"I'm Not a Millionaire": How Users' Online Behaviours and Offline Behaviours Impact Their Privacy

Michael Lutaaya
michael@lutaaya.com
Carleton University
Ottawa, Ontario, Canada

Khadija Baig
khadijabaig@cmail.carleton.ca
Carleton University
Ottawa, Ontario, Canada

Sana Maqsood
sana.maqsood@carleton.ca
Carleton University
Ottawa, Ontario, Canada

Sonia Chiasson
chiasson@scs.carleton.ca
Carleton University
Ottawa, Ontario, Canada

ABSTRACT

Compromising the privacy of personally identifiable information (PII) can leave users vulnerable to risks, such as identity theft. We conducted a study with 27 participants in which we examined the types of publicly available PII they could locate on their social media accounts, and through a web search. We interviewed participants about the online and offline behaviours they employ to manage their PII. Participants leaked significant amounts of PII through their online presence, and potentially further exposed it through their offline behaviours. Many were surprised at the amount of PII they came across, and immediately took rectifying actions.

CCS CONCEPTS

• **Security and privacy → Human and societal aspects of security and privacy**;

KEYWORDS

usable security, privacy, user study, digital identity

ACM Reference Format:
Michael Lutaaya, Khadija Baig, Sana Maqsood, and Sonia Chiasson. 2021. "I'm Not a Millionaire": How Users' Online Behaviours and Offline Behaviours Impact Their Privacy. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3411763.3451603

1 INTRODUCTION

Users frequently disclose personal information about themselves online. The persistent nature of most digital disclosures means that this information can remain visible long after the user has forgotten about it. While some disclosures are harmless, others have potential to compromise privacy and lead to security vulnerabilities. Most notably, *Personally Identifiable Information (PII)* consists of any information that can be used to distinguish or trace an individual's identity, either alone or when combined with other information connected to the individual [10]. Compared to other types of user data (e.g.: browsing behaviours), disclosures of PII can have significant consequences: it can be used in fraudulent activities, such as identity theft [9, 27], and to compromise security applications which utilize PII (e.g.: password recovery) [14]. Given the serious consequences associated with PII disclosures, our work explores users' publicly available PII online which they are able to locate themselves, and their perceptions of the availability of this information.

Three questions guide our research: **(RQ1):** What PII is easily obtainable online? **(RQ2):** How visible are users' social media accounts to others? **(RQ3):** What online and offline protection strategies do users take to protect their PII? We interviewed 27 participants and, through a novel methodology, conducted a manual analysis of participants' social media accounts and the results of targeted web searches to quantify the degree to which their PII can be obtained. Most participants had PII visible to others on social media and appearing within web search results. Our interviews highlighted that many users prioritized convenience over protecting PII despite some awareness of the risks, and that others had inaccurate mental models of the risks, which led them to disregard protective actions. We also identify how users often ignore physical protection strategies despite their direct relationship to protecting PII.

2 RELATED WORK

Personally identifiable information (PII). Research has shown that publicly or semi-publicly available PII — such as birthdate, gender, location, and social media footprints — can be aggregated to identify an individual through a process known as re-identification [1, 4, 11, 24, 25], in which anonymised data is used to determine its owner. For example, by analysing United States Census data, Sweeney demonstrated that 87% of the population could be re-identified using only three attributes: their gender, birthdate, and postal code [24]. Similarly, Irani et al. found that social media footprints can be used to re-identify individuals, and that having a bigger online social footprint makes users more susceptible [14]. For example, users with one social media account leaked 34% of the PII attributes needed

CHI '21 Extended Abstracts, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8095-9/21/05…$15.00
https://doi.org/10.1145/3411763.3451603

for an identity attack, while those with six accounts leaked 90% of the required attributes [14]. Interviews with incarcerated identity thieves revealed that they employ various strategies, including physical ones, to collect PII [6]. Strategies include purchasing information from the black market, searching trash, and using social engineering to obtain it from friends, family and acquaintances [6]. In addition to identity theft, the PII disclosed by users online can be utilized to compromise various privacy and security processes, such as password recovery using security questions [12, 14].

Behavioural factors and users' privacy perceptions of PII. Previous research has explored the behavioural factors influencing users' PII disclosure practices online [2, 5, 15–18, 26]. Mesch [18] found that increased trust in an online entity or organization encourages users to disclose more PII. Venkatanathan et al. [26] studied the effect of reciprocity on PII (full name, date of birth) disclosure with strangers on social media, and found that users were more likely to do so when the experimental profile first shared this information with them. Reciprocity may be leveraged by attackers to extract PII from users. Based on regression analysis of survey data, users who perceive privacy risks associated with PII disclosure have been shown to refrain from disclosing PII [18]. However, little research has actually explored the prevalence of users who perceive risks associated with offline or online disclosures of PII. Munson et al. [19] surveyed users with publicly available information about political campaign contributions and real-estate transactions; while users were aware of providing their PII to these services, they were uncomfortable that it was publicly available. This highlights the disconnect between user expectations and current practices relating to PII.

Research Gap: Our work contributes by (i) exploring the privacy considerations of the physical management of PII (whereas the existing literature is primarily concentrated on online activities), (ii) providing an updated outlook on the prevalence and visibility of individuals' PII (since previous work on this [6, 14] was published at a time when social media was in its infancy [22] and smartphones were not yet widely adopted [21]), and (iii) providing a qualitative perspective on users' behaviours and attitudes (whereas previous work heavily focused on quantitative data).

3 METHODOLOGY

We conducted a lab study[1] to address our three research questions. Our methods were developed in collaboration with a local technology company and cleared by our University's Research Ethics Board (REB).

Participants and recruitment: 27 participants (12 male, 15 female) were recruited through posters placed in public locations across [anonymized city] and social media groups. They were between 18 and 35 years old (M = 23.5, SD = 4.8). 18 were students and 9 were employed in various industries. For mobile phone authentication, 22 used a PIN or password, 17 had enabled fingerprint authentication, and 16 had configured their phones to automatically lock after a specific time period. Their most frequently used social media platforms were Facebook (N = 22), Instagram (N = 22), and YouTube (N = 11). Participants were compensated with $25 and reimbursed for parking expenses.

Procedure: Prior to the session, participants viewed the informed consent online and named their three most used social media accounts. They were asked to bring their login credentials to the session or a personal device which was already logged into their accounts. During the 90-minute in-person session, participants completed a demographics questionnaire, a semi-structured interview, and two hands-on tasks meant to find their publicly accessible PII online. During the tasks, we noted (i) the presence or absence of the PII, and (ii) relevant contextual information. The PII itself was never recorded.

Semi-structured interview: We asked participants about how they keep physical personal belongings containing PII and electronic devices safe, their experiences with data breaches, and attitudes towards technologies like mobile wallets and digital identities.

Social media task: For each social media account, participants looked up their general account information (e.g.: friend count, privacy settings), searched for specific types of PII as directed by the researcher (e.g.: name, phone number), and recorded the presence (yes/no) of these PII. These activities were completed while participants were signed into their accounts, so we could note any relevant security and privacy information (e.g.: privacy restrictions) associated with their accounts.

Web-search task: Participants entered provided search queries in Google to find their PII online, opening relevant results to verify whether they contained any of their PII.

The types of PII included (Appendix A) were based on what government agencies and firms offering identity proofing services would be capable of verifying [20]. As such, the information could, in theory, be used for identity theft.

Data Analysis: We analyzed interview data using qualitative content analysis [8]. This analysis was conducted by the primary researcher and a research assistant, who met frequently to compare coding decisions. After eight rounds of coding, percent agreement was 99%. Any remaining discrepancies were resolved through discussion. In reporting interview results, we deliberately avoided quantitative measures (e.g.: prevalence of a theme), because the literature [7] warns that there is "no simple relationship" between quantitative measures and the significance of an excerpt.

We classified our results according to the six terms defined in Table 1.

4 RESULTS

4.1 Addressing RQ1: Discoverability of PII

Figure 1 tabulates the quantitative results from the social media and web-search tasks, which identified the types of PII readily visible to others based on social media disclosures and simple web searches. We organize the types of information into six categories, as indicated at the bottom of Figure 1.

PII Discoverability through Social Media.

[1] Data collection occurred prior to the COVID-19 pandemic

Classification	Definition
Identified	The attribute was located, was correct, and was clearly an attribute of the participant's identity (e.g.: the participant's first name was found).
Deducible	The attribute was located but was incomplete or missing an explicit indication that it was a definite attribute of the participant's identity (e.g.: inferring an individual's birth year from birthday greetings seen on their profile).
Outdated	The attribute was located and was correct at some point in the past but is no longer true or valid (e.g.: email address that is no longer valid).
Fake Data	The attribute was located but was never true or valid at any point (e.g.: incorrect birth date).
Not Identified	The attribute was not located.
Not Applicable	The attribute did not apply to the participant (e.g.: participants without middle names).

Table 1: Terms used to classify results of social media and search result exercises.

(1) **Identifiable:** the most identifiable attributes were *Personal Information*, of which *first names*, *gender*, and *photos* were the most easily identifiable,
(2) **Deducible:** the easiest category to deduce was *Birth Information*, with *place of birth* being the most common.
(3) **Not identifiable:** *Vehicle Information* was the hardest to identify, only being identified once through a picture of a participant's car on Facebook.
(4) **Outdated:** although relatively infrequent, *Contact Information* was most frequently *outdated*.
(5) **Faked:** similarly infrequent, *day of birth*, *month of birth*, *marital status*, and *last name* were each *faked* once.

PII Discoverability through Web Searches.

(1) **Identifiable:** similar to the social media exercise, the most identifiable attributes were *Personal Information*, of which *first names*, *gender*, and *photos* were the most identified attributes.
(2) **Deducible:** again, the most deducible category was *Birth Information*, with *year of birth* and *place of birth* being the most easily deduced.
(3) **Not identifiable:** all *Vehicle Information* was classified as *not identifiable* (or *not applicable*). *Birth Information* was also rarely identified.
(4) **Outdated:** occurred rarely, with only a few instances across all categories.
(5) **Faked:** *Birth Information* was the only category with a faked attribute: *year of birth*, which was only faked once.

Participants rarely identified fake or outdated data visible to others, suggesting that they value accuracy in their self-portrayal when they do choose to disclose PII.

4.2 Addressing RQ2: Visibility of social media accounts

Facebook and Instagram were the most used Social Networking Sites (SNS), together representing 83% of the accounts analysed (N = 19 and N=15 from N = 41, respectively). On average, **Facebook** accounts had 692 friends (SD = 665, Mdn = 449)[2]. **Instagram** accounts averaged 330 followers (SD = 222, Mdn = 302) while accounts with **other SNS** averaged only 100 followers (SD = 113, Mdn = 78).

To address RQ2, we explore for each social media platform: the visibility of account content, the ease of looking up the account via the associated email address, and the ease of looking up the account via the associated phone number.

4.2.1 *Content visibility.*

Facebook: 68% of Facebook accounts (13 out of 19) had configured their future posts to be *publicly viewable*; the remainder had an assortment of more *restrictive* visibility settings.
Instagram: 57% of Instagram accounts (8 out of 14) restricted content to their *followers only*, with the remaining accounts *public*.
Other SNS: 71% of these accounts (5 out of 7) had their content viewable to *Public*; only two accounts restricted their content to *Other Registered Users* of the service.

4.2.2 *Email address lookup.*

Facebook: 57% of accounts (11 out of 19) accounts had the *most restrictive* option (*Friends* only) enabled, versus the remaining 8 accounts allowing *All Users* to do so.
Instagram: Users cannot restrict who can look them up by email and email addresses are mandatory at registration; thus it would appear that all Instagram users can be found by another user who has their email address.
Other SNS: Of the remaining four accounts, two had restricted lookup to *Nobody*, one had restricted it to *Friends of Friends* while another had allowed *All Users*.

4.2.3 *Phone number lookup.*

Facebook: 52% of Facebook accounts (10 out of 19) allowed *Everyone* while the remaining nine accounts had restricted this capability to *Friends*. Notably, upon seeing that their

[2]Facebook's friend count sometimes varied from screen to screen [28], possibly due to infrastructure reasons [29]. The results presented are based on the lowest friend count we observed for a participant.

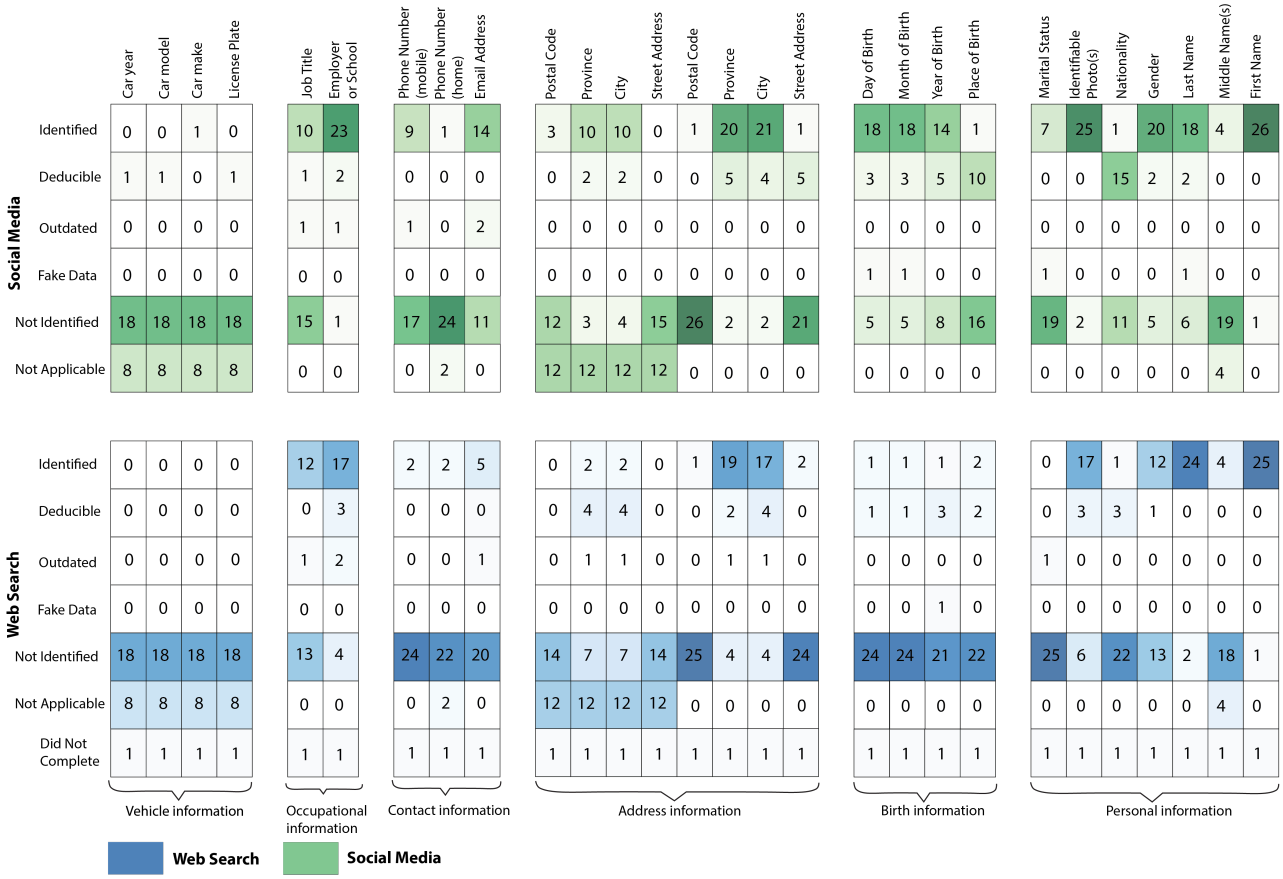

Figure 1: Discoverability of identity attributes on participants' social media and within search results for their PII.

account was configured to allow *Everyone* to look them up by their phone number, P8 immediately restricted this setting.
Instagram: As before, Instagram users cannot restrict who can look them up by their phone number and, as such, all users could be looked up by the phone number associated with their account if provided.
Other SNS: Of the four remaining accounts, three allowed *Nobody* and one allowed *Friends of Friends* to look them up using their phone number.

4.3 Addressing RQ3: Protective Strategies

To address RQ3, we discuss participants' attitudes and behaviours towards protection of their physical documents, electronic devices, and digital information, and perceptions of these behaviours.

4.3.1 *Physical possessions.*

(1) **Storage of possessions:** Participants often made efforts to keep their personal belongings containing PII (identity documents, wallets) close to them or in locations with restricted access (such as locked storage). Participants also used what they felt were more obscure locations at home, such as suitcases and drawers for clothing, although we observed common patterns. Conversely, many also reported leaving various items (e.g.: wallets, purses) unattended in a variety of contexts (e.g.: in vehicles, in coffee shops while using the washroom). Some participants defended these actions by mentioning their intention to return quickly or their trust of the people in the environment. P6 mentioned particularly risky tendencies, explaining that he would typically "*leave [his entire] wallet in [his] car*", only retrieving it when he needed money.

(2) **Factors affecting storage:** Participants' personal lives influenced how they stored PII-containing documents (if at all), but privacy was rarely mentioned in these decisions. Participants weighed whether they anticipated needing the document and how challenging it would be to replace it. P10 did not carry his birth certificate because replacing it would entail *"[going] back to [his] country"*, while P1 pointed to the complex application process for passports as a reason for keeping his at home. Many also reported carrying their Social Insurance Number (SIN) card[3] on their person or in their wallet. P28 explained that her living situation prevented her from storing these documents safely, while P25 explained she "*[didn't] have [her number] memorized.*"

[3]Similar to the US's Social Security Number, Canada's Social Insurance Number is a unique identifier used for government purposes.

4.3.2 *Electronic devices.*

(1) **Authentication:** Participants relied on a variety of authentication schemes (e.g.: passwords, biometrics, and pattern locks) to prevent their devices from being accessed by unintended users. Best practices were not always followed; for example, P23 uses a single-character password: *I would say [my password is] really easy [to guess] but [an attacker] would not think about it. . . . [My password is] literally [redacted character].* Others chose to forego authentication altogether because it was either a *"wastage of time"* (P10), or inaccessible (for example: P28 was *"on a lot of medication and sometimes [mistyped] things"*), or participants felt that had they *"nothing to hide"* (P20).

(2) **Location:** Another common protective strategy participants used was to keep their devices nearby. This was most often seen with smartphones: *"my smartphone is on me 24/7"* (P4), *"I keep my mobile phone all the time in my pocket"* (P10). Larger devices, however, were more likely to be left behind, as P1 explains: *"I work freelance. So I work from cafés. And, like, classrooms and things. So I do leave [my belongings] unattended. But only because it'd be impossible for somebody to get by me. . . . I'd be able to see them. There's one entrance in and out of the room so they can't really take [my belongings] without me seeing"* (P1).

(3) **Undesirable technology:** One participant offered this unique strategy of protecting his device: *Don't nobody want access to my Samsung S3. This thing's old. . . . I kinda sorta purposefully keep this phone so that way it deters people from wanting to steal my possession* (P13).

4.3.3 *Digital information.*

(1) **Authentication:** To protect what they considered more personal information, some participants would sign out of their applications and online accounts when not using them, or request device-based authentication with every use (like P11, who would use biometrics to *"reload the [Starbucks] card"*). P22 explains: *"I have to sign in pretty much every time for most of the things I care about—at least my emails and anything that has more personalised information."*. A few also mentioned using multi-factor authentication to protect what was especially valuable: *"My banking information is then again protected by a different passcode [than my phone's passcode]—a stronger one—and a fingerprint"* (P24). Other participants saw no need for such safeguards, with several not requiring authentication to access their device's applications. For example, P28 mentioned that because she does not have a password on her phone, *"[someone] could buy things off [her] Amazon and [her] Sephora."*

(2) **Storage of digital information:** Multiple participants avoided storing information that they deemed valuable on their devices, such as payment details for their debit or credit cards: *"I don't keep . . . my SIN number [on my smartphone]. I don't keep credit card information, debit, whatever. . . you're just gonna find a bunch of memes."* (P20). Many participants, though, made use of their web browsers' autofill functionality to pre-fill credit card information, account credentials for various online services (e.g.: financial institutions, email providers, social media services), and other commonly used form-fields requesting PII. Participants seemed aware of the risks but found the convenience more important: *"on my laptop, . . . passwords are auto-saved so if they were to go on to the online banking, they could sign in because the password is auto-saved"* (P27), *"I let Google save my credit card information so really [an attacker] could go to anything that allows you to have Google auto-save from Chrome to put in my credit card number"* (P24).

4.3.4 *Perception of risk.*

We found great variation in how participants evaluate the level of risk associated with their behaviours. Some participants acknowledged the possibility of being targeted by an attacker. For example, P8 felt that an experienced hacker could unlock her phone but that a *"regular, run-of-the-mill person"* would find it difficult to bypass the lockout mechanism that activates after multiple incorrect PIN entries. Meanwhile, despite being a mobile wallet user, P3 doubted its security: *"It's sensitive information [the information stored in a wallet app] and it's never one hundred percent safe. . . . And there's always the potential that it could . . . fall into the wrong hands."*

Other participants were unconcerned. For instance, P1 said, *"I don't view myself as that much of a target. . . . I'm not a millionaire, right?"* P12 echoed this sentiment, explaining that being *"the poorest of the people who works [at her place of employment]"* would mean that co-workers would not be interested in her wallet and so she does not need to lock it away. She voiced no related privacy concerns. Similarly, P14 believed that attackers would go after celebrities and politicians instead of ordinary people.

5 DISCUSSION

Research Questions. Returning to our research questions, we find that participants were generally truthful when revealing PII online. Many participants were unaware of how much PII they have available online, of the extent to which PII was visible to others, and of the implications of their online and offline behaviours. As seen in previous literature [13], the security and privacy of PII was frequently an afterthought, with participants deliberately choosing convenience over security and privacy. The results of our study suggest that a determined attacker could obtain a considerable amount of an individual's PII by conducting simple searches to locate their information online, by having an established relationship with the individual, or by taking advantage of opportunities of physical proximity (e.g.: having access to someone's belongings in a coffee shop). This availability of PII facilitates identity attacks relating to impersonation or breaching authentication protocols relying on PII, such as secret questions.

While reviewing their social media accounts, several participants were surprised to find their privacy settings to be relatively permissive. Some immediately restricted these settings (e.g.: by making the information visible to fewer people or revoking app permissions). They were unaware of the information they have disclosed or the extent to which it is available to others. We argue that under these circumstances, individuals are not truly in control of their own PII.

Design considerations. While educating users of the risks and how to protect themselves seems like a viable solution, it often has

limited impact [3]. Defending against identity threats should not be the sole responsibility of the users, and we must be careful not to simply shift blame onto potential victims [13]. System designers may need to take on a greater role. Their influence over the structure of a system may be leveraged to restrict user error and promote privacy and security conscious dynamics. For instance, the Government of Canada no longer issues wallet-sized SIN cards but instead provides the numbers on letter-sized paper to discourage individuals from using their SIN as an identity document and to reduce opportunities for theft [23]. Although, it remains to be seen whether this instead encourages users to take a digital photo of their document and shift the risk from physical theft to online capture. Recognizing that competing forces are at play, we nonetheless echo the need for altering the default settings of social media platforms to be privacy-preserving as opposed to public to help prevent leakage of PII.

Study design. We also highlight our study design as a novel way to collect data relating to PII in a consistent manner. Our methods enabled participants to feel comfortable with the process and remain in control of their data.

Limitations. Our study was conducted with a relatively small number of participants, which may not be representative of the general population. While running this study with more participants would be valuable, we anticipated challenges in doing so (e.g.: scaling up the manual reviews of social media and web searches). A follow-up study could explore the use of crawlers although this contravenes the Terms of Service for many platforms. For replicability, our methods explored specific PII using a consistent search methodology across participants and, as such, offers a lower bound on what information is available online about individuals.

6 CONCLUSION

By examining the availability of PII online through social media and search engines, we found a wide range of participants' PII readily visible online, sometimes without their knowledge. We also explored how participants managed their PII through online and offline behaviours. Many unknowingly exhibited behaviours and attitudes that jeopardise their personal security and privacy. Limiting user error by adjusting system design may help protect their privacy.

REFERENCES

[1] Alessandro Acquisti and Ralph Gross. 2009. I Just Found 10 Million SSNs. Paper presented at Black Hat USA 2009, Las Vegas, NV. https://www.blackhat.com/presentations/bh-usa-09/ACQUISTI/BHUSA09-Acquisti-GrossSSN-PAPER.pdf.

[2] Mahdi Nasrullah Al-Ameen, Tanjina Tamanna, Swapnil Nandy, MA Manazir Ahsan, Priyank Chandra, and Syed Ishtiaque Ahmed. 2020. We Don't Give a Second Thought Before Providing Our Information: Understanding Users' Perceptions of Information Collection by Apps in Urban Bangladesh. In *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies*. ACM, New York, NY, USA, 32–43.

[3] Maria Bada, Angela Sasse, and Jason R. C. Nurse. 2015. Cyber Security Awareness Campaigns: Why do they fail to change behaviour?. In *International Conference on Cyber Security for Sustainable Society*. N/A, Coventry, UK, 118–131. https://arxiv.org/abs/1901.02672.

[4] Michael Barbaro and Tom Zeller Jr. 2006. A Face Is Exposed for AOL Searcher No. 4417749. *The New York Times* N/A, N/A (2006), N/A. Retrieved June 24, 2019 from https://www.nytimes.com/2006/08/09/technology/09aol.html.

[5] Juan Pablo Carrascal, Christopher Riederer, Vijay Erramilli, Mauro Cherubini, and Rodrigo de Oliveira. 2013. Your browsing behavior for a big mac: Economics of personal information online. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, New York, NY, US, 189–200.

[6] Heith Copes and Lynne Vieraitis. 2007. *Identity Theft: Assessing Offenders' Strategies and Perceptions of Risk*. Report. Inter-university Consortium for Political and Social Research. Retrieved December 12, 2017 from https://www.ncjrs.gov/pdffiles1/nij/grants/219122.pdf.

[7] Hilde Van den Bulck. 2002. Tools for Studying the Media. In *The Media Book*, Chris Newbold, Oliver Boyd-Barrett, and Hilde Van den Bulck (Eds.). Arnold, N/A, Chapter 2, 55–100.

[8] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *Journal of Advanced Nursing* 62, 1 (2008), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x https://doi.org/10.1111/j.1365-2648.2007.04569.x.

[9] Kristin M Finklea. 2009. *Identity theft: Trends and issues*. DIANE Publishing, Darby, PA, U.S.A.

[10] JOINT TASK FORCE and TRANSFORMATION INITIATIVE. 2010. Guide for applying the risk management framework to federal information systems. *NIST special publication* 800 (2010), 37.

[11] Philippe Golle. 2006. Revisiting the Uniqueness of Simple Demographics in the US Population. In *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society (WPES '06)*. ACM, New York, NY, US, 77–80. https://doi.org/10.1145/1179601.1179615.

[12] Virgil Griffith and Markus Jakobsson. 2005. Messin'with texas deriving mother's maiden names using public records. In *International Conference on Applied Cryptography and Network Security*. Springer, Springer Berlin Heidelberg, Berling, Heidelberg, 91–103.

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

[14] Danesh Irani, Steve Webb, Kang Li, and Calton Pu. 2011. Modeling Unintended Personal-Information Leakage from Multiple Online Social Networks. *IEEE Internet Computing* 15, 3 (2011), 13–19. https://doi.org/10.1109/MIC.2011.25.

[15] Yongick Jeong and Erin Coyle. 2014. What are you worrying about on facebook and twitter? an empirical investigation of young social network site users' privacy perceptions and behaviors. *Journal of Interactive Advertising* 14, 2 (2014), 51–59.

[16] Hanna Krasnova, Nicole Eling, Oleg Schneider, Helena Wenninger, Thomas Widjaja, and Peter Buxmann. 2013. Does this app ask for too much data? The role of privacy perceptions in user behavior towards Facebook applications and permission dialogs. In *European Conference on Information Systems (ECIS)*. Association for Information Systems, Atlanta, GA, USA.

[17] Philipp K Masur and Michael Scharkow. 2016. Disclosure management on social network sites: Individual privacy perceptions and user-directed privacy strategies. *Social Media+ Society* 2, 1 (2016), 2056305116634368.

[18] Gustavo S Mesch. 2012. Is online trust and trust in social institutions associated with online disclosure of identifiable information online? *Computers in Human Behavior* 28, 4 (2012), 1471–1477.

[19] Sean A. Munson, Daniel Avrahami, Sunny Consolvo, James Fogarty, Batya Friedman, and Ian Smith. 2011. Attitudes Toward Online Availability of US Public Records. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*. ACM, New York, NY, USA, 2–9. https://doi.org/10.1145/2037556.2037558.

[20] National Institute of Standards and Technology. 2015. *Measuring Strength of Identity Proofing*. Discussion draft. Inforation Technology Laboratory, NIST. Retrieved June 25, 2019 from https://www.nist.gov/sites/default/files/nstic-strength-identity-proofing-discussion-draft.pdf.

[21] Pew Research Center. 2019. Mobile Fact Sheet. Retrieved January 20, 2020 from https://www.pewresearch.org/internet/fact-sheet/mobile/.

[22] Pew Research Center. 2019. Social Media Fact Sheet. Retrieved September 7, 2019 from https://www.pewinternet.org/fact-sheet/social-media/.

[23] Standing Committee on Finance. 2012. *Evidence*. Transcript 60. House of Commons of Canada. Retrieved May 24, 2019 from https://www.ourcommons.ca/DocumentViewer/en/41-1/FINA/meeting-60/evidence.

[24] Latanya Sweeney. 2000. Simple demographics often identify people uniquely. *Health (San Francisco)* 671, 2000 (2000), 1–34.

[25] Latanya Sweeney. 2013. Matching Known Patients to Health Records in Washington State Data. *Journal of Technology Science* N/A, N/A (2013), N/A. Retrieved September 7, 2019 from https://techscience.org/a/2015092903.

[26] Jayant Venkatanathan, Vassilis Kostakos, Evangelos Karapanos, and Jorge Gonçalves. 2014. Online disclosure of personally identifiable information with strangers: Effects of public and private sharing. *Interacting with Computers* 26, 6 (2014), 614–626.

[27] Linda Wilbanks. 2007. The impact of personally identifiable information. *IT Professional* 9, 4 (2007), 62–64.

[28] Alexis Wilke. 2017. Getting Rid of Deactivated Facebook Friends | Made to Order Software Corporation. Retrieved April 30, 2019 from https://www.m2osw.com/delete-deactivated-facebook-friends.

[29] Josh Wolford. 2012. Facebook Including Deactivated Accounts In Total Friend Count [UPDATED]. Retrieved April 30, 2019 from https://www.webpronews.com/facebook-including-deactivated-accounts-in-total-friend-count/.

Appendices

A SEARCH TERMS

- First and last names (in quotes); full name (in quotes) (e.g., `"jane doe" OR "jane emily doe"`)
- Name; current city; city resided in the longest (e.g., `jane doe paris OR new york`)
- Name; current or most recent employer (e.g., `jane doe air canada`)
- Name; current or most recent school (e.g., `jane doe acme university`)
- Primary email address (in quotes), typical username (in quotes) (e.g., `"jane.doe@gmail.com" OR "janedoe123"`)
- Your phone number; your longest-held phone number (e.g., `2125551234 OR 2125557890`)