28.34
3	.13	.11	< .001	.97	(9, 291) = 4.99
4	.26	.24	< .001	1.11	(9, 291) = 11.30
5	.01	-.01	.858	1.03	(6, 294) = 0.43
6	.04	.02	.045	1.26	(6, 294) = 2.18

Model 1: Predictors: (constant), ST, ITS, GIPC. (Frequent)

Model 2: Predictors: (constant), ST, ITS, GIPC. (Infrequent)

Model 3: Predictors: (constant), Age, Gender, Education, Employment, Experience, Mobile Platform, ST, ITS, GIPC. (Frequent)

Model 4: Predictors: (constant), Age, Gender, Education, Employment, Experience, Mobile Platform, ST, ITS, GIPC. (Infrequent)

Model 5: Predictors: (constant), Age, Gender, Education, Employment, Experience, Mobile Platform. (Frequent)

Model 6: Predictors: (constant), Age, Gender, Education, Employment, Experience, Mobile Platform. (Infrequent)

Table 14 Correlation between demographics, trust, privacy concern and self-disclosure

*** p < 0.01, * p < 0.05, ST = simple trust scale, ITS = Individualized Trust Scale, GIPC = Global Information Privacy Concern, SD = self-disclosure*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1-Gender	1												
2-Age	.000	1											
3-Education	-.039	.228**	1										
4-Experience	-.109	.098	-.022	1									
5-Employment	-.025	-.250**	-.295**	-.025	1								
6-Mobile	.146*	.120*	-.067	.010	-.100	1							
7-ST (Frequent)	-.063	-.014	.004	.036	.118*	-.087	1						
8-ST (Infrequent)	-.054	-.033	.071	-.162**	.058	-.085	.279**	1					
9-ITS (Frequent)	-.117*	-.037	-.072	.129*	.105	-.070	.671**	.122*	1				
10-ITS (Infrequent)	-.089	-.041	.013	-.087	.060	-.037	.136*	.677**	.145*	1			
11-GIPC	-.104	.185**	.046	.086	-.070	.081	.120*	-.003	.166**	-.093	1		
12-SD (Frequent)	-.052	-.011	-.007	-.066	.019	.067	.299**	.216**	.251**	.074	.175**	1	
13-SD (Infrequent)	-.046	.056	-.020	-.115*	-.099	.070	.105	.421**	.020	.262**	.127*	.539**	1

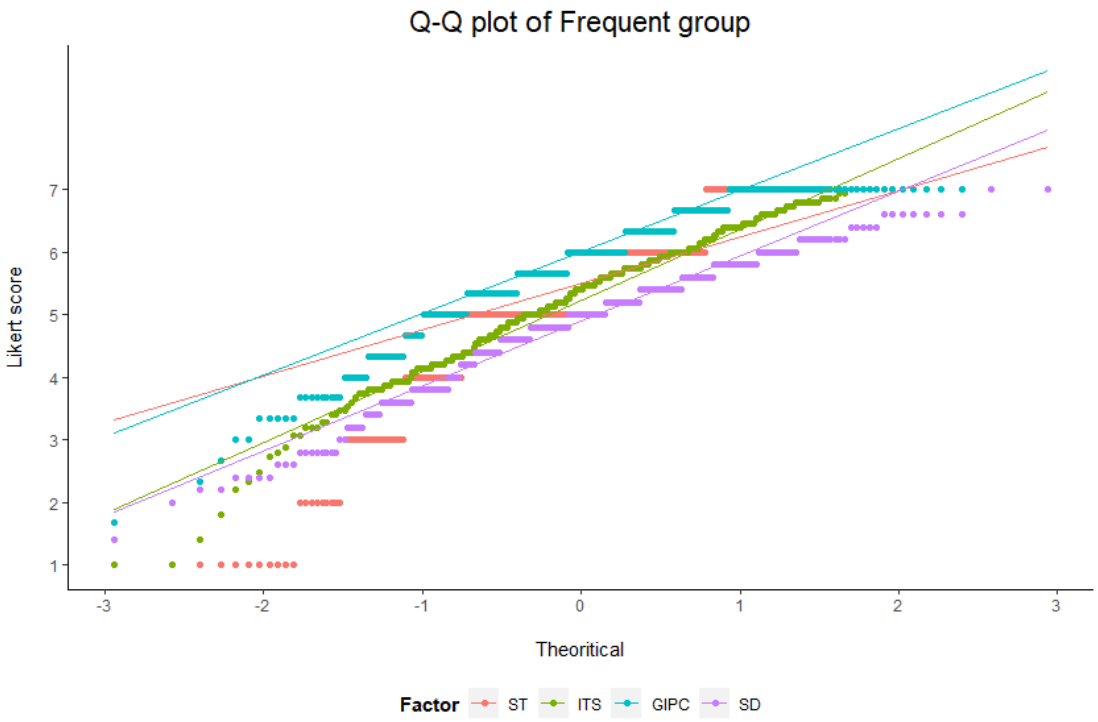

Figure 9 Q-Q plot of Frequent group

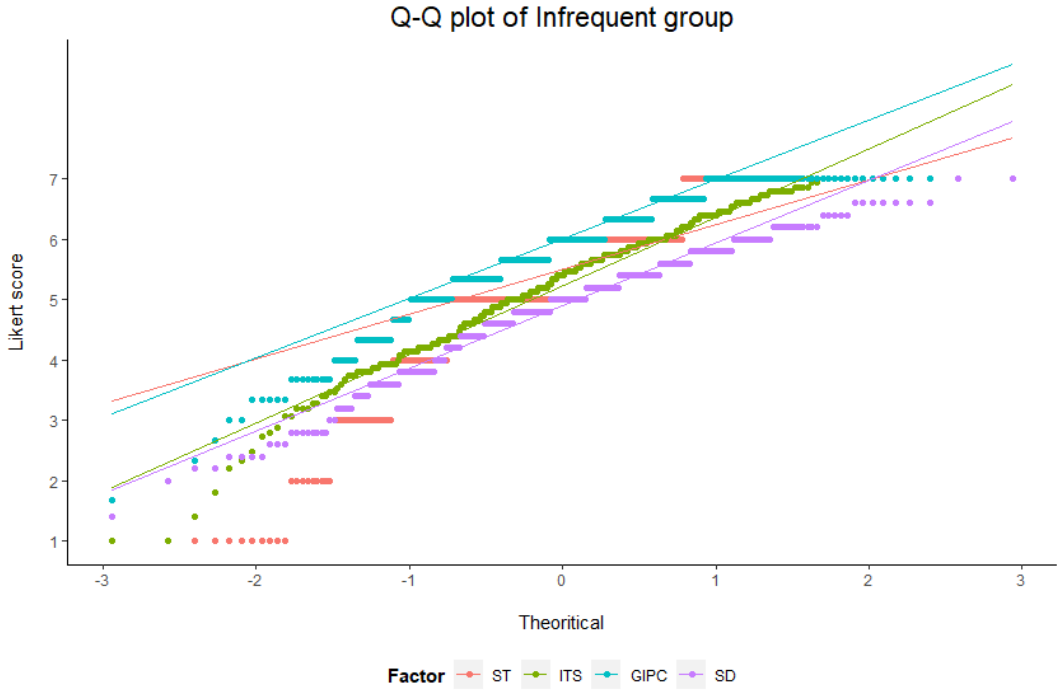

Figure 10 Q-Q plot of Infrequent group

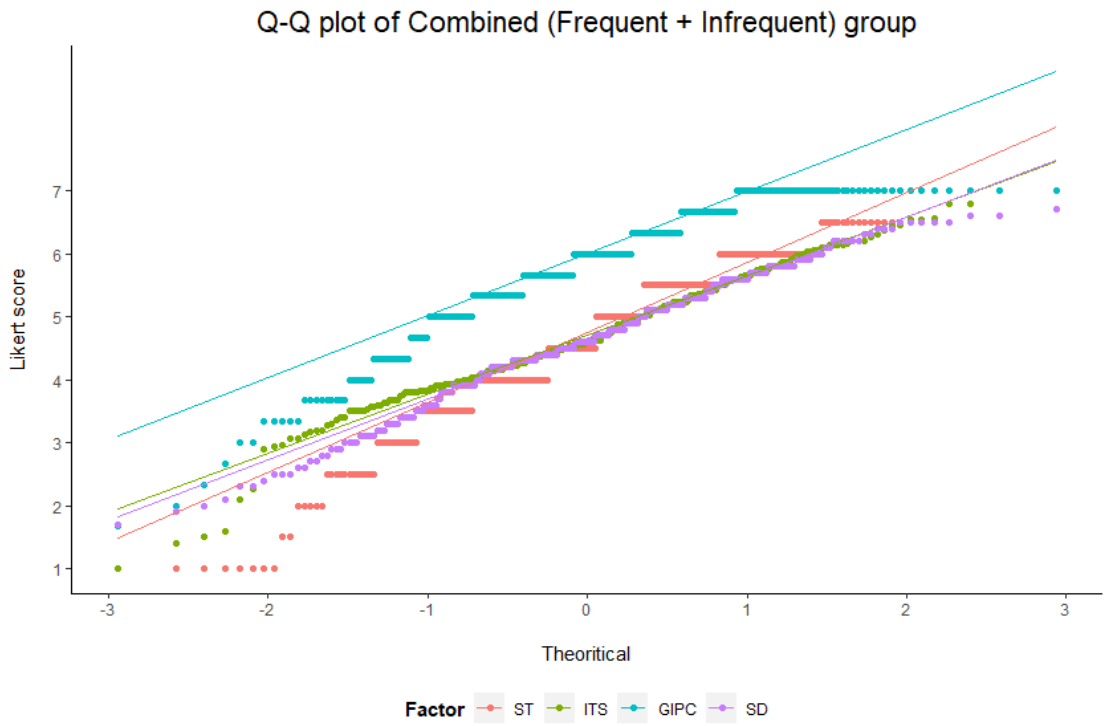

Figure 11 Q-Q plot of combined group

Table 15 Results of regression model 1 and 2

Model	**1**	**2**
Criterion	**Self-Disclosure (Frequent)**	**Self-Disclosure (Infrequent)**
ST	.14**	.35***
ITS	.08	-.06
GIPC	.16**	.19**
R^2	.12***	.22***

3.4.4 Moderation and mediation

3.4.4.1 Moderation

We tested whether trust moderates the relationship between privacy concern and self-disclosure. We conducted the tests using hierarchical multiple regression analysis. Initially, we separated the analysis with different trust measures (ST and ITS), frequent and infrequent groups. Later we found the two trust measures, despite being conceptually similar, the diagnostics did not suggest multicollinearity and autocorrelation. As such, we also conducted analyses with those trust measured together.

In the first step with privacy concern and trust, significantly accounted between 22% and 10% of the variance in self-disclosure. The addition of an interaction term in the second step increased the prediction to 23% and 11% of self-disclosure's variance, ranging from 1.4% increase to static. The increments, however, are not significant, thus

could not support the hypothesis of trust moderating relationship between privacy concern and self-disclosure. Table 16

Table 16 Moderation effect comparison between frequent and infrequent groups

and Table 17 summarise the results.

Table 16 Moderation effect comparison between frequent and infrequent groups

Group	**Frequent**			**Infrequent**		
DV	**SD**					
IV	**GIPC**					
MV	**ST**	**ITS**	**SD+ITS**	**ST**	**ITS**	**SD+ITS**
First step: (without interaction term)						
R^2	$.12^{***}$	$.10^{***}$	$.12^{***}$	$.22^{***}$	$.10^{***}$	$.22^{***}$
F-statistic	F(2,298)=19.44	F(2,298)=16.1	F(3,297)=13.44	F(2,298)=42.23	F(2,298)=17.41	F(3,297)=28.34
Second step: (with interaction term)						
R^2	$.12^{***}$	$.11^{***}$	$.13^{***}$	$.22^{***}$	$.11^{***}$	$.23^{***}$
F-statistic	F(3,297)=12.92	F(3,297)=111.71	F(5,295)=9.08	F(3,297)=28.51	F(3,297)=11.68	F(5,295)=17.23
ΔR^2	.000	.008	.014	.003	.001	.004

Table 17 Moderation effect with or without separation of ST & ITS

	Separate ST & ITS		**ST + ITS**	
Group	**Frequent**	**Infrequent**	**Frequent**	**Infrequent**
DV	**SD**			
Without interaction term				
GIPC	$.17^{**}$	$.20^{**}$	$.16^{**}$	$.19^{**}$
ST	$.18^{***}$	$.32^{***}$	$.14^{**}$	$.35^{***}$
GIPC	$.16^{**}$	$.25^{***}$	NA	
ITS	$.21^{***}$	$.26^{***}$	.08	-.06
With interaction term				
GIPC	.18	.07	$.53^{*}$	-.01
ST	.19	.11	-.26	.20
GIPC×ST	-.002	.04	.07	.03
GIPC	$.53^{*}$	.12	NA	
ITS	$.63^{*}$	.08	$.91^{*}$	-.21
GIPC×ITS	-.07	.03	$-.15^{*}$	.02

3.4.4.2 Mediation

We examined the possibility of trust mediating the relationship between privacy concern and self-disclosure. Mediation can be identified through a series of regression models. Since there were two measures of trust and main groups (frequent and infrequent), each series of models have four possible combinations.

In the first step, privacy concern significantly predicted trust in most models, except for the infrequent group and ST combination. In the second step, privacy concern significantly predicted self-disclosure in all models. In the third and fourth steps, including both trust and privacy concern, they both significantly predicted self-disclosure. Mediation of trust on privacy concern is suggested as the effect of privacy concern on self-disclosure is lower (0.04 to 0.01) including significance in the fourth step than in the second, except for infrequent group and ITS combination. We consider the mediation to be partial as the effect of privacy concern, controlling for trust, did not drop to zero and was still significant. We did not observe any mediation effect from any of the demographics, individually nor combined. The results are summarised in Figure 12.

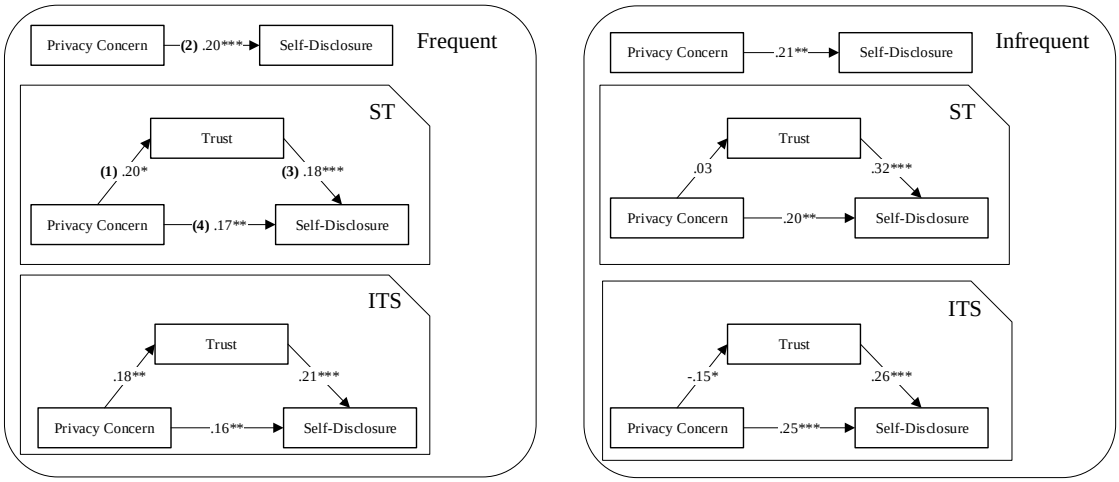

Figure 12 Mediation effects

3.5 Discussion

The main aim of this study is to investigate the relevance of the contextual integrity to the mobile ecosystem, particularly the privacy aspect. The contextual integrity emphasises on the influence of contextual factors in our every day's mobile usage. One such contextual factor is the recipients—user's attitude towards them. While there could be various dimensions of attitude, our results suggest trust having a significant influence on the user's disclosure behaviour, particularly on the relationship between privacy concern and self-disclosure. The mediation effect of trust in our results suggest its significant role in determining users' self-disclosure despite the existence of privacy concern. Our results, to some extent, are in line with an SNS study that argued that privacy concern might not necessarily inhibit self-disclosure (Heravi, Mubarak & Choo 2018; Taddicken 2014).

A contrasting result in a privacy paradox study (Norberg, Horne & Horne 2007) shows that the effect of trust on self-disclosure is insignificant. The disclosure targets were not explicitly named and only known by the industry in both of our study and their study. They alleged the resulting ambiguity could not sufficiently inform the participants from clear judgement on information disclosure. While this study also did not name the organisations, we believe the inclusion of groups of people (e.g. *colleague*, *friends* and *family*) along with their popularity—15 most popular groups from the pilot study—are much more relatable to the participants, thus could have offered better clarity. The difference in users' association with target groups is also evident in our mediation tests, whereby we observed stronger mediation effects in frequent groups compared to the infrequent groups. Those discrepancies that were mentioned, in effect, calls for caution over the use of non-empirical measures that did not account for the degree of association to the users.

The partial mediation of trust, instead of complete mediation, in addition to its insignificant moderation, suggest the possibility of additional factors. Possible factors are such as perceived benefits (Acquisti & Grossklags 2005), perceived control (Chen, J et al. 2009) and self-efficacy (Chen, HT & Chen 2015; Yao, Rice & Wallis 2007) as featured in SNS studies. Future studies can further explore other possible factors. A study (Joinson et al. 2010) shows a somewhat contrasting result, whereby trust has slight moderation but insignificant mediation effect on the relationship on privacy concern and non-disclosure. A plausible explanation for this discrepancy is that the study is concerned with users' tendency to *withhold* information, in contrast with information *disclosure* that is the focus of this study. Another study (Lin, S-W & Liu 2012) shows trust has significant moderation but insignificant mediation on the relationship between privacy concern and information disclosure. We suspect the discrepancy could be similarly explained as the previous paragraph. This could also suggest users respond differently in different use cases (i.e. mobile device vs. SNS) and the results are not necessarily applicable to one another.

Our results show significant demographical differences on trust, privacy concern and self-disclosure. Female users have a higher tendency to disclose information compared to the male counterpart on a mobile device, which is somewhat in line with the results in Li, K, Lin and Wang (2015) but a contrast to results in Xie and Kang (2015). However, our results could not be directly compared to those studies. In this study, we

treat the *information* in a generic sense (and we believe so did the participants as well), whereas those studies differentiate it in sensitivity and type. We also observed that female users tend to exhibit greater privacy concern and trust. We found that age has a positive and significant association with privacy concern but insignificant with trust and self-disclosure. This result is in contrary with Li, K, Lin and Wang (2015) whereby the results (of that study) suggest older users tend to disclose less; but our non-result is consistent with Nicholas et al. (2019). We speculate the contrasting results could be attributed to the additional constructs, i.e. trust and privacy concern, suggesting a potential interplay between them. A future study could investigate more in-depth of such relationship.

On experience with the mobile device, we observed users with more experience tend to have more privacy concern, but we did not find any significant difference in trust and self-disclosure. The non-result is in line with Martin and Shilton (2016b) which suggest more experienced users are more dependent on contextual factors—through a *combination* of them—compared to more recent users. On the overall results from a demographic perspective, there is evidence of demographical differences on trust, privacy concern and self-disclosure; however, we did not find any evidence to suggest demographic backgrounds are related nor the ability to predict those three factors. The lack of evidence suggests it may not be helpful to categorise users and caution the use of privacy profiling adopted in privacy recommendation systems. The mediation effect as evidenced in our result was significant, regardless of demographic. Our findings are consistent with Martin and Nissenbaum (2016) that show consumer across those categories could share a similar view on privacy expectations.

We conducted the survey based on self-reported data from participants that have the potential for cognitive bias and may not translate to actual behaviour. Future research could opt for the Experience Sampling Method, which solicits responses while users are actively using a mobile device (Larson & Csikszentmihalyi 2014; Shih 2015). The survey design has a potential priming effect on participants, mainly due to the privacy concern measure. However, aside from demographical background questions, the survey contained only five questions. While the short length was initially intended to prevent participation fatigue, it could also help to minimise the priming effect.

3.6 Conclusion

The study presented in this chapter suggest the influence of *recipient* on the users' privacy attitude, which as far as we are aware of, has not been considered in the currently implemented permissions management. The results also suggest that the different propensity of trust towards recipients can influence self-disclosure, despite having privacy concern. As Nissenbaum (2010) argued, "…there is no paradox in caring deeply about privacy and, at the same time, eagerly sharing information as long as the sharing and withholding conform with the principled conditions prescribed by governing contextual norms". That aside, our study is also more applicable to the mobile device, whereas many studies (as discussed in the literature review section) generally focus on SNS. While mobile apps that provide access to SNS may be popular on a mobile device, those apps are not necessarily more popular than others (Sensor Tower 2019b, 2019a).

The current approach of PET on mobile platforms through permissions management is arguably insufficient in protecting users' privacy. While fine-grained permission manager and privacy recommendation systems are useful improvements, they also further underscore the fundamental issues of the approach with the lack of grasp from the users (Felt et al. 2012b; Kelley, Patrick G et al. 2012) and its prone to abuse (Felt et al. 2011). Later studies have also shed light on usability issues faced by older adults when interacting with these systems (Frik et al. 2019; Huang, H-Y 2019). The understanding that we obtained from the results can help advances future mobile platforms that are user-centred and incorporate privacy-by-default and privacy-by-design principles.

4 Information Disclosure in Mobile Device: Examining the Influence of Information Relevance and Recipient

Abstract

Guided by Nissenbaum's framework of contextual integrity (CI), we conducted two studies as part of this research to investigate the influence of contextual factors in users' mobile usage. Specifically, we inquire about the influence of recipient and information type on mobile users' attitude. In Study 2A, we compiled 15 most common types of information from a sample (n = 390) of mobile users. In Study 2B (n = 2889), we investigated the influence of relevance of information types on the willingness of disclosure towards typical groups of recipient. While the results suggest a significant relationship between information relevance (of different information) and willingness to disclose (to different recipients), closer examination reveals the relationship is not always clear-cut, and there is a potential influence of recipient. Therefore, incorporating the recipient factor can serve as a potential improvement to the existing approach in privacy management in the mobile device.

4.1 Introduction

Existing studies have shown users often assess an information flow based on diverse contextual factors. A series of studies (Lin, J et al. 2014; Lin, J et al. 2012) showed a significant influence of *purpose* on users' subjective judgement. This is also in line with Zimmer, JC et al. (2010) that showed users are more willing to disclose information when it is perceived to be *relevant* to the function provided by the receiving service provider. These studies, in a way, also suggest users are increasingly demanding mobile apps to be more upfront about information request. This is evident in a study (Wijesekera et al. 2017) where the results suggest users consider app visibility as an essential factor in deciding on permission request, as users are usually not comfortable with an app collecting data in the background. A study on personal health data (Nicholas et al. 2019) showed participants considered not only the recipient but also the data type before disclosure. The result is also in line with Martin and Nissenbaum (2016) which showed the influence of the type of information, contextual actor (recipient) and purpose of information; the study also showed 'sensitivity' is subjectively influenced by contextual factors.

In this chapter, we expand upon previous chapter to examine additional factors. We venture on the following research question:

RQ1: What are the effects of the relevance of information types to different recipient, on the willingness to disclose? (Figure 13)

Continuing from our previous chapter which showed the influence of recipient, in this chapter, we undertake a study to investigate the relationship of data type and its relevance on the willingness to disclose to specific groups of recipients. Distinct from another similar studies (Marmion et al. 2019; Martin & Nissenbaum 2016) which utilize generic data types, our study is more specific to mobile device usage where we derive data types from mobile users.

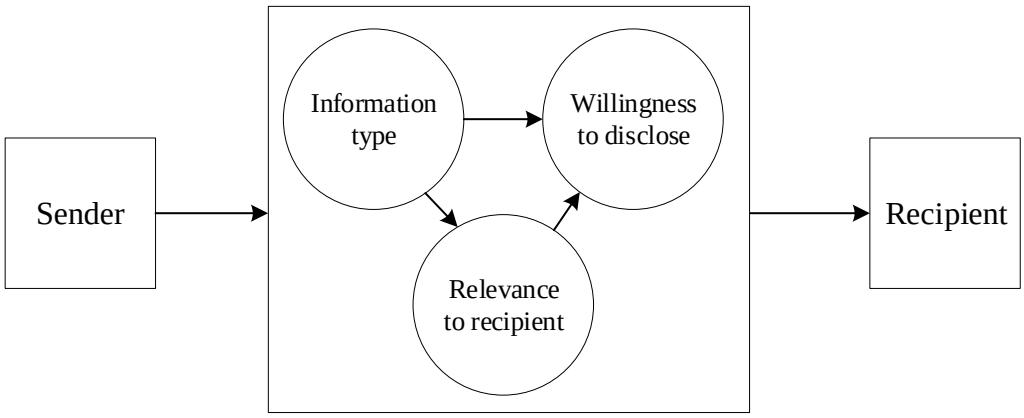

Figure 13 Influence of information relevance

The rest of this chapter is organised as follows. Section 2 reports on Study 2A. Section 3 contains Study 2B. Section 4 discusses the results of the user studies. Section 5 concludes this chapter.

4.2 Study 2A

4.2.1 Methodology

We located existing studies (Marmion et al. 2019; Martin & Nissenbaum 2016) that are closest to the purpose of our study, to examine a varying willingness of disclosure on the different data type. The lists of data type adapted in those studies were derived from Madden et al. (2014) and World Economic Forum (2012), respectively. We initially considered to adapt the measures from those sources; however, we later found the

derivation methods behind Madden et al. (2014) and World Economic Forum (2012) to be not sufficiently transparent. We also consider the lists to be generic and may not be pervasive in mobile device usage. This entails the necessity of enumerating a list of information types commonly disclosed by mobile users, so that RQ1 can be addressed based on empirical results.

To improve the relevance of the responses, we pre-tested the questionnaire over several iterations, each time with improvement on the question's clarity. To avoid priming the participants, we took precaution to avoid "privacy" keyword in our questionnaire's title and description, and in the questions (refer to Appendix B for the questionnaire).

We advertised the survey on Mechanical Turk (MTurk) for nine days in May 2019. Participants were asked to respond to our survey that we implemented on LimeSurvey. Participants spent 3 min and 57 seconds on average (median = 3 minutes 15 seconds) to complete the survey. Participants were paid USD 0.10 for completing the survey.

We utilized the following measures to minimise irrelevant data:

1 The survey is only shown to Turkers from the US location. Location is also part of the demographic questions, and only responses that specified the US were considered valid.
2 Respondents were required to input a password that was only shown at completion to get paid. We cross-checked responses from MTurk and LimeSurvey to identify invalid responses with a blank or incorrect password. Respondents were not able to leave any blank answer.
3 We identified incomplete or out of topic responses.
4 We identified responses with unrealistic completion times.
5 We identified responses that have the same IP address. We were aware that respondents could share a public IP address when behind a Network Address Translation (NAT) gateway. They are further inspected using measure 1-4 to verify their validity.

4.2.2 Results

We had a total of 435 responses from LimeSurvey. With all the measures above, we removed 45 responses and had 390 usable responses. Table 18 summarises participant demographics in Study 2A.

Table 18 Demographics of Study 2A

Attribute	**Distribution**
Gender	Male (31.03%, n = 121), Female (68.97%, n = 269)
Age	18-25 (20.77%, n = 81), 26-35 (37.95%, n = 148), 36-45 (21.79%, n = 85), 46-55 (13.33%, n = 52), 56 or above (6.15%, n = 24)
Education	Less than high school (1.42%, n = 4), High school (34.04%, n = 96), Bachelor's (48.23%, n = 136), Honours/Master's (14.18%, n = 40), Doctorate (2.13%, n = 6)
Employment	Student (5.38%, n = 21), Employed (58.97%, n = 230), Self-employed (13.33%, n = 52), Employed student (6.15%, n = 24), Unemployed (12.057%, n = 47), Retired (4.1%, n = 16)
Mobile	Android (49.49%, n = 193), iOS (42.31%, n = 165), Android and iOS (4.62%, n = 18), Others (3.59%, n = 14)
Experience	0-1 year (2.82%, n = 11), 2-4 years (15.13%, n = 59), 5-7 years (31.03%, n = 121), 8 years or more (51.03%, n = 199)

We asked the respondents to list the names of each group of their contacts. The responses were given in free text form, resulting in a wide variety of names. We combined the responses from those two questions and performed validation; the word frequencies of all groups fits a power-law distribution with $\alpha = 1.83$, $p = 0.02$ (Figure 14). It is similar to observed distributions for English word frequencies (i.e. Moby Dick ($\alpha = 1.95$) (Clauset, Shalizi & Newman 2009, p.684)). When counting the names, capitalisation and punctuation differences were ignored, but no stemming was performed.

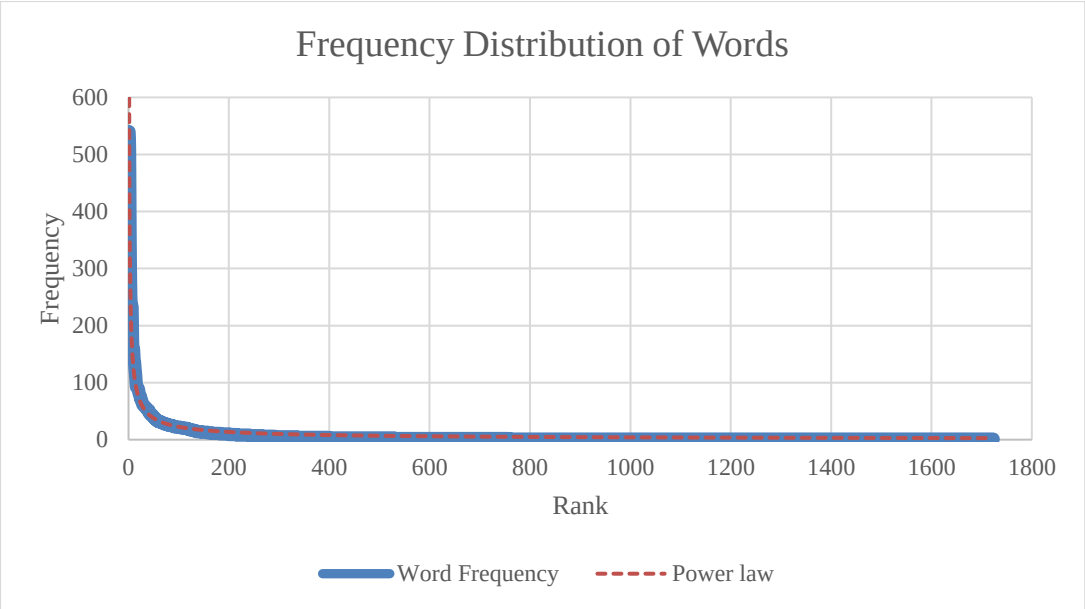

Figure 14 Power-law distribution (Study 2A)

Questionnaire:

1 List five types of information/data that you put into your mobile device.

2 What other identifying information does your mobile device capture about you?

Next, related types were identified and combined for a smaller and more practical list. We coded specific apps into their relevant categories. Some categories are further aggregated together by similar functionality or synonyms to reduce the number of groups. Table 19 illustrates some examples. This combination resulted in 43 types where each type has a frequency of at least 10. Table 20 shows the 15 most popular types of information.

Table 19 Compilation of types

Types	New types	Final types
photos of family	photos of family	personal photos
pictures of me and my children		
photos of my dog	photos of pet	
photos of my cat		
my facebook information	facebook	social media
my tweets on twitter	twitter	
snapchat videos and photos	snapchat	
my physical activity	fitness	health
step counter	body movement	
how i sleep	health	
heart beats per minute		

Table 20 15 most popular types

Types of information	Frequency
personal photos	325
social media	285
location	236
contacts	197
health	146
entertainment	136
photos	127
banking	107
emails	103
texts	97
games	97
shopping	96
chat	95
passwords	80
browsing history	79

4.3 Study 2B

4.3.1 Measures

RQ1: What are the effects of the relevance of information types to different recipient, on the willingness to disclose?

We investigate the influence of recipient and type of information on mobile device users. Specifically, we examine the propensity to disclose certain types of information to particular recipients and how much do they think the information is necessary or relevant to that recipient.

We asked participants to rate their willingness to disclose certain types of information towards each contacts group and how necessary do they think. To measure willingness to disclose, we adapted four 7-point scales from Malhotra, Kim and Agarwal (2004). We measure perceived relevance by using three 7-point scales adapted from Zimmer, JC et al. (2010) (see Appendix B for complete questionnaire). We assessed their reliability and deemed the constructs to have an acceptable level of internal consistency (Nunnally, cited in Dinev & Hart 2004; Kline 2000), i.e. Cronbach's α values are 0.94 and 0.90 respectively. During the study, each respondent was given three vignettes to respond, where each vignette is a combination of types of information and contact groups.

There were five possible types of information and 15 possible contact groups that are compiled from Studies 1A and 2A. Since the resulting 75 combinations were too large to fit into a questionnaire, we divided them into three questionnaires instead. In each sub-questionnaire, we used five out of the 15 contact groups, while the types of information remained constant, resulting in 25 possible combinations.

To avoid repeat participations, the sub-questionnaires were conducted consecutively. We utilized TurkPrime (later rebranded as CloudResearch) to distribute surveys on MTurk. TurkPrime enabled us to exclude previous participants (Turkers) from participating in the subsequent studies.

4.3.2 Methodology

We advertised the questionnaires on MTurk for eight days in July 2019. Participants were asked to respond to our survey that we implemented on LimeSurvey. Participants spent 2 min and 20 seconds on average (median = 2 minutes 4 seconds) to complete the survey. Participants were paid USD 0.10 for completing the survey. We utilized similar measures as Study 2A's to minimise junk data. The questionnaire was approved by the Human Research Ethics Committee of our institution (equivalent to IRB approval in the US) before the recruitment of participants. Participants were presented with a participant information sheet (Appendix C) before responding to the questionnaire.

We took several measures suggested previously (Kuziemko et al. 2015; Page, Kobsa & Knijnenburg 2012; Shay et al. 2014) to minimise junk data. These measures are:

1 The survey is only advertised to Turkers located in the US. Location is also part of the demographic questions, and only responses with this location are considered valid.
2 Respondents were required to input a password that was only shown at completion to get paid. We cross-checked responses from MTurk and LimeSurvey to identify invalid responses with a blank or incorrect password. Respondents were not able to leave any blank answer.
3 We identified incomplete or out of topic responses.
4 The Likert scales are reversed alternately.
5 We identified responses with unrealistic completion times. They are not entirely invalid since those with good computer "reflex" could finish faster (Downs et al. 2010). They are further inspected using measures 1-5 to verify they are invalid.
6 We identified responses from the same IP address and further verified using measures 1-5.

We performed several regression diagnostics to validate the regression analysis. The Durbin-Watson statistic value was 1.99 ($p > 0.6$), suggesting no significant presence of autocorrelation. The Cook's distance value was 0.002, thus no evidence to suggest there were highly influential outliers.

Table 21 Demographics of Study 2B

Attribute	**Distribution**
Gender	Male (36.76%, n = 1062), Female (63.24%, n = 1827)
Age	18-25 (22.26%, n = 643), 26-35 (40.15%, n = 1160), 36-45 (20.84%, n = 602), 46-55 (10.76%, n = 311), 56 or above (5.99%, n = 173)
Education	Less than high school (0.69%, n = 20), High school (41.36%, n = 1195), Bachelor's (43.86%, n = 1267), Honours/Master's (12.22%, n = 353), Doctorate (1.87%, n = 54)
Employment	Student (7.41%, n = 214), Employed (57.29%, n = 1655), Self-employed (11.15%, n = 322), Employed student (7.75%, n = 224), Self-employed student (1.14%, n = 33), Unemployed (12.77%, n = 369), Retired (2.49%, n = 72)
Mobile	Android (49.43%, n = 1428), iOS (44.58%, n = 1288), Android and iOS (5.02%, n = 145), Others (0.97%, n = 28)
Experience	0-1 year (2.28%, n = 66), 2-4 years (11.46%, n = 331), 5-7 years (32.43%, n = 937), 8 years or more (53.82%, n = 1555)

We had a total of 3444 responses from LimeSurvey. With all the measures above, we removed 555 responses and remained with 2889 usable responses. Before the data analysis, we converted the 7-point Likert to a range of -3 to +3. The following Table 21 shows participants demographics.

4.3.3 Demographics

We compared the willingness to disclose among the demographics (Figure 15 & Table 22). We conducted Kruskal-Wallis (one-way ANOVA on ranks) to detect any differences. Kruskal-Wallis test was significant on age, suggesting at least one significant difference among age groups. Subsequent test between age groups using Conover test with Bonferroni adjustment was significant to suggest 18-25 age group is significantly higher than the rest of the group, except for the 26-35 age group; 26-35 is significantly higher than 56 or above.

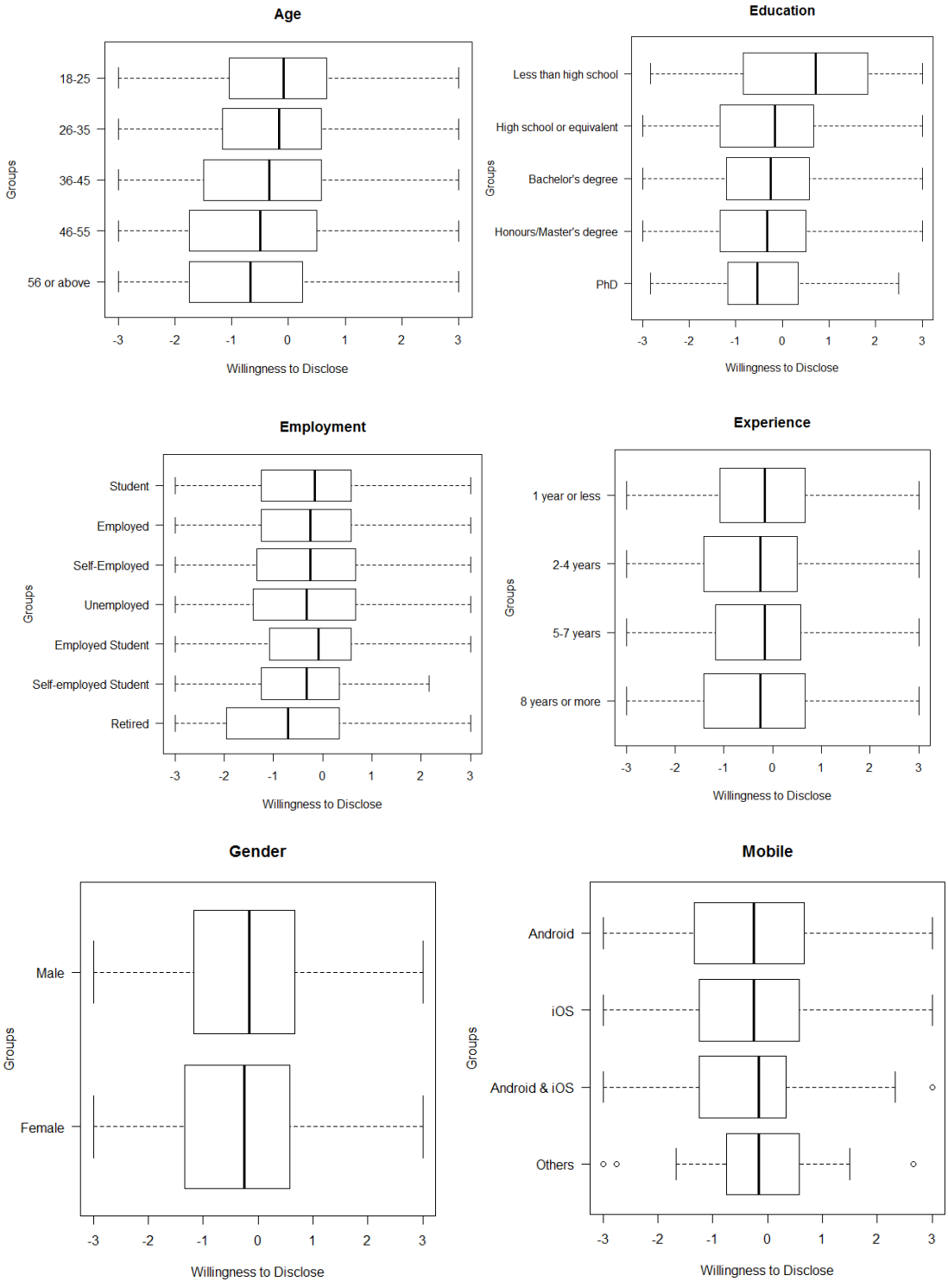

Figure 15 Willingness to disclose across demographics

Table 22 Demographics differences in willingness to disclose

Groups	Mean	Std.Dev	Test statistics
Age:			H = 27.997, df = 4, p < .001
18-25	-0.18	1.30	
26-35	-0.27	1.41	
36-45	-0.41	1.47	
46-55	-0.53	1.49	
56 or above	-0.67	1.48	
Education:			H = 6.768, df = 4, p = .149
Less than high school	0.41	1.83	
High school or equivalent	-0.34	1.46	
Bachelor's degree	-0.31	1.39	
Honours/Master's degree	-0.43	1.36	
PhD	-0.45	1.30	
Employment:			H = 8.823, df = 6, p = .184
Student	0.71	1.33	
Employed	0.68	1.42	
Self-Employed	0.62	1.42	
Unemployed	0.62	1.52	
Employed Student	0.79	1.28	
Self-employed Student	0.51	1.20	
Retired	0.32	1.55	
Experience:			H = 3.325, df = 3, p = .344
1 year or less	0.87	1.49	
2-4 years	0.60	1.43	
5-7 years	0.72	1.32	
8 years or more	0.64	1.47	
Gender:			p = .181
Male	0.71	1.39	
Female	0.64	1.44	
Mobile:			H = 0.435, df = 3, p = .933
Android	0.68	1.78	
iOS	0.64	1.55	
Android & iOS	0.70	1.40	
Others	0.66	0.66	

Table 23 Average indexes in different groups

Note: Each index column is colour-coded separately

Group	Frequency	Disclosure Index	Relevance Index
Acquaintances	942	-0.32	-0.04
Commercial Organizations	970	-0.99	0.15
Education Institutions	938	-0.39	0.15
Employers	964	-0.59	-0.16
Family	950	0.84	0.74
Financial Institutions	991	-1.13	-0.45
Friends	1004	0.55	0.47
Healthcare Organizations	958	-0.20	0.18
Non-profit Organizations	950	-0.76	-0.15

Disclosure Index

-1.2 0.8

Relevance Index

-0.5 0.7

Table 24 Average indexes of each information type

Note: each index column is colour-coded separately

Type	Frequency	Disclosure Index	Relevance Index
Contacts	1740	-0.73	-0.03
Health-related Information	1703	-0.16	0.34
Location	1827	0.15	0.42
Personal Photos	1734	-0.54	-0.14
Social Media Activity	1663	-0.41	-0.12

Disclosure Index

-0.8 0.2

Relevance Index

-0.15 0.5

Table 25 Average disclosure index

Disclosure	Contacts	Health-related Information	Location	Personal Photos	Social Media Activity
Acquaintances	-0.70	-0.76	-0.63	0.32	0.15
Commercial Organisations	-1.61	-0.99	-0.29	-1.36	-0.85
Education Institutions	-0.76	0.02	0.32	-1.17	-0.47
Employers	-0.76	-0.01	0.12	-1.26	-1.11
Family	0.49	1.19	1.04	0.80	0.71
Financial Institutions	-1.60	-1.25	0.18	-1.70	-1.28
Friends	0.09	0.19	0.58	1.13	0.67
Healthcare Organisations	-0.54	0.95	0.47	-1.03	-0.72
Non-profit Organisations	-1.15	-0.49	-0.45	-1.11	-0.69

-1.7 1.2

Table 26 Average relevance index

Relevance	Contacts	Health-related Information	Location	Personal Photos	Social Media Activity
Acquaintances	-0.03	-0.20	-0.05	0.11	-0.02
Commercial Organisations	0.07	-0.01	0.55	-0.01	0.07
Education Institutions	0.00	0.62	0.40	-0.36	0.07
Employers	-0.32	0.54	0.15	-0.67	-0.55
Family	0.44	1.30	1.02	0.65	0.30
Financial Institutions	-0.34	-0.69	0.34	-1.00	-0.63
Friends	0.28	0.42	0.65	0.62	0.35
Healthcare Organisations	0.03	1.16	0.66	-0.37	-0.50
Non-profit Organisations	-0.32	0.09	-0.01	-0.39	-0.17

-1 1.4

Table 27 Differences in disclosure and relevance indexes

Disclosure-Relevance	Contacts	Health-related Information	Location	Personal Photos	Social Media Activity
Acquaintances	0.67	0.56	0.58	0.21	0.17
Commercial Organisations	1.68	0.98	0.84	1.34	0.92
Education Institutions	0.76	0.60	0.08	0.81	0.54
Employers	0.45	0.55	0.03	0.59	0.56
Family	0.05	0.10	0.02	0.14	0.42
Financial Institutions	1.26	0.56	0.17	0.70	0.64
Friends	0.19	0.23	0.07	0.50	0.32
Healthcare Organisations	0.57	0.21	0.19	0.66	0.22
Non-profit Organizations	0.83	0.58	0.43	0.72	0.52

0.02 1.7

Correlation analysis showed that perceived relevance is significantly correlated with self-disclosure in both frequent and infrequent groups (Spearman $r = 0.48$, $p < 0.001$). The regression model showed relevance explained 26% of the variance in willingness to disclose (Table 28).

Table 28 Regression effect of relevance on willingness to disclose

Criterion	**Willingness to disclose**
Relevance	0.52 ($p < 0.001$)
R^2	.26
Adjusted R^2	.26
Significance	<0.001
Standard Error of Estimate	1.679
F-statistic	(1,8665) = 2972

4.4 Discussion

As part of our investigation on the relevance of the contextual integrity to the mobile ecosystem, especially the privacy aspect. In the previous chapter (Chapter 3), we conducted a study to investigate the influence of recipients—a contextual factor—on the users' privacy attitude. The results suggest that the different propensity of trust towards recipients can influence self-disclosure, despite having a privacy concern.

In this chapter, we studied the effect of a combination of contextual factors—recipients and type of information—on users' attitude. Specifically, we investigated how a combination of those factors can affect users' willingness to disclose and their perception of information relevance. From the results, we observed another form of privacy paradox—higher sensitivity does not necessarily result in lower disclosure. For instance, information types that are considered to be highly sensitive like health-related information and location (Madden et al. 2014) are not ranked in the lower half of the disclosure index (Table 24). Those types even rank higher in disclosure index than social media information, which was previously considered to be low sensitivity (Markos, Labrecque & Milne 2018). Previous studies posit that the paradox can be explained by information relevance (Nicholas et al. 2019; Zimmer, JC et al. 2010) which is a focus of our study.

We investigated the relationship between willingness to disclose and perceived relevance. The result suggests the user is more likely to disclose a piece of information when it is perceived to be relevant and mostly in line with existing studies. While the results suggest a significant relationship, it does not necessarily hold true in some instances. For instance, participants tend to perceive health-related information to be quite related on average, yet there is a slight resistance in disclosure (Table 24). When looking at different combinations of information type and recipient, we notice that while participants perceived "Contacts" and "Personal Photos" to be slightly relevant to "Commercial Organisations", yet they reacted strongly against disclosing those pieces of information to that group (Table 27). While the recipient group with the highest relevance index also has the highest disclosure index and vice versa, we do not observe a similar trend in information type. The information type with the highest relevance index also has the highest disclosure index, but the one with the lowest relevance index does not have the lowest disclosure index (Table 23 & Table 24).

Disclosure index may seem to be distinct between information types (Table 23). However, when we split it into different groups of the recipient, the distinction becomes erratic. For instance, when we compare "Contacts"—the information type with the lowest disclosure index (-0.73) on average—across different recipients, the value ranges from -1.61 to 0.49 (Table 25). Even though it is the lowest on average, when comparing across recipients, we notice it is not necessarily the lowest. In fact, it is only the lowest in two out of nine recipients. A similar discrepancy is also apparent in the Relevance index. Take

"Location" for example, which has the highest relevance index (0.42), when divided into varying recipients, the value ranges from -0.05 to 1.02 (Table 26). It is highest only in three out of nine recipient groups.

4.5 Conclusion

Findings from our studies in this chapter highlighted the influence of contextual factors—recipient and information type—on information exchange within the mobile ecosystem. The findings consequently lead to two practical implications; first, our results cast doubt over the established effects of "sensitivity" and its usefulness in PET. Existing studies (Milne et al. 2017; Mothersbaugh et al. 2011) posit that the significant relationship between sensitivity and willingness to disclose. If this assumption holds true, we can expect a consistent response in willingness to disclose a type of information across recipients. This study, however, could not reproduce such consistency (Table 25) and further demonstrate that sensitivity can vary according to the intended recipient. Second, while there is an evidence of a significant relationship between information relevance and disclosure, several discrepancies showed the relationship is not always clear-cut. Thus, we urge researchers to practice caution over the use of generic information relevance in predicting the tendency to disclose.

While not part of the main research question, we also examined the demographical differences. In this study, we did not find any significant difference between genders in propensity in disclosing information, nor in most demographics. This is contrary to our previous study and in turn, a study by Li, K, Lin and Wang (2015). We theorise that the initial difference information disclosure behaviour diminishes and reacted similarly as users take into consideration of information relevance. A notable exception is that there is evidence of a significant difference between age groups. Future study can examine more closely in how different age groups perceive information relevance.

5 Discussion and Conclusion

The overall aim of this thesis is to identify the factors that are essential to improve the current Privacy Enhancing Technologies (PET) in mobile platforms. The prevailing approach of privacy preservation in mobile devices through permissions management alone is not optimal due to the gap between flexibility and usability. Accommodating the diversity in contexts and also users' privacy preferences is complicated by privacy paradox.

Privacy paradox can lead to two issues. First, privacy recommendation system that relies on privacy profiling (e.g. (Knijnenburg 2014; Lin, J et al. 2014; Liu, B et al. 2016)) may not be as accurate. Privacy profiling typically works by typecasting a user into a particular category based on specific characteristics. When certain components, particularly privacy concern and trust, lost their prediction powers, this will subsequently affect the effectiveness of the privacy recommendation system. Second, it can exacerbate the vicious growth of deceptive privacy options and excessive data collection as businesses can assume users would continue surrendering their data (Walker 2016) under a flawed privacy policy (Acquisti & Grossklags 2005)—regardless of privacy concern.

In the first study (Chapter 3), we examined the prevalence of privacy paradox through the lens of the framework of contextual integrity (CI) (Nissenbaum 2010). The contextual integrity emphasises on the influence of contextual factors in our every day's mobile usage. We examined one such contextual factor is the recipients—user's attitude towards them. While there could be various dimensions of mentality, the results (from a sample of 301 users) suggest trust having a significant influence on the user's disclosure behaviour, particularly on the relationship between privacy concern and self-disclosure. The mediation effect of trust in our results suggest its significant role in determining users' self-disclosure despite the existence of privacy concern. The findings offer a meaningful explanation behind privacy paradox and corroborate with other related studies that suggest privacy concern does not necessarily inhibit self-disclosure (Heravi, Mubarak & Choo 2018; Taddicken 2014); where a user is more likely to disclose to a trusted recipient, despite having privacy concern.

In Chapter 3, we observed significant demographical differences on trust, privacy concern and self-disclosure. Female users have a higher tendency to disclose information compared to the male counterpart on a mobile device. We also observed that female users tend to exhibit greater privacy concern and trust. We found that age has a positive and significant association with privacy concern but insignificant with trust and self-disclosure. We speculate the contrasting results could be attributed to the additional constructs, i.e. trust and privacy concern, suggesting a potential interplay between them. A future study could investigate more in-depth of such relationship.

In the second study (Chapter 4), we observed another form of privacy paradox—higher sensitivity does not necessarily result in lower disclosure. Aside from privacy concern, previous studies (Nicholas et al. 2019; Zimmer, JC et al. 2010) have also posited the effects of information relevance on the self-disclosure. We examined the impact of two contextual factors—recipient and information type—on the relationship between information relevance and self-disclosure. While there is an evidence of a significant relationship between information relevance and disclosure, several discrepancies showed the relationship is not always clear-cut.

Our results highlight users' attitude on disclosure within the mobile ecosystem is often fraught with nuances and the use of generic information relevance in predicting the tendency to disclose may not be as effective as expected. Our results from the second study also cast doubt over the established effects of "sensitivity" and its usefulness in PET. We observed inconsistent response in willingness to disclose a type of information across recipients. This further demonstrates that sensitivity can vary according to the intended recipient. We conducted Study 1 and 2 in Chapter 3 and 4 respectively based on our framework as illustrated in Figure 16.

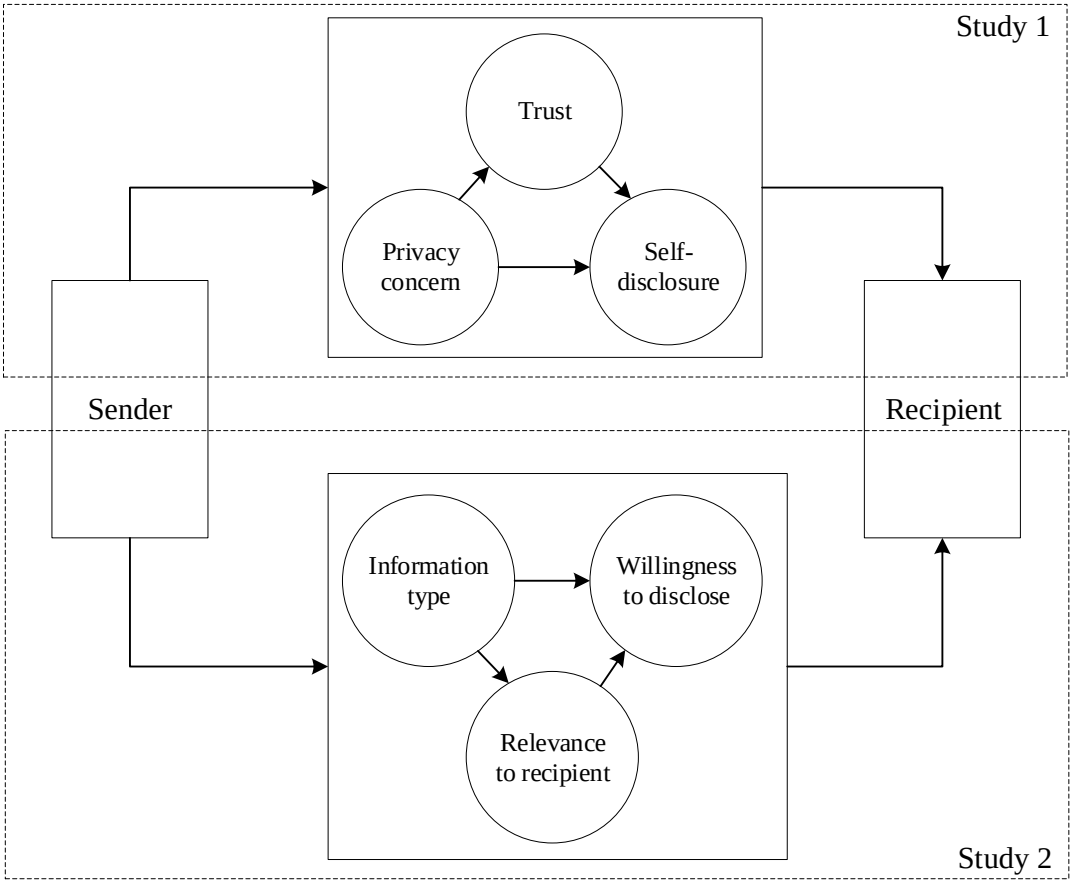

Figure 16 Contextual Privacy Framework

This thesis also contributes categorisations that are derived empirically (See Table 29 for the lists). In Chapter 3, we enumerated 15 groups of recipients most typically found in mobile devices from a sample of 282 users, before investigating the influence of trust and privacy concern on self-disclosure with the recipient. In Chapter 4, we compiled a list of 15 most commonly disclosed information types from a sample of 390 mobile users, for the investigation on the influence of relevance of information types on the willingness of disclosure towards typical groups of recipient. These empirically-derived lists add more weight to the results of the user studies that we have conducted and signify a significant improvement to previous studies that are often based on the investigator's assumptions.

Table 29 Empirically-derived classifications

Groups	Types of information
friends	personal photos
colleague	social media
family	location
financial	contacts
healthcare	health
relatives	entertainment
social media	photos
business	banking
retail	emails
employment	texts
school	games
npo	shopping
classmates	chat
acquaintances	passwords
strangers	browsing history

PET in mobile device is in dire need of a major overhaul when current mobile users have developed "learned helplessness" (Shklovski et al. 2014) and data "surrender" (Walker 2016) behaviours, numb to the invasive data activity (despite availability of PET) and simply acquiescing to any data request regardless of relevance. These behaviours are arguably stemmed from decision fatigue (Svirsky 2019; Utz et al. 2019), having been overwhelmed by a myriad of privacy options. However, this presents a dilemma, whereby despite users' desire for more control (Wijesekera et al. 2015), having more privacy options can induce more complexity causing the system to be less intuitive to use (Felt et al. 2012b; Liu, B et al. 2016) and subsequently leads to decision fatigue.

Despite attempts to ease the user's burden in permission management, this thesis uncovered several weaknesses in current approaches. Mobile platforms like Google Android tries to cut down on consent dialogue prompts by separating permission into *dangerous* and *normal* categories; dangerous permissions are prompted while normal permissions are granted straightaway. Privacy risks aside (Alepis & Patsakis 2018; Kywe et al. 2016), it is a dichotomy of sensitivity: sensitive and non-sensitive. In Chapter 4, we observed sensitivity received an inconsistent response from the participants and it can vary across intended recipients.

In addition to the existing approach in current mobile platforms, researchers have proposed privacy profiling systems. In profiling, a person is typecast into a particular category based on certain characteristics (e.g. privacy concern) and preferences of the

users. When applied to permission management, each profile would have distinct permission settings. The most obvious issue is that an 'unconcerned' profile would have highly permissive settings rendering the permission manager ineffective. We also doubt the effectiveness of profiling based on individual's privacy concern as we did not observe any significant effect of privacy concern in Chapter 3.

Overall, this thesis demonstrates the relevance of the CI framework in the mobile space and its potential to improve the current approach in PET, particularly the privacy recommendation system. The privacy recommendation system is a promising answer to the dilemma of having too little or too much privacy control. We believe by incorporating a crucial metric, "recipient", in addition to other contextual factors, the privacy recommendation system can advance its effectiveness. By taking into account of users' interactions with their recipients, the metric enables the ability to accommodate the ever-changing contexts and the diversity of users' privacy preferences, which are the weaknesses of the current system as detailed in Chapter 2. Future studies could take a deep dive upon the remaining contextual factors: *sender*, *subject* and *type* of information, and purpose of information collection.

The insights that we gained on mobile users' privacy attitudes can help integrating artificial intelligence (AI) into privacy recommendation systems. This integration helps to ensure continued improvements to PET's effectiveness, while minimising cognitive burden to the users. AI also can be utilised to incorporate decision heuristic of the CI framework (Nissenbaum 2010) into privacy recommendation or other areas of PET in the mobile platforms. Successful integration helps to create an ideal mobile platform that is feature-rich while respecting user privacy in dynamic contexts. This thesis is also relevant to the consumer tech providers to gain an understanding of their customers' privacy behaviour to design and incorporate PET that is more customised to the user's privacy needs. Companies have started to use stringent data privacy practice to gain a competitive advantage (Trindade 2020) by gaining customers' satisfaction and trust.

In this thesis, we recruited participants through a crowdsourcing platform. Future work could consider more crowdsourcing or recruitment platforms to obtain larger datasets. Our recruitment process did not involve choosing sample users randomly and might lead to selection bias. Alternative approaches that enable the use of random sampling include web scraping and application programming interfaces (APIs) provided

by the social media platforms that we can utilise to gauge public sentiments on the desired topics. Larger datasets combining with more sophisticated modelling could help uncover constructs that are not observable from the limited datasets utilised in this work. Since the participants involved in this work only expressed their views at a certain point in time, a longitudinal study can be conducted to evaluate whether the preferences could change over time.

Appendix A Questionnaires for Studies 1A and 1B

Study 1A

Simple Trust: Five-point scales anchored with "not at all" and "very strongly" (Molm, Takahashi & Peterson 2000)

1 How much do you trust the following mobile contact groups?

Study 1B

Simple Trust: Seven-point scales anchored with "not at all" and "very strongly"

1 How much do you trust the following mobile contact groups?

Individualized Trust Scale: Seven-point semantic scales (Wheeless & Grotz 1977)

On the scales that follow, please indicate your perception of the following mobile contact groups: <GROUP>. Select an option that represents your immediate "feelings".

1 Trustworthy/Untrustworthy
2 Distrustful of this group/Trustful of this group (r)
3 Confidential/Divulging
4 Exploitative/Benevolent (r)
5 Safe/Dangerous
6 Deceptive/Candid (r)
7 Not deceitful/Deceitful
8 Tricky/Straightforward (r)
9 Respectful/Disrespectful
10 Inconsiderate/Considerate (r)
11 Honest/Dishonest
12 Unreliable/Reliable (r)
13 Faithful/Unfaithful
14 Insincere/Sincere (r)
15 Careful/Careless

Privacy Concern: Seven-point scales anchored with “strongly disagree” and “strongly agree” (Malhotra, Kim & Agarwal 2004)

1 I am sensitive about the way online companies handle my personal information.
2 To me, it is the most important thing to keep my privacy intact form online companies.
3 I am concerned about threats to my personal privacy today.

Self-Disclosure: Seven-point scales anchored with “strongly disagree” and “strongly agree” (Wheeless & Grotz 1976)

1 My self-disclosures on mobile devices are always accurate reflections of who I really am.
2 I always feel completely sincere when I reveal my own feelings and experiences on mobile device.
3 I feel that I sometimes do not control what information I reveal about myself on mobile device.
4 I am confident that my expressions of my own feelings, emotions and experiences on mobile device are true reflections of myself.
5 I intimately disclose who I really am, openly and fully on mobile device.

(r): Reverse item

Appendix B Questionnaires for Studies 2A and 2B

Study 2A

1 List five types of information/data that you put into your mobile device.
2 What other identifying information does your mobile device capture about you?

Study 2B

Disclosure: Seven-point semantic scales (Malhotra, Kim & Agarwal 2004)

Please specify the extent to which you would reveal <TYPE> to <GROUP>, on the scales that follow.

1 Unlikely / likely
2 Not probable / probable
3 Possible / impossible (r)
4 Willing / unwilling (r)

Relevance: Seven-point semantic scales (Zimmer, JC et al. 2010)

Please indicate the extent of each factor for your above response.

5 Irrelevant / Relevant
6 Important / Unimportant (r)
7 Unnecessary / Necessary

(r): Reverse item

Appendix C Participant Information Sheet

Name: Ming Di Leom
Qualification: PhD Candidate in Information & Computer Science
Contact: [redacted]
Project title: User Privacy Preservation on Mobile Devices

Invitation to participate

You are all invited to participate in this research study by answering an anonymous online questionnaire which will take less than 10 minutes. Participation is voluntary and highly appreciated. Following provides detailed information about the research.

Purpose of this study

The aim of the proposed research is to provide a balanced approach to privacy that can ensure adequate and usable privacy while flexible to the diversity of privacy preference. Findings from a planned user survey on the topic of mobile device privacy will contribute to the development of a privacy preservation model. Based on the model, this research aims to evaluate existing tools and provide the necessary privacy preservation enhancements for mobile platforms.

Collected information

- The questionnaire is anonymous and no personal information will be asked nor collected.
- The researcher will take every care to remove responses from any identifying material as early as possible.
- Individual responses will remain confidential and no information which could lead to the identification of any individual will be released- unless required by law.
- As soon as practicable after data collection is complete, the research will delete the data from the online platform. However, the confidentiality of anonymous data stored on the Internet survey platform cannot be guaranteed.
- The questionnaire only contains questions regarding users' opinion about privacy in the mobile device.

- Survey results will be stored on the survey platform and the manager may have access to that data. However, since data collected is anonymous, the platform manager will not be able to identify participants.

What the participant will be expected to do:

It is expected that participants only answer the anonymous online questionnaire.

Possible risks:

- Since the questionnaire is anonymous, it is not anticipated that there are any risks to participation in this study beyond those encountered during everyday life.
- Participants are free to withdraw from the research project at any stage without affecting their status now or in the future.
- Once you submit your survey response, we are unable to remove your response as it will be impossible to identify your individual data.

Ethics

This project has been approved by the University of South Australia's Human Research Ethics Committee as No. 200270. If you have any ethical concerns about the project or questions about your rights as a participant, please contact the Executive Officer of this Committee,
Tel: [redacted]; Email: [redacted]

Data Storage

The data will be exported to the researcher's workstation and stored as computer files (Comma-separated values (CSV) files or Microsoft Office Excel spreadsheets) for 5 years, in accordance with the Australian Code of Responsible Research.

By completing and submitting the online survey, you are indicating that you have read and understood the Participant Information and give your consent to be involved in the research. After completion and submission of this survey with valid input of the completion code, you shall be compensated with the amount of USD 0.10 for your time and effort.

References

Abdella, J, Özuysal, M & Tomur, E 2016, 'CA-ARBAC: privacy preserving using context-aware role-based access control on Android permission system', *Security and Communication Networks*, vol. 9, no. 18, pp. 5977-95.

Acquisti, A & Grossklags, J 2005, 'Privacy and rationality in individual decision making', *IEEE Security and Privacy Magazine*, vol. 3, no. 1, pp. 26-33.

Acquisti, A, Taylor, CR & Wagman, L 2015, 'The economics of privacy', *Journal of Economic Literature*, vol. 52, no. 2, pp. 1-64.

Agarwal, Y & Hall, M 2013, 'ProtectMyPrivacy: detecting and mitigating privacy leaks on iOS devices using crowdsourcing', in H-H Chu & P Huang (eds), *Annual International Conference on Mobile Systems, Applications, and Services*, ACM, pp. 97-110.

Ahern, S, Eckles, D, Good, NS, King, S, Naaman, M & Nair, R 2007, 'Over-exposed?: privacy patterns and considerations in online and mobile photo sharing', in B Begole et al (eds), *Conference on Human Factors in Computing Systems*, ACM, pp. 357-66.

Ahn, J 2011, 'The effect of social network sites on adolescents' social and academic development: Current theories and controversies', *Journal of the American Society for Information Science and Technology*, vol. 62, no. 8, pp. 1435-45.

Alaqra, AS & Wästlund, E 2019, 'Reciprocities or incentives? Understanding privacy intrusion perspectives and sharing behaviors', *HCI for Cybersecurity, Privacy and Trust*, vol. 11594, pp. 355-70.

Alepis, E & Patsakis, C 2018, 'Unravelling security issues of runtime permissions in Android', *Journal of Hardware and Systems Security*, vol. 3, no. 1, pp. 45-63.

Allmer, T 2011, 'A critical contribution to theoretical foundations of privacy studies', *Journal of Information, Communication and Ethics in Society*, vol. 9, no. 2, pp. 83-101.

Altman, I 1975, *The environment and social behavior: privacy, personal space, territory, crowding*, Wadsworth Publishing, Belmont, CA.

Alvarez, R, Levenson, J, Sheatsley, R & McDaniel, P 2019, 'Application transiency: towards a fair trade of personal information for application services', in S Chen et al (eds), *Security and Privacy in Communication Networks*, vol. 305, Springer, Cham, pp. 47-66.

Amadeo, R 2015, *Android 6.0 Marshmallow, thoroughly reviewed*, Ars Technica, viewed 11 September 2016, <http://arstechnica.com/gadgets/2015/10/android-6-0-marshmallow-thoroughly-reviewed/>.

Amadeo, R 2017, *Android 8.0 Oreo, thoroughly reviewed*, Ars Technica, viewed 1 October 2019, <https://arstechnica.com/gadgets/2017/09/android-8-0-oreo-thoroughly-reviewed/>.

Amazon 2018, *Accceptable use policy*, viewed 6 March 2019, <https://www.mturk.com/worker/acceptable-use-policy>.

Apple 2013, *Protecting the user's privacy*, viewed 19 January 2019, <https://developer.apple.com/documentation/uikit/core_app/protecting_the_user_s_privacy>.

Apthorpe, N, Shvartzshnaider, Y, Mathur, A, Reisman, D & Feamster, N 2018, 'Discovering smart home internet of things privacy norms using contextual integrity', in S Santini (ed), *Interactive, Mobile, Wearable and Ubiquitous Technologies*, ACM, pp. 1-23.

Bai, G, Gu, L, Feng, T, Guo, Y & Chen, X 2010, 'Context-aware usage control for Android', in S Jajodia & J Zhou (eds), *Security and Privacy in Communication Networks*, vol. 50, Springer, Berlin, Heidelberg, pp. 326-43.

Baokar, A 2016, 'A contextually-aware privacy-preserving Android permission model', Master's thesis, Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA.

Barkhuus, L 2012, 'The mismeasurement of privacy: using contextual integrity to reconsider privacy in HCI', in JA Konstan (ed), *Conference on Human Factors in Computing Systems*, ACM, pp. 367-76.

Barnes, SB 2006, *A privacy paradox: social networking in the United States*, First Monday, viewed 23 March 2018, <https://doi.org/10.5210/fm.v11i9.1394>.

Barth, A, Datta, A, Mitchell, JC & Nissenbaum, H 2006, 'Privacy and contextual integrity: framework and applications', in H Orman (ed), *Symposium on Security and Privacy*, IEEE, pp. 1-15.

Behrend, TS, Sharek, DJ, Meade, AW & Wiebe, EN 2011, 'The viability of crowdsourcing for survey research', *Behav Res Methods*, vol. 43, no. 3, Sep, pp. 800-13.

Benisch, M, Kelley, PG, Sadeh, N & Cranor, LF 2010, 'Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs', *Personal and Ubiquitous Computing*, vol. 15, no. 7, pp. 679-94.

Benthall, S, Gürses, S & Nissenbaum, H 2017, 'Contextual integrity through the lens of computer science', *Foundations and Trends® in Privacy and Security*, vol. 2, no. 1, pp. 1-69.

Bokhorst, M 2016, *XPrivacy*, viewed 5 June 2016, <https://www.xprivacy.eu/>.

Bonné, B, Peddinti, ST, Bilogrevic, I & Taft, N 2017, 'Exploring decision making with Android's runtime permission dialogs using in-context surveys', in ME Zurko (ed), *Symposium on Usable Privacy and Security*, USENIX, pp. 195-210.

Bosu, A, Liu, F, Yao, D & Wang, G 2017, 'Collusive data leak and more: large-scale threat analysis of inter-app communications', in R Karri et al (eds), *Symposium on Information, Computer and Communications Security*, ACM, pp. 71-85.

Boyd, A 2019, *What half of iPhone users don't know about their privacy: new poll*, Mozilla Foundation, viewed 12 Nov 2019, <https://foundation.mozilla.org/en/blog/what-half-of-iphone-users-dont-know-about-their-privacy-new-poll/>.

Brandimarte, L, Acquisti, A & Loewenstein, G 2012, 'Misplaced confidences: privacy and the control paradox', *Social Psychological and Personality Science*, vol. 4, no. 3, pp. 340-47.

Brandtzæg, PB, Lüders, M & Skjetne, JH 2010, 'Too many Facebook "friends"? Content sharing and sociability versus the need for privacy in social network sites', *International Journal of Human-Computer Interaction*, vol. 26, no. 11-12, pp. 1006-30.

Brunton, F & Nissenbaum, H 2019, *The fantasy of opting out*, MIT Press, viewed 2 Nov 2019, <https://thereader.mitpress.mit.edu/the-fantasy-of-opting-out/>.

Buchanan, T, Paine, C, Joinson, AN & Reips, U-D 2007, 'Development of measures of online privacy concern and protection for use on the Internet', *Journal of the American Society for Information Science and Technology*, vol. 58, no. 2, pp. 157-65.

Buhrmester, M, Kwang, T & Gosling, SD 2011, 'Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data?', *Perspect Psychol Sci*, vol. 6, no. 1, Jan, pp. 3-5.

Burke, D 2019, *Introducing Android Q beta*, Android Developers Blog, viewed 1 October 2019, <https://android-developers.googleblog.com/2019/03/introducing-android-q-beta.html>.

Butler, JK & Cantrell, RS 2016, 'A behavioral decision theory approach to modeling dyadic trust in superiors and subordinates', *Psychological Reports*, vol. 55, no. 1, pp. 19-28.

Bylund, CL, Peterson, EB & Cameron, KA 2012, 'A practitioner's guide to interpersonal communication theory: an overview and exploration of selected theories', *Patient Education and Counseling*, vol. 87, no. 3, pp. 261-67.

Camp, LJ 1999, 'Web security and privacy: an American perspective', *The Information Society*, vol. 15, no. 4, pp. 249-56.

Casler, K, Bickel, L & Hackett, E 2013, 'Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing', *Computers in Human Behavior*, vol. 29, no. 6, pp. 2156-60.

Chakraborty, S, Shen, C, Raghavan, KR, Shoukry, Y, Millar, M & Srivastava, M 2014, 'ipShield: a framework for enforcing context-aware privacy', in R Mahajan & I Stoica (eds), *Symposium on Networked Systems Design and Implementation*, USENIX, pp. 143-56.

Chan, YE, Stalker, LLH, Lyon, D, Pavlov, A, Sharpe, J, Smith, E, Trottier, D & Zureik, E 2008, *The globalization of personal data project: an international survey on privacy and surveillance*, The Surveillance Project, Queen's University. viewed 12 January 2017, <http://www.sscqueens.org/sites/default/files/2008_Surveillance_Project_International_Survey_Findings_Summary.pdf>.

Chen, HT & Chen, W 2015, 'Couldn't or wouldn't? The influence of privacy concerns and self-efficacy in privacy management on privacy protection', *Cyberpsychology, Behavior and Social Networking*, vol. 18, no. 1, pp. 13-19.

Chen, J, Ping, W, Xu, Y & Tan, BC 2009, 'Am I afraid of my peers? Understanding the antecedents of information privacy concerns in the online social context', in H Chen & S Slaughter (eds), *International Conference on Information Systems*, Association for Information Systems, pp. 1-18.

Civil Society 2009, 'Madrid declaration', *31st annual meeting of the International Conference of Privacy and Data Protection Commissioners*. <https://ec.europa.eu/energy/sites/ener/files/documents/Madrid%20declaration.pdf>.

Clauset, A, Shalizi, CR & Newman, MEJ 2009, 'Power-law distributions in empirical data', *SIAM Review*, vol. 51, no. 4, pp. 661-703.

Colnago, J, Feng, Y, Palanivel, T, Pearman, S, Ung, M, Acquisti, A, Cranor, LF & Sadeh, N 2020, 'Informing the design of a personalized privacy assistant for the Internet of Things', *Conference on Human Factors in Computing Systems*, ACM.

Committee on Energy and Commerce 2001, 'What consumers have to say about information privacy', *Hearing before the Subcommittee on Commerce, Trade and Consumer Protection, 107th US Congress*.

Conti, M, Nguyen, VTN & Crispo, B 2011, 'CRePE: context-related policy enforcement for Android', in M Burmester et al (eds), *Information Security*, vol. 6531, Springer, Berlin, Heidelberg, pp. 331-45.

Crowston, K 2012, 'Amazon Mechanical Turk: A Research Tool for Organizations and Information Systems Scholars', vol. 389, pp. 210-21.

Culnan, MJ & Armstrong, PK 1999, 'Information privacy concerns, procedural fairness, and impersonal trust: an empirical investigation', *Organization Science*, vol. 10, no. 1, pp. 104-15.

Cunningham, A 2016, *iOS 9.3 brings multi-user mode to iPads*, Ars Technica, viewed 21 Jan 2018, <https://arstechnica.com/gadgets/2016/01/ios-9-3-brings-multi-user-mode-to-ipads-along-with-more-features-and-fixes/>.

Cyphers, B & Gebhart, G 2019, *Behind the one-way mirror: a deep dive into the technology of corporate surveillance*, Electronic Frontier Foundation, viewed 15 Dec 2019, <https://www.eff.org/wp/behind-the-one-way-mirror>.

DeCew, J 2015, *Privacy*, Spring 2015 edn, The Stanford Encyclopedia of Philosophy, viewed 23 March 2016, <https://plato.stanford.edu/archives/spr2015/entries/privacy/>.

Dinev, T & Hart, P 2004, 'Internet privacy concerns and their antecedents - measurement validity and a regression model', *Behaviour & Information Technology*, vol. 23, no. 6, pp. 413-22.

Doorey, AM, Wilcox, GB & Easin, MS 2017, 'Consumer privacy and the new mobile commerce', in AC Scheinbaum (ed), *The Dark Side of Social Media: A Consumer Psychology Perspective*, Routledge, New York, NY, pp. 179-200.

Downs, JS, Holbrook, MB, Sheng, S & Cranor, LF 2010, 'Are your participants gaming the system? Screening Mechanical Turk workers', in E Mynatt, K Edwards & T Rodden (eds), *Conference on Human Factors in Computing Systems*, ACM, pp. 2399-402.

Fang, Z, Han, W & Li, Y 2014, 'Permission based Android security: issues and countermeasures', *Computers & Security*, vol. 43, pp. 205-18.

Fazlioglu, M 2019, 'Beyond the nature of data obstacles to protecting sensitive information in the European Union and the United States', *Fordham Urban Law Journa*, vol. 46, no. 2, pp. 271-306.

Felt, AP, Chin, E, Hanna, S, Song, D & Wagner, D 2011, 'Android permissions demystified', in Y Chen, G Danezis & V Shmatikov (eds), *Conference on Computer and Communications Security*, ACM, pp. 627-38.

Felt, AP 2012, 'Towards comprehensible and effective permission systems', PhD thesis, University of California, Berkeley, CA.

Felt, AP, Egelman, S, Finifter, M, Akhawe, D & Wagner, D 2012a, 'How to ask for permission', *Workshop on Hot Topics in Security*, USENIX, pp. 1-6.

Felt, AP, Egelman, S & Wagner, D 2012, 'I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns', in T Yu, W Enck & X Jiang (eds), *Workshop on Security and Privacy in Smartphones and Mobile Devices*, ACM, pp. 33-44.

Felt, AP, Ha, E, Egelman, S, Haney, A, Chin, E & Wagner, D 2012b, 'Android permissions: user attention, comprehension, and behavior', in LF Cranor (ed), *Symposium on Usable Privacy and Security*, ACM, pp. 1-14.

Flaherty, DH 2014, *Protecting privacy in surveillance societies: the Federal Republic of Germany, Sweden, France, Canada, and the United States*, University of North Carolina Press, Chapel Hill, NC.

Fogues, RL, Such, JM, Espinosa, A & Garcia-Fornes, A 2018, 'Tie and tag: a study of tie strength and tags for photo sharing', *PLoS One*, vol. 13, no. 8, pp. 1-22.

Frik, A, Nurgalieva, L, Bernd, J, Lee, JS, Schaub, F & Egelman, S 2019, 'Privacy and security threat models and mitigation strategies of older adults', in HR Lipford (ed), *Symposium on Usable Privacy and Security*, USENIX, pp. 21-40.

FTC 2013, *Android flashlight app developer settles FTC charges it deceived consumers*, Federal Trade Commission, viewed 29 July 2016, <https://www.ftc.gov/news-events/press-releases/2013/12/android-flashlight-app-developer-settles-ftc-charges-it-deceived>.

FTC 2016, *FTC issues warning letters to app developers using 'Silverpush' code*, Federal Trade Commission, viewed 18 March 2016, <https://www.ftc.gov/news-events/press-releases/2016/03/ftc-issues-warning-letters-app-developers-using-silverpush-code>.

FTC 2017, *The Equifax data breach: what to do*, Federal Trade Commission, viewed 30 Oct 2019, <https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-do>.

Gao, H, Guo, C, Huang, D, Hou, X, Wu, Y, Xu, J, He, Z & Bai, G 2020, 'Autonomous permission recommendation', *IEEE Access*, vol. 4, pp. 1-15.

Ginosar, A & Ariel, Y 2017, 'An analytical framework for online privacy research: what is missing?', *Information & Management*, vol. 54, no. 7, pp. 948-57.

Gogus, A & Saygın, Y 2019, 'Privacy perception and information technology utilization of high school students', *Heliyon*, vol. 5, no. 5, pp. 1-9.

Google 2017, *Supporting multiple users*, viewed 21 Jan 2018, <https://source.android.com/devices/tech/admin/multi-user>.

Google 2018, *App permissions best practices*, viewed 19 January 2019, <https://developer.android.com/training/permissions/usage-notes>.

Grace, MC, Zhou, W, Jiang, X & Sadeghi, A-R 2012, 'Unsafe exposure analysis of mobile in-app advertisements', in M Krunz et al (eds), *Conference on Security and Privacy in Wireless and Mobile Networks*, ACM, pp. 101-12.

Greene, K & Faulkner, SL 2002, 'Expected versus actual responses to disclosure in relationships of HIV-positive African American adolescent females', *Communication Studies*, vol. 53, no. 4, pp. 297-317.

Grob, M 2017, *We are making on-device AI ubiquitous*, Qualcomm, viewed 16 May 2020, <https://www.qualcomm.com/news/onq/2017/08/16/we-are-making-device-ai-ubiquitous>.

Grodzinsky, FS & Tavani, HT 2008, 'Online file sharing: resolving the tensions between privacy and property interests', *ACM SIGCAS Computers and Society*, vol. 38, no. 4, pp. 28-39.

Grodzinsky, FS & Tavani, HT 2010, 'Applying the "contextual integrity" model of privacy to personal blogs in the blogoshere', *International Journal of Internet Research Ethics*, vol. 3, pp. 38-47.

Grodzinsky, FS & Tavani, HT 2011, 'Privacy in "the cloud": applying Nissenbaum's theory of contextual integrity', *ACM SIGCAS Computers and Society*, vol. 41, no. 1, pp. 38-47.

Grossklags, J & Acquisti, A 2007, 'When 25 cents is too much: an experiment on willingness-to-sell and willingness-to-protect personal information', *Workshop on the Economics of Information Security*, Carnegie Mellon University, pp. 1-22.

Habib, H, Zou, Y, Jannu, A, Sridhar, N, Swoopes, C, Acquisti, A, Cranor, LF, Sadeh, N & Schaub, F 2019, 'An empirical analysis of data deletion and opt-out choices on 150 websites', in HR Lipford (ed), *Symposium on Usable Privacy and Security*, USENIX, pp. 387-406.

Habib, H, Pearman, S, Wang, J, Zou, Y, Acquisti, A, Cranor, LF, Sadeh, N & Schaub, F 2020, '"It's a scavenger hunt": usability of websites' opt-out and data deletion choices', in R Bernhaupt & F Mueller (eds), *Conference on Human Factors in Computing Systems*, ACM, pp. 1-12.

Hann, I-H, Hui, K-L, Lee, S-Y & Png, I 2007, 'Overcoming online information privacy concerns: an information-processing theory approach', *Journal of Management Information Systems*, vol. 24, no. 2, pp. 13-42.

Hanson, J, Wei, M, Veys, S, Kugler, M, Strahilevitz, L & Ur, B 2020, 'Taking data out of context to hyper-personalize ads: crowdworkers' privacy perceptions and decisions to disclose private information', in R Bernhaupt & F Mueller (eds), *Conference on Human Factors in Computing Systems*, ACM, pp. 1-13.

Heravi, A, Mubarak, S & Choo, K-KR 2018, 'Information privacy in online social networks: uses and gratification perspective', *Computers in Human Behavior*, vol. 84, pp. 441-59.

Hoofnagle, CJ & Urban, JM 2014, 'Alan Westin's privacy homo economicus', *Wake Forest Law Review*, vol. 49, pp. 261-317.

Howe, DC & Nissenbaum, H 2017, 'Engineering privacy and protest: a case study of AdNauseam', *International Workshop on Privacy Engineering*, vol. 1873, pp. 1-8.

Huang, DY, Apthorpe, N, Acar, G, Li, F & Feamster, N 2019, 'IoT inspector: crowdsourcing labeled network traffic from smart home devices at scale', *arXiv:1909.09848*, pp. 1-14.

Huang, H-Y 2019, 'Examining older users' online privacy-enhancing experience from a human-computer interaction perspective', PhD thesis, University of Illinois at Urbana-Champaign, Urbana, IL.

Isaac, M 2017, *Uber C.E.O. plays with fire*, New York Times, viewed 23 April 2017, <https://www.nytimes.com/2017/04/23/technology/travis-kalanick-pushes-uber-and-himself-to-the-precipice.html>.

Jeon, J, Micinski, KK, Vaughan, JA, Fogel, A, Reddy, N, Foster, JS & Millstein, T 2012, 'Dr. Android and Mr. Hide: fine-grained permissions in Android applications', in W Enck & X Jiang (eds), *Workshop on Security and Privacy in Smartphones and Mobile Devices*, ACM, pp. 3-14.

Jia, YJ, Chen, QA, Wang, S, Rahmati, A, Fernandes, E, Mao, ZM & Prakash, A 2017, 'ContexIoT: Towards providing contextual integrity to appified IoT platforms', in J Amann (ed), *Network and Distributed System Security Symposium*, Internet Society, pp. 1-15.

Johnson-George, C & Swap, WC 1982, 'Measurement of specific interpersonal trust: construction and validation of a scale to assess trust in a specific other', *Journal of Personality and Social Psychology*, vol. 43, no. 6, pp. 1306-17.

Johnson, M 2012, 'Toward usable access control for end-users: a case study of Facebook privacy settings', PhD thesis, Graduate School of Arts and Sciences, Columbia University, New York City, NY.

Joinson, A, Reips, U-D, Buchanan, T & Schofield, CBP 2010, 'Privacy, trust, and self-disclosure online', *Human-Computer Interaction*, vol. 25, no. 1, pp. 1-24.

Jung, J, Han, S & Wetherall, D 2012, 'Enhancing mobile application permissions with runtime feedback and constraints', in W Enck & X Jiang (eds), *Workshop on Security and Privacy in Smartphones and Mobile Devices*, ACM, pp. 45-50.

Jung, K & Park, S 2013, 'Context-aware role based access control using user relationship', *International Journal of Computer Theory and Engineering*, vol. 5, no. 3, pp. 533-37.

Kaldestad, ØH & Myrstad, F 2018, *Deceived by design*, Norwegian Consumer Council. <https://www.forbrukerradet.no/side/facebook-and-google-manipulate-users-into-sharing-personal-data/>.

Kehr, F, Kowatsch, T, Wentzel, D & Fleisch, E 2015, 'Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus', *Information Systems Journal*, vol. 25, no. 6, pp. 607-35.

Kelley, PG, Benisch, M, Cranor, LF & Sadeh, N 2011a, 'When are users comfortable sharing locations with advertisers?', in D Tan, B Begole & W Kellogg (eds), *Conference on Human Factors in Computing Systems*, ACM, pp. 2449-52.

Kelley, PG, Brewer, R, Mayer, Y, Cranor, LF & Sadeh, N 2011b, 'An investigation into Facebook friend grouping', in P Campos et al (eds), *Human-Computer Interaction - INTERACT*, vol. 6948, Springer, Berlin, Heidelberg, pp. 216-33.

Kelley, PG, Consolvo, S, Cranor, LF, Jung, J, Sadeh, N & Wetherall, D 2012, 'A conundrum of permissions: installing applications on an Android smartphone', *Financial Cryptography and Data Security*, pp. 68-79.

Kim, E 2015, *Creating better user experiences on Google Play*, Android Developers Blog, viewed 12 April 2017, <https://android-developers.googleblog.com/2015/03/creating-better-user-experiences-on.html>.

King, J 2012, *How come I'm allowing strangers to go through my phone? Smartphones and privacy expectations*, SSRN Electronic Journal, viewed 13 June 2019, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2493412>.

King, J 2014, 'Taken out of context: an empirical analysis of Westin privacy scale', in A Acquisti et al (eds), *Workshop on Privacy Personas and Segmentation*, USENIX, pp. 1-8.

King, J 2018, 'Privacy, disclosure, and social exchange theory', PhD thesis, University of California, Berkeley.

King, J & Katsanevas, A 2019, 'Blending contextual integrity and social exchange theory: assessing norm building under conditions of "Informational inequality"', in S Egelman, H Nissenbaum & Y Shvartzshnaider (eds), *Symposium on Applications of Contextual Integrity*, ICSI, pp. 1-4.

Kline, P 2000, *Handbook of psychological testing*, 2nd edn, Routledge, New York.

Knausenberger, J, Hellmann, JH & Echterhoff, G 2015, 'When virtual contact is all you need: subtle reminders of Facebook preempt social-contact restoration after exclusion', *European Journal of Social Psychology*, vol. 45, no. 3, pp. 279-84.

Knijnenburg, BP 2014, 'Information disclosure profiles for segmentation and recommendation', in A Acquisti et al (eds), *Workshop on Privacy Personas and Segmentation*, USENIX, pp. 1-4.

Kokolakis, S 2017, 'Privacy attitudes and privacy behaviour: a review of current research on the privacy paradox phenomenon', *Computers & Security*, vol. 64, pp. 122-34.

Krupa, Y & Vercouter, L 2012, 'Handling privacy as contextual integrity in decentralized virtual communities: the PrivaCIAS framework', *Web Intelligence and Agent Systems: An International Journal*, vol. 10, no. 1, pp. 105-16.

Kumaraguru, P & Cranor, LF 2005, *Privacy indexes: a survey of Westin's studies*, Institute for Software Research International, Carnegie Mellon University, viewed 2 August 2018, <https://www.cs.cmu.edu/~ponguru/CMU-ISRI-05-138.pdf>.

Kuziemko, I, Norton, MI, Saez, E & Stantcheva, S 2015, 'How elastic are preferences for redistribution? evidence from randomized survey experiments', *American Economic Review*, vol. 105, no. 4, pp. 1478-508.

Kywe, SM, Li, Y, Petal, K & Grace, M 2016, 'Attacking Android smartphone systems without permissions ', in H Sarrafzadeh (ed), *Annual Conference on Privacy, Security and Trust*, IEEE, pp. 147-56.

Larson, R & Csikszentmihalyi, M 2014, 'The experience sampling method', *Flow and the Foundations of Positive Psychology*, Springer, Netherlands, pp. 21-34.

Larzelere, RE & Huston, TL 1980, 'The dyadic trust scale: toward understanding interpersonal trust in close relationships', *Journal of Marriage and the Family*, vol. 42, no. 3, pp. 595-604.

Laufer, RS, Prohansky, HM & Wolfe, M 1973, 'Some analytic dimensions of privacy', in R Küller (ed), *Architecutral Psychology*, Studentlitteratur AB, Scania, Sweden.

Laufer, RS & Wolfe, M 1977, 'Privacy as a concept and a social issue: a multidimensional developmental theory', *Journal of Social Issues*, vol. 33, no. 3, pp. 22-42.

Li, K, Lin, Z & Wang, X 2015, 'An empirical analysis of users' privacy disclosure behaviors on social network sites', *Information & Management*, vol. 52, no. 7, pp. 882-91.

Li, Y 2012, 'Theories in online information privacy research: a critical review and an integrated framework', *Decision Support Systems*, vol. 54, no. 1, pp. 471-81.

Lin, J, Sadeh, N, Amini, S, Lindqvist, J, Hong, JI & Zhang, J 2012, 'Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing', in A Dey, H-H Chu & G Hayes (eds), *Conference on Ubiquitous Computing*, ACM, pp. 501-10.

Lin, J, Liu, B, Sadeh, N & Hong, JI 2014, 'Modeling users' mobile app privacy preferences: restoring usability in a sea of permission settings', in LF Cranor (ed), *Symposium On Usable Privacy and Security*, USENIX, pp. 199-212.

Lin, K-Y & Lu, H-P 2011, 'Why people use social networking sites: An empirical study integrating network externalities and motivation theory', *Computers in Human Behavior*, vol. 27, no. 3, pp. 1152-61.

Lin, S-W & Liu, Y-C 2012, 'The effects of motivations, trust, and privacy concern in social networking', *Service Business*, vol. 6, no. 4, pp. 411-24.

Liu, B, Andersen, MS, Schaub, F, Almuhimedi, H, Zhang, S, Sadeh, N, Acquisti, A & Agarwal, A 2016, 'Follow my recommendations: a personalized privacy assistant for mobile app permissions', in ME Zurko (ed), *Symposium on Usable Privacy and Security*, USENIX, pp. 27-41.

Liu, B 2019, 'Can Machine Learning Help People Configure Their Mobile App Privacy Settings?', PhD thesis, Carnegie Mellon University, Pittsburgh PA.

Liu, R 2015, 'Mitigating privacy risks of smartphones in mobile computing', Master's thesis, Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong.

Liu, R, Cao, J, Yang, L & Zhang, K 2015, 'PriWe: recommendation for privacy settings of mobile apps based on crowdsourced users' expectations', *IEEE Software,* vol. 32, no. 2, pp. 150-57.

Liu, R, Cao, J, VanSyckel, S & Gao, W 2016, 'PriMe: Human-centric privacy measurement based on user preferences towards data sharing in mobile participatory sensing systems', in M Kumar & A Seneviratne (eds), *International Conference on Pervasive Computing and Communications*, IEEE, pp. 1-8.

Lord, B 2016, *An important message about Yahoo user security*, viewed 20 Dec 2018, <https://yahoo.tumblr.com/post/150781911849/an-important-message-about-yahoo-user-security>.

Madden, M, Rainie, L, Zickuhr, K, Duggan, M & Smith, A 2014, *Public perceptions of privacy and security in the post-Snowden era,* Pew Research Center, viewed 5 Nov 2019, <https://www.pewresearch.org/internet/2014/11/12/public-privacy-perceptions/>.

Malhotra, NK, Kim, SS & Agarwal, J 2004, 'Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model', *Information Systems Research,* vol. 15, no. 4, pp. 336-55.

Markos, E, Labrecque, LI & Milne, GR 2018, 'A new information lens: the self-concept and exchange context as a means to understand information sensitivity of anonymous and personal identifying information', *Journal of Interactive Marketing,* vol. 42, pp. 46-62.

Marmion, V, Millard, DE, Gerding, EH & Stevenage, SV 2019, 'The willingness of crowds: cohort disclosure preferences for personally identifying information', in J Pfeffer et al (eds), *International AAAI Conference on Web and Social Media*, AAAI, pp. 358-68.

Martin, K & Nissenbaum, H 2016, 'Measuring privacy: an empirical test using context to expose confounding variables', *Columbia Science and Technology Law Review,* vol. 18, no. 1, pp. 176-218.

Martin, K & Shilton, K 2016a, 'Putting mobile application privacy in context: An empirical study of user privacy expectations for mobile devices', *The Information Society,* vol. 32, no. 3, pp. 200-16.

Martin, K & Shilton, K 2016b, 'Why experience matters to privacy: how context-based experience moderates consumer privacy expectations for mobile applications', *Journal of the Association for Information Science and Technology*, vol. 67, no. 8, pp. 1871-82.

Martin, K & Shilton, K 2018, 'Mobile privacy expectations: how privacy is respected with mobile devices', in E Selinger, J Polonetsky & O Tene (eds), *The Cambridge Handbook of Consumer Privacy*, Cambridge University Press, Cambridge, UK, pp. 85-101.

Mass, CF & Madaus, LE 2014, 'Surface pressure observations from smartphones: a potential revolution for high-resolution weather prediction?', *Bulletin of the American Meteorological Society*, vol. 95, no. 9, pp. 1343-49.

Masur, PK & Scharkow, M 2016, 'Disclosure management on social network sites: individual privacy perceptions and user-directed privacy strategies', *Social Media + Society*, vol. 2, no. 1, pp. 1-13.

Mathur, A, Acar, G, Friedman, MJ, Lucherini, E, Mayer, J, Chetty, M & Narayanan, A 2019, 'Dark patterns at scale: findings from a crawl of 11k shopping websites', in E Gilbert & K Karahalios (eds), *Conference on Computer-Supported Cooperative Work and Social Computing*, ACM, pp. 1-32.

McDonald, N & Forte, A 2020, 'The politics of privacy theories: moving from norms to vulnerabilities', *Conference on Human Factors in Computing Systems*, ACM.

Mehrnezhad, M, Toreini, E, Shahandashti, SF & Hao, F 2017, 'Stealing PINs via mobile sensors: actual risk versus user perception', *International Journal of Information Security*, vol. 17, pp. 291-313.

Mejias, UA & Couldry, N 2019, 'Datafication', *Internet Policy Review*, vol. 8, no. 4.

Miettinen, M, Heuser, S, Kronz, W, Sadeghi, A-R & Asokan, N 2014, 'ConXsense: automated context classification for context-aware access control', in S Moriai, T Jaeger & K Sakurai (eds), *Symposium on Information, Computer and Communications Security*, ACM, pp. 293-304.

Milne, GR, Pettinico, G, Hajjat, FM & Markos, E 2017, 'Information sensitivity typology: mapping the degree and type of risk consumers perceive in personal data sharing', *Journal of Consumer Affairs*, vol. 51, no. 1, pp. 133-61.

Misra, G & Such, JM 2015, 'Social computing privacy and online relationships', in YJ Erden, R Giovagnoli & G Dodig-Crnkovic (eds), *AISB Social Aspects of Cognition and Computing Symposium*, University of Kent, pp. 1-5.

Moghaddam, HM, Acar, G, Burgess, B, Mathur, A, Huang, DY, Feamster, N, Felten, EW, Mittal, P & Narayanan, A 2019, 'Watching you watch: the tracking ecosystem of over-the-top TV streaming devices', in X Wang & J Katz (eds), *Conference on Computer and Communications Security*, ACM.

Mohammad, HJ 2019, 'Google or privacy, the inevitable trade-off', MSc thesis, Faculty of Humanities, Social Sciences and Education, The Arctic University of Norway.

Molm, LD, Takahashi, N & Peterson, G 2000, 'Risk and trust in social exchange: an experimental test of a classical proposition', *American Journal of Sociology*, vol. 105, no. 5, pp. 1396-427.

Moor, JH 1997, 'Towards a theory of privacy in the information age', *Computers and Society*, vol. 27, no. 3, pp. 27-32.

Mothersbaugh, DL, Foxx, WK, Beatty, SE & Wang, S 2011, 'Disclosure antecedents in an online service context', *Journal of Service Research*, vol. 15, no. 1, pp. 76-98.

Mourey, J & Waldman, AE 2020, 'Past the privacy paradox: the importance of privacy changes as a function of control and complexity', *Journal of the Association for Consumer Research*, https://doi.org/10.1086/708034.

Muhammad, SS, Dey, BL & Weerakkody, V 2017, 'Analysis of factors that influence customers' willingness to leave big data digital footprints on social media: a systematic review of literature', *Information Systems Frontiers*, vol. 20, no. 3, pp. 559-76.

Nehf, JP 2011, 'The FTC's proposed framework for privacy protection online: a move toward substantive controls or just more notice and choice?', *William Mitchell Law Review*, vol. 37, no. 4, pp. 1727-44.

Neisse, R, Steri, G, Geneiatakis, D & Nai Fovino, I 2016, 'A privacy enforcing framework for Android applications', *Computers & Security*, vol. 62, pp. 257-77.

Nicholas, J, Shilton, K, Schueller, SM, Gray, EL, Kwasny, MJ & Mohr, DC 2019, 'The role of data type and recipient in individuals' perspectives on sharing passively collected smartphone data for mental health: cross-sectional questionnaire study', *JMIR Mhealth Uhealth*, vol. 7, no. 4, Apr 5, pp. 1-10.

Nissenbaum, H 2004, 'Privacy as contextual integrity', *Washington Law Review*, vol. 79, pp. 119-58.

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

Nissenbaum, H & Patterson, H 2016, 'Biosensing in context: health privacy in a connected world', in D Nafus (ed), *Quantified: Biosensing Technologies in Everyday Life*, MIT Press, Cambridge, MA, pp. 79-100.

Norberg, PA, Horne, DR & Horne, DA 2007, 'The privacy paradox: personal information disclosure intentions versus behaviors', *Journal of Consumer Affairs*, vol. 41, no. 1, pp. 100-26.

Norouzizadeh Dezfouli, F, Dehghantanha, A, Eterovic-Soric, B & Choo, K-KR 2015, 'Investigating social networking applications on smartphones detecting Facebook, Twitter, LinkedIn and Google+ artefacts on Android and iOS platforms', *Australian Journal of Forensic Sciences*, vol. 48, no. 4, pp. 1-20.

Nouwens, M, Liccardi, I, Veale, M, Karger, D & Kagal, L 2020, 'Dark patterns after the GDPR: scraping consent pop-ups and demonstrating their influence', *Conference on Human Factors in Computing Systems*, ACM.

OAIC 2014, *Guide to undertaking privacy impact assessments*, Australian Government Office of the Australian Information Commissioner, viewed 22 March 2017, <https://www.oaic.gov.au/agencies-and-organisations/guides/guide-to-undertaking-privacy-impact-assessments>.

OECD 2013, *The OECD Privacy Framework*, Organisation for Economic Co-operation and Development, viewed 13 March 2017, <https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf>.

Olmstead, K & Smith, A 2017, *Americans and Cybersecurity*, Pew Research Center, viewed 15 February 2017, <http://www.pewinternet.org/2017/01/26/americans-and-cybersecurity/>.

Oppenheimer, DM, Meyvis, T & Davidenko, N 2009, 'Instructional manipulation checks: detecting satisficing to increase statistical power', *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 867-72.

Osbankk, P 2007, 'A privacy enhancing infrastructure for context awareness', PhD thesis, University of Kent at Canterbury, Kent, UK.

Page, X, Kobsa, A & Knijnenburg, BP 2012, 'Don't disturb my circles! Boundary preservation is at the center of location-sharing concerns', in J Breslin et al (eds), *International AAAI Conference on Weblogs and Social Media*, Association for the Advancement of Artificial Intelligence, pp. 266-73.

Paolacci, G, Chandler, J & Ipeirotis, PG 2010, 'Running experiments on Amazon Mechanical Turk', *Judgment and Decision Making*, vol. 5, no. 5, pp. 411-19.

Paolacci, G & Chandler, J 2014, 'Inside the Turk: understanding Mechanical Turk as a participant pool', *Current Directions in Psychological Science*, vol. 23, no. 3, pp. 184-88.

Popiel, P 2019, *Terms of public service: framing mobile privacy discourses*, First Monday, viewed 15 December 2019, <https://doi.org/10.5210/fm.v24i11.10005>.

Preibusch, S 2013, 'Guide to measuring privacy concern: review of survey and observational instruments', *International Journal of Human-Computer Studies*, vol. 71, no. 12, pp. 1133-43.

Rashidi, B, Fung, C & Vu, T 2015, 'Dude, ask the experts!: Android resource access permission recommendation with RecDroid', in R Boutaba & W Almuhtadi (eds), *IFIP/IEEE International Symposium on Integrated Network Management*, IEEE, pp. 296-304.

Raynes-Goldie, K 2010, *Aliases, creeping, and wall cleaning: understanding privacy in the age of Facebook*, viewed 4 June 2018, <https://firstmonday.org/ojs/index.php/fm/article/view/2775/2432>.

Raynes-Goldie, K 2012, 'Privacy in the age of Facebook: discourse, architecture, consequences', PhD thesis, Curtin University, Perth, Australia.

Razaghpanah, A, Nithyanand, R, Vallina-Rodriguez, N, Sundaresan, S, Allman, M, Kreibich, C & Gill, P 2018, 'Apps, trackers, privacy, and regulators: a global study of the mobile tracking ecosystem', in P Traynor & A Oprea (eds), *Network and Distributed System Security Symposium*, Internet Society, pp. 1-15.

Ren, J, Lindorfer, M, Dubois, DJ, Rao, A, Choffnes, D & Vallina-Rodriguez, N 2018, 'Bug fixes, improvements, ... and privacy leaks: a longitudinal study of PII leaks across Android app versions', in P Traynor & A Oprea (eds), *Network and Distributed System Security Symposium*, Internet Society, pp. 1-15.

Ren, J, Dubois, DJ, Choffnes, D, Mandalari, AM, Kolcun, R & Haddadi, H 2019, 'Information exposure from consumer IoT devices: a multidimensional, network-informed measurement approach', in P Gill & R Beverly (eds), *Internet Measurement Conference*, ACM, pp. 267-79.

Rohrer, F, Zhang, Y, Chitkushev, L & Zlateva, T 2013, 'DR BACA: dynamic role based access control for Android', in CN Payne (ed), *Annual Computer Security Applications Conference*, ACM, pp. 299-308.

Rosenberg, M 2018, *Facebook says Cambridge Analytica harvested data of up to 87 million users*, New York Times, viewed 30 Oct 2019, <https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html>.

Rotter, JB 1967, 'A new scale for the measurement of interpersonal trust', *Journal of Personality*, vol. 35, no. 4, pp. 651-65.

Schnorf, S, Sedley, A, Ortlieb, M & Woodruff, A 2014, 'A comparison of six sample providers regarding online privacy benchmarks', *Workshop on Privacy Personas and Segmentation*, USENIX.

Selleck, E 2017, *iOS 11 will better notify users when an app is using location in the background*, iPhoneHacks, viewed 1 October 2019, <http://www.iphonehacks.com/2017/06/ios-11-location-sharing-warning.html>.

Sensor Tower 2019a, *Most popular Google Play app store categories from 1st quarter 2016 to 4th quarter 2018, by number of downloads (in millions)*, Statista, viewed 10 March 2019, <https://www.statista.com/statistics/256772/most-popular-app-categories-in-the-google-play-store/>.

Sensor Tower 2019b, *Most popular Apple App Store categories from 1st quarter 2016 to 4th quarter 2018, by number of downloads (in millions)*, Statista, viewed 10 March

2019, <https://www.statista.com/statistics/237336/apple-app-store-top-app-categories-by-downloads/>.

Seo, J, Kim, D, Cho, D, Kim, T & Shin, I 2016, 'FLEXDROID: Enforcing in-app privilege separation in Android', in M Antonakakis (ed), *Network and Distributed System Security Symposium*, Internet Society, pp. 1-15.

Serovich, JM, Greene, K & Parrott, R 1992, 'Boundaries and AIDS testing: privacy and the family system', *Family Relations*, vol. 41, no. 1, pp. 104-09.

Serovich, JM & Greene, K 1993, 'Perceptions of family boundaries: the case of disclosure of HIV testing information', *Family Relations*, vol. 42, no. 2, pp. 193-97.

Shay, R, Ion, I, Reeder, RW & Consolvo, S 2014, '"My religious aunt asked why i was trying to sell her viagra": experiences with account hijacking', in T Grossman & A Schmidt (eds), *Conference on Human Factors in Computing Systems*, ACM, pp. 2657-66.

Sheehan, KB 2002, 'Toward a typology of internet users and online privacy concerns', *The Information Society*, vol. 18, no. 1, pp. 21-32.

Sheldon, P 2008, 'The relationship between unwillingness-to-communicate and students' Facebook use', *Journal of Media Psychology*, vol. 20, no. 2, pp. 67-75.

Shih, F 2015, 'ContextProbe: exploring mobile privacy in context', PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Shih, F, Liccardi, I & Weitzner, D 2015, 'Privacy tipping points in smartphones privacy preferences', in W Woo & K Inkpen (eds), *Conference on Human Factors in Computing Systems*, ACM, pp. 807-16.

Shipman, FM & Marshall, CC 2020, 'Ownership, privacy, and control in the wake of Cambridge Analytica: the relationship between attitudes and awareness', *Conference on Human Factors in Computing Systems*, ACM.

Shklovski, I, Mainwaring, SD, Skúladóttir, HH & Borgthorsson, H 2014, 'Leakiness and creepiness in app space: perceptions of privacy and mobile app use', in T Grossman & A Schmidt (eds), *Conference on Human Factors in Computing Systems*, ACM, pp. 2347-56.

Shvartzshnaider, Y, Tong, S, Wies, T, Kift, P, Nissenbaum, H, Subramanian, L & MIttal, P 2016, 'Learning privacy expectations by crowdsourcing contextual informational norms', in A Ghosh & M Lease (eds), *AAAI Conference on Human Computation and Crowdsourcing*, AAAI, pp. 209-18.

Smith, B 2018, *Project Strobe: Protecting your data, improving our third-party APIs, and sunsetting consumer Google+*, Google Blog, viewed 30 Oct 2019, <https://www.blog.google/technology/safety-security/project-strobe/>.

Smith, HJ, Dinev, T & Xu, H 2011, 'Information privacy research: an interdisciplinary review', *MIS Quarterly*, vol. 35, no. 4, pp. 989-1016.

Smith v. Maryland (1979) 442 Supreme Court of United States 735.

Snoopwall 2014, *Flashlight apps threat assessment report*. viewed 29 July 2016, <http://www.snoopwall.com/wp-content/uploads/2014/10/Flashlight-Spyware-Appendix-2014.pdf>.

Solove, DJ 2006, 'A taxonomy of privacy', *University of Pennsylvania Law Review*, vol. 154, no. 3, pp. 477-564.

Stevens, R, Gibler, C, Crussell, J, Erickson, J & Chen, H 2012, 'Investigating user privacy in Android ad libraries', in H Chen, L Koved & DS Wallach (eds), *Workshop on Mobile Security Technologies*, IEEE, pp. 1-10.

Stoller, DR 2019, *Facebook, Google fund nonprofits shaping federal privacy debate*, Bloomberg, viewed 19 Nov 2019, <https://news.bloomberglaw.com/privacy-and-data-security/facebook-google-donate-heavily-to-privacy-advocacy-groups>.

Svirsky, D 2019, *Why are privacy preferences inconsistent*, Harvard Law School, John M. Olin Center for Law, Economics, and Business.

Taddicken, M 2012, 'Privacy, surveillance, and self-disclosure in the social web: exploring the user's perspective via focus groups', in C Fuchs et al (eds), *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, Routledge, New York, pp. 255-72.

Taddicken, M 2014, 'The 'privacy paradox' in the social web: the impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure', *Journal of Computer-Mediated Communication*, vol. 19, no. 2, pp. 248-73.

Tamir, DI & Mitchell, JP 2012, 'Disclosing information about the self is intrinsically rewarding', *Proceedings of the National Academy of Sciences*, vol. 109, no. 21, pp. 8038-43.

Tavani, HT 2008, 'Informational privacy: concepts, theories, and controversies', in KE Himma & HT Tavani (eds), *The Handbook of Information and Computer Ethics*, Wiley, Hoboken, NJ, pp. 131-64.

Tene, O & Polonetsky, J 2014, 'A theory of creepy: technology, privacy, and shifting social norms', *Yale Journal of Law and Technology*, vol. 16, no. 1, pp. 59-102.

Terpstra, A, Schouten, AP, de Rooij, A & Leenes, R 2019, *Improving privacy choice through design: how designing for reflection could support privacy self-management*, First Monday, viewed 11 December 2019, <https://doi.org/10.5210/fm.v24i7.9358>.

Thompson, C, Johnson, M, Egelman, S, Wagner, D & King, J 2013, 'When it's better to ask forgiveness than get permission', in LF Cranor (ed), *Symposium on Usable Privacy and Security*, ACM, pp. 1-16.

Thompson, SA & Warzel, C 2019, *Twelve million phones, one dataset, zero privacy*, New York Times, viewed 21 Dec 2019, <https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>.

Toch, E 2012, 'Crowdsourcing privacy preferences in context-aware applications', *Personal and Ubiquitous Computing*, vol. 18, no. 1, pp. 129-41.

Trindade, RR 2020, 'Data privacy as a strategic competitive advantage: a case study of how Apple and Facebook use data privacy in their branding process', Master's thesis, Escola de Administração de Empresas de São Paulo, São Paulo, Brazil.

Tsai, L, Wijesekera, P, Reardon, J, Reyes, I, Egelman, S, Wagner, D, Good, N & Chen, J-W 2017, 'Turtle Guard: helping Android users apply contextual privacy preferences', in ME Zurko (ed), *Symposium on Usable Privacy and Security*, USENIX, pp. 145-62.

Utz, C, Degeling, M, Fahl, S, Schaub, F & Holz, T 2019, '(Un)informed Consent: studying GDPR consent notices in the field', in X Wang & J Katz (eds), *Conference on Computer and Communications Security*, ACM, pp. 1-18.

Vaidya, J & Atluri, V 2008, 'Privacy, profiling, targeted marketing, and data mining', in A Acquisti et al (eds), *Digital Privacy: Theory, Technologies, and Practices*, Auerbach Publications, Boca Raton, FL, pp. 117-31.

van Knippenberg, AFM 1984, 'Intergroup differences in group perceptions', *The Social Dimension: European Developments in Social Psychology*, vol. 2, Cambridge University Press, Cambridge, UK, pp. 560-78.

Vesset, D, Morries, HD, Little, G, Borovick, L, Feldman, S, Eastwood, M, Woo, B, Villars, RL, Bozman, JS, Olofson, CW, Conway, S & Yezhkova, N 2012, *Worldwide big data technology and services 2012-2015 forecast*, IDC. viewed 23 October 2017, <http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=6242>.

Vickery, JR 2014, "I don't have anything to hide, but …': the challenges and negotiations of social and mobile media privacy for non-dominant youth', *Information, Communication & Society*, vol. 18, no. 3, pp. 281-94.

Vitaldevara, K 2020, *Safer and more transparent access to user location*, Android Developers Blog, viewed 20 February 2020, <https://android-developers.googleblog.com/2020/02/safer-location-access.html>.

Voicebot & Voicify 2019, *Smart speaker consumer adoption report*, viewed 20 April 2019, <https://voicebot.ai/smart-speaker-consumer-adoption-report-2019/>.

von Staden, H 1996, '"In a pure and holy way": personal and professional conduct in the Hippocratic oath?', *Journal of the History of Medicine and Allied Sciences*, vol. 51, no. 4, pp. 404-37.

Votipka, D, Rabin, SM, Micinski, K, Gilray, T, Mazurek, MM & Foster, JS 2018, 'User comfort with Android background resource accesses in different contexts', in ME Zurko & HR Lipford (eds), *Symposium on Usable Privacy and Security*, USENIX, pp. 1-16.

Wacks, R 2010, *Privacy: a very short introduction*, Oxford University Press, Oxford, UK.

Waldman, AE 2018, 'What does trust mean for privacy?', *Privacy as Trust: Information Privacy for an Information Age*, Cambridge University Press, pp. 61-76.

Walker, KL 2016, 'Surrendering information through the looking glass: transparency, trust, and protection', *Journal of Public Policy & Marketing*, vol. 35, no. 1, pp. 144-58.

Wang, Y, Norcie, G, Komanduri, S, Acquisti, A, Leon, PG & Cranor, LF 2011, '"I regretted the minute I pressed share": a qualitative study of regrets on Facebook', in LF Cranor (ed), *Symposium on Usable Privacy and Security*, ACM, pp. 1-13.

Warberg, L, Acquisti, A & Sicker, D 2019, 'Can privacy nudges be tailored to individuals' decision making and personality traits?', in J Domingo-Ferrer (ed), *Workshop on Privacy in the Electronic Society*, ACM, pp. 175-97.

Warren, SD & Brandeis, LD 1890, 'The right to privacy', *Harvard Law Review*, vol. 4, no. 5, pp. 193-220.

Westin, AF 2003, 'Social and political dimensions of privacy', *Journal of Social Issues*, vol. 59, no. 2, pp. 431-53.

Wheeless, LR & Grotz, J 1976, 'Conceptualization and measurement of reported self-disclosure', *Human Communication Research*, vol. 2, no. 4, pp. 338-46.

Wheeless, LR & Grotz, J 1977, 'The measurement of trust and its relationship to self-disclosure', *Human Communication Research*, vol. 3, no. 3, pp. 250-57.

White House 2012, *Consumer data privacy in a networked world*. viewed 21 February 2017, <https://obamawhitehouse.archives.gov/sites/default/files/privacy-final.pdf>.

White House 2017, *Privacy in our digital lives protecting individuals and promoting innovation*. viewed 21 February 2017, <https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/images/Documents/Privacy%20in%20Our%20Digital%20Lives.pdf>.

Wiese, J, Kelley, PG, Cranor, LF, Dabbish, L, Hong, JI & Zimmerman, J 2011, 'Are you close with me? Are you nearby? Investigating social groups, closeness, and willingness to share', in DJ Patterson, Y Rogers & X Xie (eds), *International Conference on Ubiquitous Computing*, ACM, pp. 197-206.

Wijesekera, P, Baokar, A, Hosseini, A, Egelman, S, Wagner, D & Beznosov, K 2015, 'Android permissions remystified: a field study on contextual integrity', *USENIX Security Symposium*, USENIX, pp. 499-514.

Wijesekera, P, Baokar, A, Tsai, L, Reardon, J, Egelman, S, Wagner, D & Beznosov, K 2017, 'The feasibility of dynamically granted permissions: aligning mobile privacy with user preferences', in KR Butler, Ú Erlingsson & B Parno (eds), *Symposium on Security and Privacy*, IEEE, pp. 1-17.

World Economic Forum 2012, *Rethinking personal data: strengthening trust*, viewed 3 June 2018, <https://www.weforum.org/reports/rethinking-personal-data-strengthening-trust>.

Wu, PF, Vitak, J & Zimmer, M 2019, 'A contextual approach to information privacy research', *Journal of the Association for Information Science and Technology*, https://doi.org/10.1002/asi.24232.

Xie, W & Kang, C 2015, 'See you, see me: teenagers' self-disclosure and regret of posting on social network site', *Computers in Human Behavior*, vol. 52, pp. 398-407.

Yao, MZ, Rice, RE & Wallis, K 2007, 'Predicting user concerns about online privacy', *Journal of the American Society for Information Science and Technology*, vol. 58, no. 5, pp. 710-22.

Zhang, X, Gao, Q, Khoo, CS & Wu, A 2013, 'Categories of friends on social networking sites: an exploratory study', *International Conference on Asia-Pacific Library and Information Education and Practice*, University of Khon Kaen, pp. 244-59.

Zhauniarovich, Y, Russello, G, Conti, M, Crispo, B & Fernandes, E 2014, 'MOSES: supporting and enforcing security profiles on smartphones', *IEEE Transactions on Dependable and Secure Computing*, vol. 11, no. 3, pp. 211-23.

Zimmer, JC, Arsal, RE, Al-Marzouq, M & Grover, V 2010, 'Investigating online information disclosure: effects of information relevance, trust and risk', *Information & Management*, vol. 47, no. 2, pp. 115-23.

Zimmer, M 2005, 'Surveillance, privacy and the ethics of vehicle safety communication technologies', *Ethics and Information Technology*, vol. 7, no. 4, pp. 201-10.

Zimmer, M 2008, 'Privacy on planet Google: using the theory of "contextual integrity" to clarify the privacy threats of Google's quest for the perfect search engine', *Journal of Business & Technology Law*, vol. 3, no. 1, pp. 109-26.