“My Bank Already Gets this Data”: Exposure Minimisation and Company Relationships in Privacy Decision-Making

Reuben Binns
Jun Zhao
Max Van Kleek
Nigel Shadbolt
University of Oxford
Oxford, United Kingdom
reuben.binns@cs.ox.ac.uk
jun.zhao@cs.ox.ac.uk
max.van.kleek@cs.ox.ac.uk
nigel.shadbolt@cs.ox.ac.uk

Ilaria Liccardi
Daniel Weitzner
MIT CSAIL
Cambridge, MA, USA
ilaria@csail.mit.edu
djweitzner@csail.mit.edu

CHI'17 Extended Abstracts, May 6–11, 2017, Denver, CO, USA.
ACM ISBN 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3053255

Abstract
This paper explores how individuals' privacy-related decision-making processes may be influenced by their pre-existing relationships to companies in a wider social and economic context. Through an online role-playing exercise, we explore attitudes to a range of services including home automation, Internet-of-Things and financial services. We find that individuals do not only consider the privacy-related attributes of applications, devices or services in the abstract. Rather, their decisions are heavily influenced by their pre-existing perceptions of, and relationships with, the companies behind such apps, devices and services. In particular, perceptions about a company's size, level of regulatory scrutiny, relationships with third parties, and pre-existing data exposure lead some users to choose an option which might otherwise appear worse from a privacy perspective. This finding suggests a need for tools that support users to incorporate these existing perceptions and relationships into their privacy-related decision making.

Author Keywords
Decision-making; Personal data; Mobile apps; Privacy Indicators

ACM Classification Keywords
H.5.m. [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous; K.4.1 [Computers and Society]: Privacy

Introduction

With the proliferation of new devices and applications in a wide range of contexts, from banking, media and the home, individuals face an increasing number of decisions which are likely to impact on their digital privacy. Personal devices and applications generally require the collection of data about user behaviour in order to function. Once collected, such data is usually not stored locally, but sent to remote servers, belonging to the service provider and potentially other third parties for purposes such as personalisation, advertising, and analytics.

As a result, there is a need for decision support tools which enable individuals to make choices between different devices and applications which are informed along privacy-related dimensions. Some such tools exist already; smartphones OS providers incorporate some privacy controls and a degree of transparency about apps distributed through their app stores, while various web and mobile applications aim to give individuals greater understanding and control over the collection of their data [4, 16].

Prior research has identified various shortcomings of the existing default approaches. Privacy policies that accompany applications generally fail to inform users due to multiple problems with their structure and complexity of content [9]. Permissions management in smartphones has also been found to overwhelm users due to their complexity and the volume of permissions requested [12, 7]. Some studies find differences in the level and nature of privacy concerns amongst different populations of smartphone users [3, 6], suggesting a need for variety in tools to suit different users. In addition, users' behaviour and decision processes may be highly sensitive to the timing and format of decision support tools [1, 13, 10, 17].

In response to these problems, previous work has explored the design of 'privacy indicators', which aim to present important privacy-related aspects of services by displaying information in a summary form, including numerical scores or icons [5, 8, 11, 15, 14, 2]. Privacy indicators aim help users to understand the privacy implications of a service when making a choice about whether to use it, by showing what kinds of data are collected, for what purposes, and whether the data are shared with third parties. These attributes are clearly highly important to privacy decisions. However, decision support tools which rely entirely on such information are limited in one important respect; they focus solely on the endogenous attributes of the app itself, while ignoring exogenous aspects of the individual's context that may modulate the privacy implications of choosing that application.

An important class of exogenous factors include the individual's prior perceptions, interactions and relationships with the organisations associated with the service in question. In many cases, an application will be associated with one or more entities with whom the individual already has some experience, familiarity or even a data-collecting relationship. For instance, a user may face a choice between a smartphone payments app provided by their existing bank, another provided by a well-known digital platform they have used in another context, and another provided by an otherwise unknown company. The extent to which a user may already be engaged with the first two, and the nature of these existing engagements, could have a strong bearing on their privacy-related deliberations. To complicate matters further, even if the user does not have any prior relationship with the service provider, there may be third parties with whom they does. For instance, an app they are considering installing might share data with a prevalent third party

advertising network that already gathers data about the individual through other apps they already use.

In this study, we wanted to investigate the ways in which privacy-related decision-making processes might be affected by an individual's pre-existing situation relative to the organisation(s) associated with a particular app or service. In particular, we wanted to explore the impact of the following factors on individual's privacy-related decision-making processes;

1. Pre-existing flows of data from the individual to organisation(s) associated with a service
2. The individual's background knowledge, beliefs and perceptions about organisation(s) associated with a service
3. Real or perceived relationships between organisations associated with a service (such as 'parent-subsidiary').

Methodology

In order to investigate how these factors might influence privacy decisions, we conducted an online role-playing exercise designed to elicit a variety of perspectives. Participants were presented with a variety of hypothetical scenarios in which they were asked to choose between apps with similar functionality, but different data use and sharing practices, and explain their choices. Our primary aim was to gather qualitative data in order to identify a range of considerations and strategies that users may have in these contexts. Participants were recruited via flyers posted in public spaces across a small city in the UK, via social media, and e-mail mailing lists of a major UK university

Participants were asked to imagine they had access to a privacy indicator tool which could reveal detailed, accurate information about the privacy implications of each option. In each case, a mock-up visual display derived from this hypothetical tool was provided. It displayed information about the first-party service providers, as well as associated third parties. This included the name of the company, what data they received, and what they used it for (e.g. 'advertising' or 'analytics').

The organisations involved in each option included a mixture of popular, large consumer-facing technology companies (e.g. Google), lesser known third parties (e.g. Crittercism), and fabricated companies with names invented for the purposes of this study. We included this mixture of real and fictitious entities in order to explore how the presence or absence of familiarity and pre-existing relationships might impact decision-making processes.

The exercise began with a set of survey questions to establish participant's general privacy attitudes, levels of technology use, and demographic attributes. It then presented three scenarios in which the participants were asked to role-play making a choice between different devices and services; a home automation system, a smart TV, and a smartphone payment app. These particular scenarios were chosen because they would involve a mixture of familiar and unfamiliar organisations, in relatively new application contexts; the aim was to present plausible and engaging scenarios in which consideration of pre-existing relationships with companies might play a role in deliberation. In some scenarios, additional information was introduced in stages, and participants were asked if this would change their reasoning, and why.

An iterative thematic coding process was used to identify common themes for all aspects that were mentioned in the

free-text responses associated with each decision. Two researchers convened and consolidated themes, derived theme codes, then re-coded all rounds. Cohen's kappa k was run to determine if there was inter-coder agreement between the thematic tag assigned by each coder to each piece of text.

Results

There were 27 respondents in total. 13 were aged between 22-34, 9 were between 35-44, and 3 were between 18-21; 17 male, 8 female, 1 other. Levels of education ranged from high school (4), bachelors degree (4), masters (8), doctorate (9). 14 were in full time employment, 3 self-employed and 9 were students. All participants were smartphone users (Android (55%) and iOS (45%)), with between 15 and 102 apps installed, and frequency of use ranging from a few times a day (3), a couple times an hour (12), every few minutes (8), and 'almost constantly' (3).

Our thematic analysis of participants' responses to the role-playing scenarios resulted in the set of codes described in Table 1. There was substantial agreement between the coder's judgements (Cohen's kappa k = .894; p<.000).

Scenario 1: Home Automation System

In the first role-play scenario, participants were asked to imagine they are interested in purchasing a smart thermostat for their home. They faced a choice between functionally similar systems, one from a popular brand (Google / Nest), another from a less known brand ('HeatSmart'). A majority of participants (18) chose the HeatSmart on the basis that it appeared to collect less sensitive data types (DT) (five could not decide between the two, and 4 chose the Nest). The lack of familiarity (FA) with the HeatSmart brand relative to Google was cited as an important factor, even amongst those who chose it: as one participant (P5) stated: "Based on the privacy information, I'd probably go with the HeatSmart... However, I do trust Google with privacy more than companies I haven't heard of".

Code & Description		Example
TR:	Trust	*"I trust my bank more than many other companies"*
FA:	Familiarity	*"I have heard of them"*
EM:	Exposure minimisation	*"I would try to minimize the number of parties involved"*
CA:	Company attributes	*"bigger companies tend to handle these problems better"*
RE:	Reputation	*"[Company] have a long track record of at least keeping data safe."*
DT:	Data Type	*"I'm not comfortable with conversations being recorded"*
PU:	Purpose	*"They are going to use the data only to improve my user experience"*

Table 1: Tags and tag definitions derived from thematic coding of free-text responses to the scenarios.

Scenario 2: Smart TV

The next scenario centred on a choice between two Smart TV devices, provided by two 'market leading' manufacturers, HiVision and Bartley (both fictional). The Bartley shared data with third parties for advertising purposes. Initially, participants focused primarily on the types of data each model collected (DT) in making their decisions, and the purposes for data being shared (advertising vs improving speech recognition). A subsequent step in the role-play revealed that one of the third party advertising networks associated with Bartley (AdMob) is a subsidiary of Google. For many participants, this changed the way they felt about AdMob, both positively and negatively. Participant P2 felt reassured upon learning that AdMob was owned by a more established company, because the latter's "data policy is

scrutinized internally and externally, so I'd feel more comfortable with them having data than an independent company". For others, this corporate relationship was a reason to avoid the Bartley, to prevent their data being "sucked into the all encompassing data mountain of Google and attached to my very persistent Google identity" (P18).

Scenario 3: Mobile Payments
In the final scenario, participants were asked to imagine that they had been recommended three smartphone applications which allow them to make contactless payments using their smart phone; one provided by their existing bank, one provided by Google (Wallet), and a fictitious provider ('Payzee'). In each case, their transaction data (defined as 'what you bought, when, where, and from whom') would be shared with the app provider and their existing bank. This scenario prompted a wide range of considerations, including the general level of trust (TR) placed in different industry sectors (e.g. technology companies versus banks), and the possible relevance of certain business structures and company attributes (CA), such as whether the bank was a credit union (P10) or a building society (P14). As P12 stated, 'I have slightly more of a consumer/business relationship with my bank'. Another participant believed that a bigger company would be 'more thoroughly monitored and restricted' by regulators than a new start-up (P22).

Participants commonly cited the idea that because their bank already had access to their transaction data through their credit or debit cards, using their app presented no additional privacy concerns regarding how that data would be used, unlike Google Wallet or Payzee. For instance, P17 stated that " I assume they have access to that information already... they're not really gaining any new data about me", while P27 argued "all else being equal I'd prefer to give my data to fewer parties".

User strategies

Privacy indicators aim to provide users with answers to the most basic privacy questions, such as *who* gets to know *what* about me, who will they *share* it with, and *how* will it be used? But beyond seeking answers to these fundamental questions when reasoning about their privacy, users may also rely significantly on their perceptions of a company's business models, purpose, reputation, and even the legal and regulatory regimes the company might be subject to. Also important are the nature of existing relationships between between companies and consumers.

Relatedly, some participants expressed an 'exposure minimisation' strategy, in which their choices were informed not just by the privacy practices of a particular option, but also mediated by beliefs about which entities might have access to the data in question already. In this sense, many data flows may be 'overdetermined'; the same organisation may have access to the same data about an individual through multiple different channels. A consequence of this strategy is that a user might rationally prefer an app that shares data with a greater number of third parties, if they believe that their data has already be shared with those third parties by some other app or in some other context.

If such pre-existing knowledge, which might otherwise seem orthogonal to privacy issues, plays an important role, then user interfaces ought to support these forms of deliberation. Many participants recognised that their background knowledge was limited and conjectural; rather than leave users to draw on their own, possibly flawed, background beliefs about companies, privacy tools could instead support people to incorporate such pertinant information about organisations in their privacy decisions.

Conclusion

This preliminary study explored a range of ways in which individuals' prior perceptions and relations towards organisations might mediate their privacy-related decision-making. While the sample size is too small and insufficiently representative to draw substantial conclusions, it suggests a potential design space for privacy awareness tools which allow users to incorporate such exogenous and idiosyncratic factors into their deliberation. Existing privacy indicator tools often focus on the features of an app or device in the abstract, divorced from the user's context and prior relationships with and attitudes towards the market. This means they are likely to ignore the contextual reasons why some users might choose an app whose privacy-relevant practices are apparently 'worse' when compared to others in the abstract.

The attention that participants paid to pre-existing data exposure suggests that such tools should incorporate a personalised element, for instance, accounting for the apps that a user has already installed, and map which third parties will have already accessed which types of data as a result. By incorporating knowledge of these existing information flows, they could refine the information displayed to users. When a user views the display for a new app, they could see which of the third-party information flows associated with the app are new, and which of them already exist as a result of apps they have already installed. More broadly, such tools might aim to represent data flows in the context of social processes involving multiple entities situated in particular economic and legal contexts. We have explored this design space, developing and testing such tools, in related work [18].

Acknowledgements

Max Van Kleek, Reuben Binns, Jun Zhao, Nigel Shadbolt were supported under *SOCIAM: The Theory and Practice of Social Machines.* The SOCIAM Project is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/J017728/2 and comprises the University of Oxford, the University of Southampton, and the University of Edinburgh.

Ilaria Liccardi and Daniel J. Weitzner were supported by the *William and Flora Hewlett Foundation* and the *NSF 1639994: Transparency Bridges* grant.

References

[1] Alessandro Acquisti and Jens Grossklags. 2005. Privacy and rationality in individual decision making. *IEEE Security & Privacy* 2, 2005 (2005), 24–30.

[2] Julio Angulo, Simone Fischer-Hübner, Tobias Pulls, and Erik Wästlund. 2015. Usable transparency with the data track: a tool for visualizing data disclosures. In *Proceedings of the Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1803–1808.

[3] Zinaida Benenson, Freya Gassmann, and Lena Reinfelder. 2013. Android and iOS users' differences concerning security and privacy. In *Proceedings of the Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 817–822.

[4] Federal Trade Commission and others. 2013. FTC Staff Report Recommends Ways to Improve Mobile Privacy Disclosures. (2013).

[5] Lorrie Faith Cranor. 2006. What do they indicate?: evaluating security and privacy indicators. *Interactions* 13, 3 (2006), 45–47.

[6] Serge Egelman, Janice Tsai, Lorrie Faith Cranor, and Alessandro Acquisti. 2009. Timing is everything?:

the effects of timing and placement of online privacy indicators. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 319–328.

[7] Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. 2012. Android permissions: User attention, comprehension, and behavior. In *Proceedings of Symposium on Usable Privacy and Security*. ACM, 3.

[8] Hamza Harkous, Rameez Rahman, and Karl Aberer. 2016. Data-Driven Privacy Indicators. In *Symposium on Usable Privacy and Security*. USENIX Association.

[9] Carlos Jensen and Colin Potts. 2004. Privacy Policies As Decision-making Tools: An Evaluation of Online Privacy Notices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. 471–478.

[10] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security. In *Proceedings of Symposium On Usable Privacy and Security*. 39–52.

[11] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. 2009. A nutrition label for privacy. In *Proceedings of Symposium on Usable Privacy and Security*. ACM, 4.

[12] Patrick Gage Kelley, Sunny Consolvo, Lorrie Faith Cranor, Jaeyeon Jung, Norman Sadeh, and David Wetherall. 2012. A conundrum of permissions: installing applications on an android smartphone. In *International Conference on Financial Cryptography and Data Security*. Springer, 68–79.

[13] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. 2013. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3393–3402.

[14] Ilaria Liccardi, Joseph Pato, and Daniel J. Weitzner. 2014a. Improving Mobile App Selection through Transparency and Better Permission Analysis. *Journal of Privacy and Confidentiality: Vol. 5: Iss. 2, Article 1.* (2014), 1–55.

[15] Ilaria Liccardi, Joseph Pato, Daniel J. Weitzner, Hal Abelson, and David De Roure. 2014b. No Technical Understanding Required: Helping Users Make Informed Choices About Access to Their Personal Data. In *Proceedings of International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. 140–150.

[16] Kenneth Olmstead and Michelle Atkinson. 2015. App Permissions in the Google Play Store: Chapter 1: The Majority of Smartphone Owners Download Apps. *Pew's Report: Numbers, fact and Trends Shaping the world* (November 19th 2015), 1–19.

[17] Yong Jin Park and S Mo Jang. 2014. Understanding privacy knowledge and skill in mobile communication. *Computers in Human Behavior* 38 (2014), 296–303.

[18] Max Van Kleek, Ilaria Liccardi, Reuben Binns, Jun Zhao, Daniel J. Weitzner, and Nigel Shadbolt. 2017. Better the devil you know: Exposing the data sharing practices of smartphone apps. In *forthcoming in CHI-2017*.