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

Ming Di Leom[1] **Gaye Deegan**[1] **Ben Martini**[1] **John Boland**[2]
ming_di.leom@mymail.unisa.edu.au Gaye.Deegan, Ben.Martini, John.Boland@unisa.edu.au

[1]*University of South Australia, UniSA STEM, SA, 5000, Australia*
[2]*University of South Australia, Centre for Industrial and Applied Mathematics, SA, 5000, Australia*

Abstract

Guided by Nissenbaum's framework of contextual integrity, 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 1, we compiled 15 most common types of information from a sample (n = 390) of mobile users. In Study 2 (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.

1. Introduction

Fuelled by recent high-profile scandals (Yahoo [1], Cambridge Analytica [2], Equifax [3], Google+ [4]), users are growing mistrustful of digital service providers for their inattentive handling of user data and overarching reach on consumer's everyday life. Excessive tracking and data collection [5] on the web [6, 7], to mobile device [8-11], even to television [12-15], user data is often collected under flawed notice and consent (more commonly known as privacy policy) [16] and most likely *surrendered* by the users rather unwillingly [17]. Studies conducted on Privacy Enhancing Technologies (PETs) in mobile ecosystem estimated that as much as 10% of the permissions were granted reluctantly [18]; and at least 80% participants wished they could have denied the permission request, once they knew its purpose [19].

PET in the mobile platforms often relies on permissions management to restrict undesired information flow. However, the current approach in permissions management alone is not optimal as it often regards data 'privacy' as dichotomies—sensitive and non-sensitive, risky and non-risk, private (personal) and not-private, identifiable and non-identifiable—where only one half warrant privacy consideration. In the mobile platforms, users are usually prompted with consent dialogue or permission prompt whenever an app request for 'sensitive' data for the first time.

Classifying the sensitivity or riskiness of information leads to a troubling issue. Sensitive information is often predefined by the respective OS platform. However, what information constitute as *sensitive* is subject to the users' varying privacy preferences and may also vary according to circumstances. A study [20] found data sensitivity did not significantly affect the willingness to disclose. This suggests that relying on predefined sensitive information may be impractical in serving a broad user base. Sensitive information is often deemed so because it is identifiable, but this assumption could not apply as any piece of information is potentially an identifier or at least a quasi-identifier [21]. Piecing together related quasi-identifiers would paint a more comprehensive picture of an individual, resulting in an ensuing of privacy loss, regardless of the person's intent.

When a type of information is regarded as identifiable, it can become sensitive when disclosing it "may result in harm to its subjects" [22]. However, predicting which type of information can inflict harm is subjective and may not always consistent [23, 24]. Similarly, The OECD Privacy Framework [25] 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" [26]. Users often consider "a richer space of information" before disclosing a piece of information through a mobile device, instead of just taking into account of "sensitivity" [27].

Thus, defining privacy by sensitivity alone is problematic because sensitivity is usually at the discretion of the provider, who may not always act in the consumer's best interests [7, 10, 28]. There is also an inherent limitation in *computing* sensitivity as nuances of social interaction are often abstracted away [29], bounded by statistical models and computing resources. Even back in 1969, the measure of "sensitivity" is already recognised as being vary "…depends in large measure upon the context in which it was first given, and the context in which it is later used" [30]. Another contentious issue is that there is no universal definition of "privacy" [26, 31-33], let alone the definition of "sensitivity" (in the context of PET).

Contextual integrity [34] evaluates whether the flow of information is appropriate in a given context. 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*). It suggests that public outcry will erupt whenever there is a violation of an information norm. We can utilise this property to identify privacy violation that is dependent on the current social norm, without subscribing to a rigid definition of privacy. As such, we can construe CI as a "framework for socially regulating information flows that is legitimate separately from the contest over 'privacy'" [35].

The rest of this chapter is organised as follows. Section 2 reviews related works. Section 3 reports on Study 1. Section 4 contains Study 2. Section 5 discusses the results of the user studies. Section 6 concludes this chapter.

2. Literature Review

In our previous study, while there was evidence of demographical differences on trust, privacy concern and self-disclosure, we did not find any evidence to suggest demographic backgrounds significantly 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, in a way, are consistent with Martin and Nissenbaum [23] that show consumer across those categories (including those so-called 'unconcerned') could share a similar view on privacy expectations. In a series of studies conducted by [36], the results suggested individuals' privacy preferences are not necessarily relevant to the disclosure decision. This further demonstrates classifying consumer by privacy preference or concern is not effective.

The results also 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 [37, 38].

Existing studies have shown users often assess an information flow based on diverse contextual factors. A series of studies [39, 40] showed a significant influence of *purpose* on users' subjective judgement. This is also in line with Zimmer, et al. [41] 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 [27] 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 [20] showed participants considered not only the recipient but also the data type before disclosure. The result is also in line with Martin and Nissenbaum [23] 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.

Thus, in this study, 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 1)

Continuing from our previous research which showed the influence of recipient, in this paper, 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 [23, 42] which utilize generic data types, our study is more specific to mobile device usage where we derive data types from mobile users.

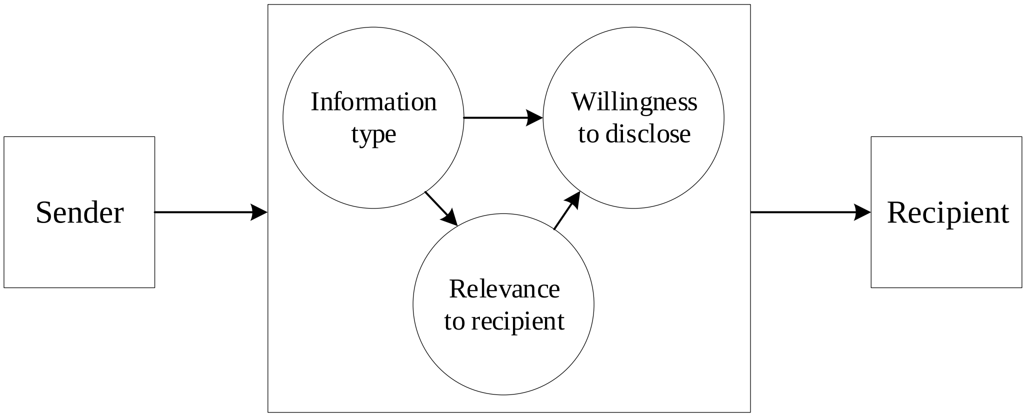

Figure 1: Influence of information relevance

3. Study 1

Methodology

We located existing studies [23, 42] 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. [43] and World Economic Forum [44], respectively. We initially considered to adapt the measures from those sources; however, we later found the derivation methods behind Madden, et al. [43] and World Economic Forum [44] 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 RQ1—to enumerate a list of information types commonly disclosed by mobile users—so that RQ3 and the rest 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 for questionnaire sample).

We advertised the survey on Mechanical Turk 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 workers 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 Mechanical Turks 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.

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 1 summarises participant demographics in Study 1.

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)

Table 1: Demographics of Study 1

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 2). It is similar to observed distributions for English word frequencies (i.e. Moby Dick ($\alpha = 1.95$) [45]). When counting the names, capitalisation and punctuation differences were ignored, but no stemming was performed.

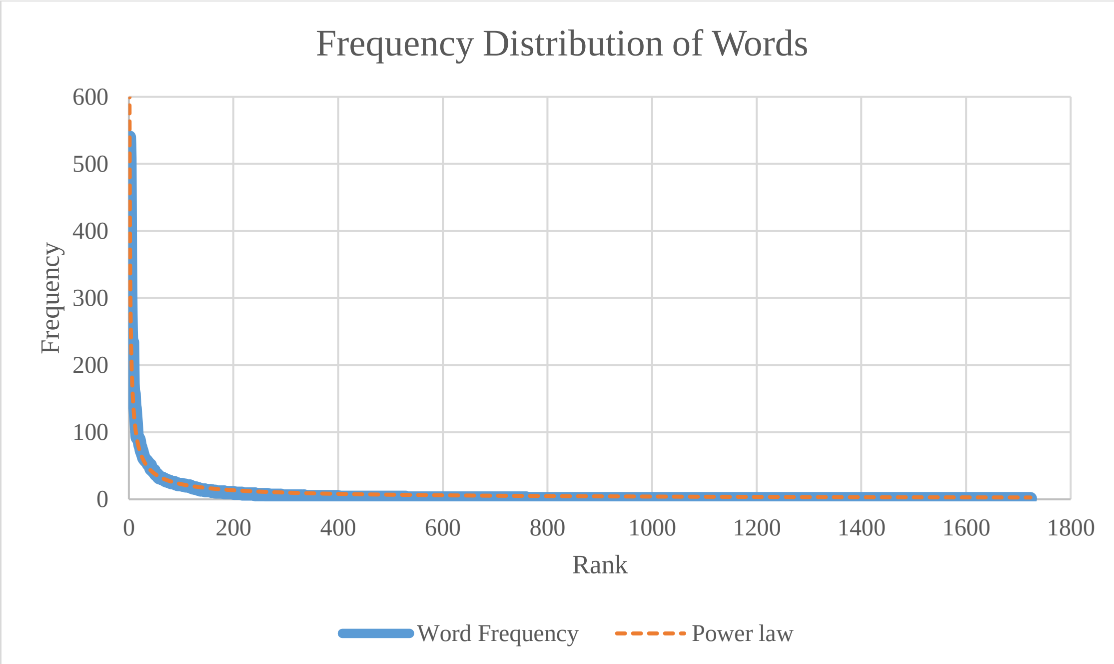

Figure 2: Power-law distribution (Study 1)

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 2 illustrates some examples. This combination resulted in 43 types where each type has a frequency of at least 10. Table 3 shows the 15 most popular types of information.

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 2: Compilation of 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

Table 3: 15 most popular types

4. Study 2

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, et al. [46]. We measure perceived relevance by using three 7-point scales adapted from Zimmer, et al. [41] (see Appendix for complete questionnaire). We assessed their reliability and deemed the constructs to have an acceptable level [47, Nunnally, cited in 48] of internal consistency, 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, each compiled from previous studies (including Study 1 of this paper) that we have conducted. 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, and we utilized TurkPrime (later rebranded as CloudResearch) to distribute surveys on MTurk. TurkPrime enabled us to exclude previous participants (Workers) from participating in subsequent studies.

Methodology

We advertised the questionnaires on Mechanical Turk 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. Table x shows the demographic. We utilized similar measures as Study 1'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.

We took several measures suggested previously [49-51] to minimise initial and remove subsequent junk data. These measures are:

1. The survey is only advertised to MTurk Workers 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 [52]. They are further inspected using measure 1-5 to verify they are invalid.
6. We identified responses from the same IP address and further verified using measure 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.

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 Likert to a range of -3 to +3. Table 4 shows the participants demographics.

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)

Table 4: Demographics of Study 2

Demographics

We compared the willingness to disclose among the demographics (Figure 3 and Table 5). 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.

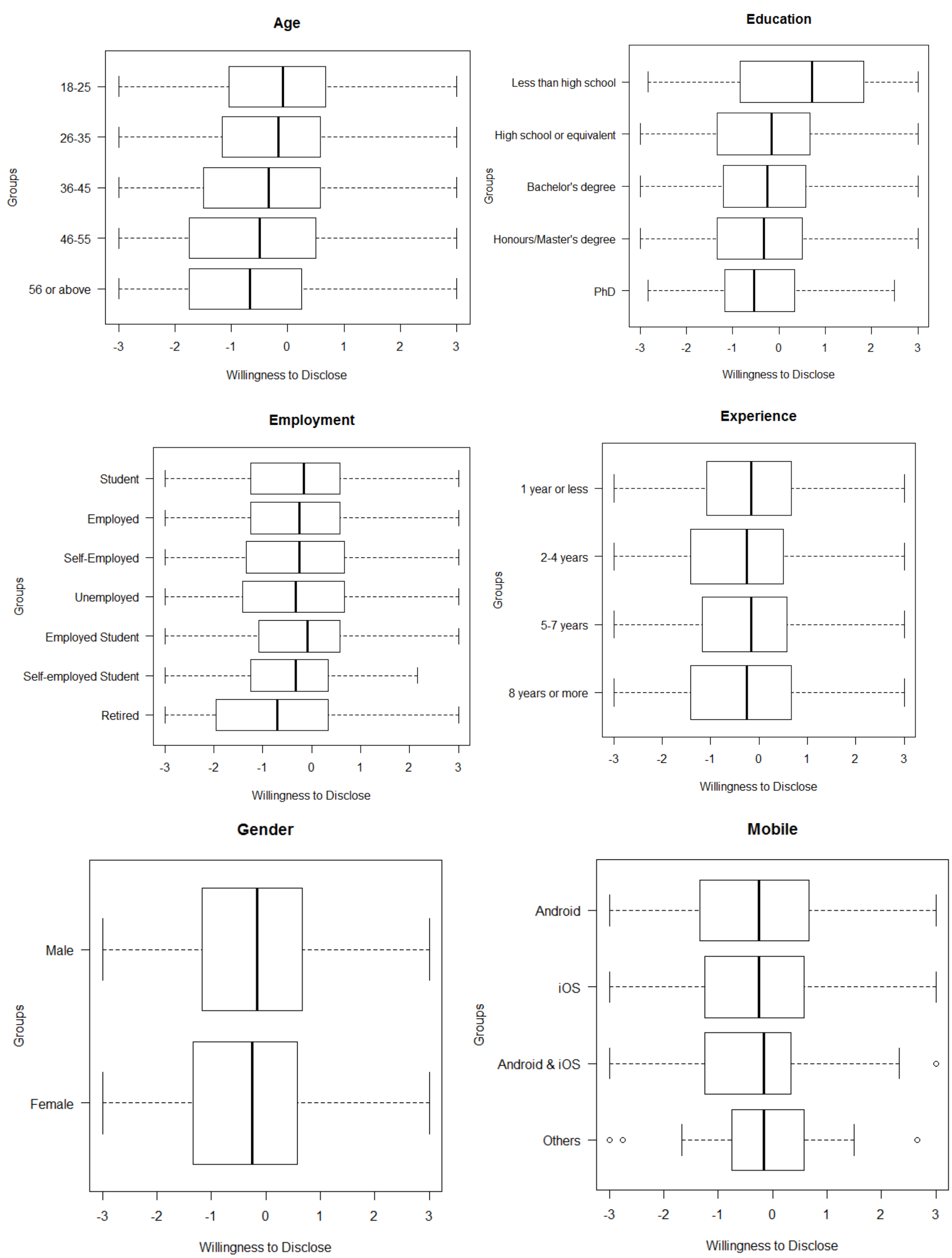

Figure 3: Willingness to disclose across demographics

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	
Other	0.66	0.66	

Table 5: Demographics differences in willingness to disclose

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

Table 6: Average indexes in different groups (each index column is colour-coded separately)

Disclosure Index

-1.2 0.8

Relevance Index

-0.5 0.7

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

Table 7: Average indexes in different types (each index column is colour-coded separately)

Disclosure Index

-0.8 0.2

Relevance Index

-0.15 0.5

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

Table 8: Average disclosure index

-1.7 1.2

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

Table 9: Average relevance index

-1 1.4

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

Table 10: Differences in disclosure and relevance indexes

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 11).

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

Table 11: Regression effect of relevance on willingness to disclose

5. Discussion

As part of our investigation on the relevance of the contextual integrity to the mobile ecosystem, especially the privacy aspect. In the previous study, we 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 paper, 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 [43] are not ranked in the lower half of the disclosure index (Table 7). Those types even rank higher in disclosure index than social media information, a type that is previously considered to be low sensitivity [53]. Previous studies posit that the paradox can be explained by information relevance [20, 41] which is a focus of this 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 as relevance 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 7). 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 10). 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 6 & Table 7).

Disclosure index may seem to be distinct between information types (Table 6). 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 8). 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 9). It is highest only in three out of nine recipient groups.

6. Conclusions

Findings from our studies in this paper 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 [54, 55] 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 8) and further demonstrate that sensitivity can vary according to the intended recipient. Second, while there is 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 of this study, 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, et al. [56]. 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.

References

[1] B. Lord. "An important message about Yahoo user security." https://yahoo.tumblr.com/post/150781911849/an-important-message-about-yahoo-user-security (accessed 20 Dec 2018.

[2] M. Rosenberg. "Facebook says Cambridge Analytica harvested data of up to 87 million users." New York Times. https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html (accessed 30 Oct 2019.

[3] FTC. "The Equifax data breach: what to do." Federal Trade Commission. https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-do (accessed 30 Oct 2019.

[4] B. Smith. "Project Strobe: Protecting your data, improving our third-party APIs, and sunsetting consumer Google+." Google Blog. https://www.blog.google/technology/safety-security/project-strobe/ (accessed 30 Oct 2019.

[5] B. Cyphers and G. Gebhart. "Behind the one-way mirror: a deep dive into the technology of corporate surveillance." Electronic Frontier Foundation. https://www.eff.org/wp/behind-the-one-way-mirror (accessed 15 Dec 2019.

[6] D. C. Howe and H. Nissenbaum, "Engineering privacy and protest: a case study of AdNauseam," *International Workshop on Privacy Engineering*, vol. 1873, pp. 1-8, 2017.

[7] A. Mathur *et al.*, "Dark patterns at scale: findings from a crawl of 11k shopping websites," presented at the Conference on Computer-Supported Cooperative Work and Social Computing, Austin, TX, 2019.

[8] A. Razaghpanah *et al.*, "Apps, trackers, privacy, and regulators: a global study of the mobile tracking ecosystem," presented at the Network and Distributed System Security Symposium, San Diego, CA, 2018.

[9] FTC. "Android flashlight app developer settles FTC charges it deceived consumers." Federal Trade Commission. https://www.ftc.gov/news-events/press-releases/2013/12/android-flashlight-app-developer-settles-ftc-charges-it-deceived (accessed 29 July 2016.

[10] Ø. H. Kaldestad and F. Myrstad, "Deceived by design," Norwegian Consumer Council, 2018. [Online]. Available: https://www.forbrukerradet.no/side/facebook-and-google-manipulate-users-into-sharing-personal-data/

[11] S. A. Thompson and C. Warzel. "Twelve million phones, one dataset, zero privacy." New York Times. https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html (accessed 21 Dec 2019.

[12] FTC. "FTC issues warning letters to app developers using 'Silverpush' code." Federal Trade Commission. https://www.ftc.gov/news-events/press-releases/2016/03/ftc-issues-warning-letters-app-developers-using-silverpush-code (accessed 18 March 2016.

[13] H. M. Moghaddam *et al.*, "Watching you watch: the tracking ecosystem of over-the-top TV streaming devices," presented at the Conference on Computer and Communications Security, London, UK, 2019.

[14] D. Y. Huang, N. Apthorpe, G. Acar, F. Li, and N. Feamster, "IoT inspector: crowdsourcing labeled network traffic from smart home devices at scale," *arXiv:1909.09848*, pp. 1-14, 2019.
[15] J. Ren, D. J. Dubois, D. Choffnes, A. M. Mandalari, R. Kolcun, and H. Haddadi, "Information exposure from consumer IoT devices: a multidimensional, network-informed measurement approach," presented at the Internet Measurement Conference, Amsterdam, Netherlands, 2019.
[16] A. Acquisti and J. Grossklags, "Privacy and rationality in individual decision making," *IEEE Security and Privacy Magazine*, vol. 3, no. 1, pp. 26-33, 2005, doi: 10.1109/msp.2005.22.
[17] K. L. Walker, "Surrendering information through the looking glass: transparency, trust, and protection," *Journal of Public Policy & Marketing*, vol. 35, no. 1, pp. 144-158, 2016, doi: 10.1509/jppm.15.020.
[18] B. Bonné, S. T. Peddinti, I. Bilogrevic, and N. Taft, "Exploring decision making with Android's runtime permission dialogs using in-context surveys," presented at the Symposium on Usable Privacy and Security, Santa Clara, CA, 2017.
[19] P. Wijesekera, A. Baokar, A. Hosseini, S. Egelman, D. Wagner, and K. Beznosov, "Android permissions remystified: a field study on contextual integrity," presented at the USENIX Security Symposium, Washington, D.C., 12-14 August 2015, 2015.
[20] J. Nicholas, K. Shilton, S. M. Schueller, E. L. Gray, M. J. Kwasny, and D. C. Mohr, "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, pp. 1-10, Apr 5 2019, doi: 10.2196/12578.
[21] K. Nissim and A. Wood, "Is privacy privacy?," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2128, 2018, doi: 10.1098/rsta.2017.0358.
[22] H. Nissenbaum, *Privacy in context: technology, policy, and the integrity of social life*. Stanford, CA: Stanford University Press, 2010.
[23] K. Martin and H. Nissenbaum, "Measuring privacy: an empirical test using context to expose confounding variables," *Columbia Science and Technology Law Review*, vol. 18, no. 1, pp. 176-218, 2016.
[24] M. Fazlioglu, "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, 2019.
[25] OECD. "The OECD Privacy Framework." Organisation for Economic Co-operation and Development. https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf (accessed 13 March 2017.
[26] R. Wacks, *Privacy: a very short introduction*. Oxford, UK: Oxford University Press, 2010.
[27] P. Wijesekera *et al.*, "The feasibility of dynamically granted permissions: aligning mobile privacy with user preferences," presented at the Symposium on Security and Privacy, San Jose, CA, 22-24 May 2017, 2017.
[28] A. M. Doorey, G. B. Wilcox, and M. S. Easin, "Consumer privacy and the new mobile commerce," *The Dark Side of Social Media: A Consumer Psychology Perspective*, A. C. Scheinbaum, Ed., New York, NY: Routledge, 2017, pp. 179-200.

[29] F. Ferra, I. Wagner, E. Boiten, L. Hadlington, I. Psychoula, and R. Snape, "Challenges in assessing privacy impact: Tales from the front lines," *Security and Privacy*https://doi.org/10.1002/spy2.101, 2019, pp. 1-19.
[30] A. R. Miller, "Personal privacy in the computer age: the challenge of a new technology in an information-oriented society," *Michigan Law Review*, vol. 67, no. 6, pp. 1089-1246, 1969, doi: 10.2307/1287516.
[31] T. Allmer, "A critical contribution to theoretical foundations of privacy studies," *Journal of Information, Communication and Ethics in Society*, vol. 9, no. 2, pp. 83-101, 2011, doi: 10.1108/14779961111148613.
[32] A. F. Westin, "Social and political dimensions of privacy," *Journal of Social Issues*, vol. 59, no. 2, pp. 431-453, 2003, doi: 10.1111/1540-4560.00072.
[33] H. T. Tavani, "Informational privacy: concepts, theories, and controversies," in *The Handbook of Information and Computer Ethics*, K. E. Himma and H. T. Tavani Eds. Hoboken, NJ: Wiley, 2008, ch. 6, pp. 131-164.
[34] H. Nissenbaum, "Privacy as contextual integrity," *Washington Law Review*, vol. 79, pp. 119-158, 2004.
[35] S. Benthall and B. Haynes, "Contexts are political: field theory and privacy," presented at the Symposium on Applications of Contextual Integrity, Berkeley, CA, 2019.
[36] K. Martin and K. Shilton, "Mobile privacy expectations: how privacy is respected with mobile devices," *The Cambridge Handbook of Consumer Privacy*, E. Selinger, J. Polonetsky, and O. Tene, Eds., Cambridge, UK: Cambridge University Press, 2018, pp. 85-101.
[37] A. Heravi, S. Mubarak, and K.-K. R. Choo, "Information privacy in online social networks: uses and gratification perspective," *Computers in Human Behavior*, vol. 84, pp. 441-459, 2018, doi: 10.1016/j.chb.2018.03.016.
[38] M. Taddicken, "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. 248-273, 2014, doi: 10.1111/jcc4.12052.
[39] J. Lin, N. Sadeh, S. Amini, J. Lindqvist, J. I. Hong, and J. Zhang, "Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing," presented at the Conference on Ubiquitous Computing, Pittsburgh, PA, 2012.
[40] J. Lin, B. Liu, N. Sadeh, and J. I. Hong, "Modeling users' mobile app privacy preferences: restoring usability in a sea of permission settings," presented at the Symposium On Usable Privacy and Security, Menlo Park, CA, 9-11 Jul 2014, 2014.
[41] J. C. Zimmer, R. E. Arsal, M. Al-Marzouq, and V. Grover, "Investigating online information disclosure: effects of information relevance, trust and risk," *Information & Management*, vol. 47, no. 2, pp. 115-123, 2010, doi: 10.1016/j.im.2009.12.003.
[42] V. Marmion, D. E. Millard, E. H. Gerding, and S. V. Stevenage, "The willingness of crowds: cohort disclosure preferences for personally identifying information," presented at the International AAAI Conference on Web and Social Media, Munich, Germany, 2019.
[43] M. Madden, L. Rainie, K. Zickuhr, M. Duggan, and A. Smith. "Public perceptions of privacy and security in the post-Snowden era." Pew Research Center.

https://www.pewresearch.org/internet/2014/11/12/public-privacy-perceptions/ (accessed 5 Nov 2019.
[44] World Economic Forum. "Rethinking personal data: strengthening trust." https://www.weforum.org/reports/rethinking-personal-data-strengthening-trust (accessed 3 June 2018.
[45] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661-703, 2009, doi: 10.1137/070710111.
[46] N. K. Malhotra, S. S. Kim, and J. Agarwal, "Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model," *Information Systems Research*, vol. 15, no. 4, pp. 336-355, 2004, doi: 10.1287/isre.1040.0032.
[47] P. Kline, *Handbook of psychological testing*, 2nd ed. New York: Routledge, 2000.
[48] T. Dinev and P. Hart, "Internet privacy concerns and their antecedents - measurement validity and a regression model," *Behaviour & Information Technology*, vol. 23, no. 6, pp. 413-422, 2004, doi: 10.1080/01449290410001715723.
[49] I. Kuziemko, M. I. Norton, E. Saez, and S. Stantcheva, "How elastic are preferences for redistribution? evidence from randomized survey experiments," *American Economic Review*, vol. 105, no. 4, pp. 1478-1508, 2015, doi: 10.1257/aer.20130360.
[50] R. Shay, I. Ion, R. W. Reeder, and S. Consolvo, ""My religious aunt asked why i was trying to sell her viagra": experiences with account hijacking," presented at the Conference on Human Factors in Computing Systems, Toronto, Canada, 2014.
[51] X. Page, A. Kobsa, and B. P. Knijnenburg, "Don't disturb my circles! Boundary preservation is at the center of location-sharing concerns," presented at the International AAAI Conference on Weblogs and Social Media, Dublin, Ireland, 2012.
[52] J. S. Downs, M. B. Holbrook, S. Sheng, and L. F. Cranor, "Are your participants gaming the system? Screening Mechanical Turk workers," presented at the Conference on Human Factors in Computing Systems, Georgia, GA, 2010.
[53] E. Markos, L. I. Labrecque, and G. R. Milne, "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, 2018, doi: 10.1016/j.intmar.2018.01.004.
[54] D. L. Mothersbaugh, W. K. Foxx, S. E. Beatty, and S. Wang, "Disclosure antecedents in an online service context," *Journal of Service Research*, vol. 15, no. 1, pp. 76-98, 2011, doi: 10.1177/1094670511424924.
[55] G. R. Milne, G. Pettinico, F. M. Hajjat, and E. Markos, "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. 133-161, 2017, doi: 10.1111/joca.12111.
[56] K. Li, Z. Lin, and X. Wang, "An empirical analysis of users' privacy disclosure behaviors on social network sites," *Information & Management*, vol. 52, no. 7, pp. 882-891, 2015, doi: 10.1016/j.im.2015.07.006.

Appendix

Study 1

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 2

Disclosure: Seven-point semantic scales [46]

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 [41]

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

1. Irrelevant / Relevant
2. Important / Unimportant (r)
3. Unnecessary / Necessary

(r): Reverse item