Perceived risks of personal data sharing

Anya Skatova[1],
[1]Horizon Digital Economy Research
University of Nottingham,
NG7 2TU, UK. +44 (0) 115 82 32553
anya.skatova@nottingham.ac.uk

Jaspreet Johal[1], Robert Houghton[1], Richard Mortier[1], Neelam Bhandari[1], Tom Lodge[1], Christian Wagner[1], James Goulding[1]

Jon Crowcroft[2], Anil Madhavapeddy[2],
[2] University of Cambridge,
Cambridge, UK.

ABSTRACT
With ubiquitous deployment of digital technology, the amount of data that is recorded about us and retained in various virtual and physical places is enormous. These data have economic value for various governmental organisations and private companies, and we are constantly under pressure to share our personal data, whether for free or for benefit. However, it is still not clear whether this data passively recorded with technology (e.g., bank statements, geospatial location) has value to the individuals themselves.

We explore the perceived risks of personal data sharing through two online surveys, the results of which we outline here. We find significant differences in the way people treat different types of personal data: some personal data is perceived as risky to share (e.g., bank statements, geospatial location), other data requires medium protection (e.g., broadband usage, Internet browsing history), and there are types of data for which sharing has minimal perceived risk (e.g., loyalty cards information, household bills). We discuss the different decision rules individuals use when estimating the risks of sharing personal data, and the implications this behaviour has for data sharing practices and design of digital services.

Keywords
Personal data, data sharing, perceived risk, risk assessment

1. INTRODUCTION
Personal data is a rather abstract concept for most people, yet new Digital Economy services and business models increasingly require individuals to make decisions about sharing and managing their personal data. We know very little about peoples' understanding of their data or how they attribute subjective value to information like their health records or records of their domestic energy use. In particular, it is not clear whether people can adequately assess the risks in sharing data, enabling them to grant informed consent for use of their personal information. Further, when sharing multiple sets of personal data, the impact upon privacy is likely to be synergistic given the mining techniques that can be unleashed upon correlated datasets. Other research has demonstrated that individuals underestimate synergistic risks, and this underestimation is invariant across hazard domain (Dawson et al, 2012). This has implications for development of risk communications to inform and increase awareness of risks to sharing multiple types of personal data.

2. CURRENT RESEARCH
We investigated whether users understand their data as having a value and if valuations across different types of data consistent across individuals. We conducted two online surveys to study (1) whether individuals differentiate risks of sharing different types of personal data and (2) whether individuals are consistent in judging the data sharing risks of single (e.g. bank statement) versus synergetic types (bank statement and geospacial location at the same time).

3. Study 1

3.1 Participants
144 volunteers were recruited via departmental mailing lists and word of mouth to fill in an online survey for no monetary incentive. The survey was distributed using Qualtics software. 60 participants completed the survey. The mean age was 30 years old, with a range of 19 to 61; 65% of participants were female; 66.7 % claimed English as their first language; 86.7% reside in UK; and 85% used a smart phone on regular basis.

3.2 Materials and Procedure
Participants read a vignette asking them to imagine that they just bought a new smartphone. In addition to having all the usual functions (camera, phone, Internet), this smartphone also had a special application that aimed to assist in their everyday lives by tracking items of personal data. This personal data could reflect various aspects of their daily behaviours and any transactions they made (including where they'd been via GPS data, their health records, their bank records, their household bills, their social networking activities etc). Participants were told that when they first used the personal data app they would be given two choices: either (A) to pay a given amount of money for the service for their data to remain private and never accessed by the company providing the service or (B) to run the app in the "free" mode. In the latter case, they did not have to pay anything but the company providing the app anonymously collected personal data about them when the participant used it.

After reading this background vignette, participants received 72 items in random order but all with the same structure, asking them to make a forced hypothetical choice between the use of service for free when some data about them was collected by the company, or paying an amount of money to use the app without sharing personal data with the company. The type of data collected and the money they were asked to pay varied across 72 items. The following types of data were used in this study: bank account statement, health records, social networking sites profiles and activities, physical location history, household and mobile phone bills, loyalty cards, online advertising click history, Internet browsing history, Internet search history, and demographic information. The amounts of money participants were required to pay to protect their data were 50 pence, £1, £5, £10, £15, £20.

3.3 Results
We estimated perceived risks in sharing different types of personal data by calculating the percentage of participants who were ready to pay a high price (£20), moderate price (from 50 pence to £15) or nothing at all to secure the privacy of each type of personal data (see results in Figure 1). We found three distinct groups of personal data. First, those personal data that were highly protected: more than 70% of participants chose to pay the

maximum possible price (£20) to secure the privacy of bank statements and digital communication history data. Second, those personal data that were moderately secured, such as health records, social networking site profiles and activities, and physical location history, with about 50% of participants ready to pay the maximum price to secure this data, while about 20% of participants were unprepared to pay anything. Household bills, online purchasing history, Internet browsing and search history, and demographic information were perceived as risky to share by only around 20% of participants, with the majority either unwilling to pay anything or willing to pay only a small amount. Finally, loyalty cards data and online advertising click history was not perceived as risky to share for 70% of participants.

These results demonstrate that not all types of personal data were perceived as equally risky/non-risky to share, with some data categories more likely to be protected than others.

4. Study 2

4.1 Participants

In the second study, 1317 volunteers were recruited via a variety of mailing lists that reached university and non-university employees, as well as university students. 853 completed the survey and, if they wished, were entered in a prize draw for £50. The survey was distributed using Qualtics software.

The mean age was 26 years old, ranging from 18 to 64; 63.2% - female, 79.2 % claimed English as their first language, 98.5% reside in UK, 77.6% used a smartphone on regular basis.

4.2 Materials and Procedure

Participants were given the same vignette about their new smartphone as in Study 1, but now with three modifications.

First, unlike Study 1 where participants had to make judgements about all types of personal data, in Study 2 each participant was presented with just one combination of two data items from the list. Specifically, on the first occasion they were asked about Item 1, on the second occasion about Item 2, and on the third occasion they were asked whether or not they would pay to protect both items in conjunction. For example, they might have been asked whether or not they would pay to protect bank account statement information, followed by a separate choice to pay to protect physical location history or not, followed by a third choice to pay to protect the combination of bank account statement information and physical location history, or not. The pairs of items participants were asked to give judgements about were randomized between participants.

Second, based on the findings of Study 1 we restricted the overall list of items we presented to bank account statements, social networking site profiles and activities, physical location history, electricity bill, broadband usage bill, mobile phone bill, loyalty cards, Internet browsing history, and demographic information. Third, we increased the potential monetary value to pay to protect personal data to a range from £1 to £100.

4.3 Results

To estimate perceived risk to share different types of personal data we calculated an average amount of money that individuals were ready to pay to protect each type (see Figure 2). The results replicated findings from Study 1: there were four different types of data by perceived risks to share, with the items in the categories being similar to the ones in Study 1. The highest perceived risk belonged to bank account statement data: individuals were ready to pay almost £30 to protect it, with this being significantly higher than for any other type of data. Social networking profiles and activity and location history data were assessed as moderately risky to share: individuals were ready to pay just £20 to protect it, statistically significant compared to the other types of data. The lowest perceived risk belonged to electricity data: individuals were prepared to pay just £10 to protect it, and this was significantly lower than for any other type of data in our study.

We further looked whether participants perceived sharing a combination of two types of data as higher risk. Our results showed that participants mostly underestimated the risk of sharing two types of data simultaneously when compared to the sum of the risks their assigned to each single type of data (see Figure 3 for representative results). For example, when judging social networking site profiles and activities, and bank account data separately, participants assigned £20 and £30 respectively to each. When asked about sharing both sets of data simultaneously, they were ready to pay under £30, significantly lower than £50, the sum of values for each item.

We also investigated the decision rules that guided people's judgements for synergic risks (see Figure 4). The most popular strategy, adopted by 33% of individuals, was to state an equal amount to share any type of data, whether alone or in combination ("constant", Figure 4). For example, such individuals would indicate that they were ready to pay £10 to protect bank statement data, £10 to protect loyalty card data, and £10 to protect the combination of bank statement and loyalty card data. The second most prominent strategy, adopted by 32% of participants, was to use the value of one item to define the value of both items ("maximum", Figure 4). For example, they were ready to pay £20 to protect bank statement data, £5 to protect loyalty card data, and £20 to protect the combination. Finally, only 14% of participants used an "additive" strategy by combining the value of two separate items to calculate the perceived risks of sharing two items at the same time. For example, if they stated they were paying £20 to protect bank statement data and £5 to protect loyalty card data, they would state that they were paying £25 to avoid sharing both bank statement and loyalty card data. 21% of participants did not use a strategy when deciding about risks of synergetic data.

5. CONCLUSIONS

The main contribution of this work is to demonstrate that different risks are associated with sharing different types of personal data: privacy is reliably valued differently depending upon the type of personal data involved. Our results show that there were consistent differences in the way people make decisions about sharing their private data and assessing the risks of sharing. Further, there are individual differences in the perceived risks of sharing personal data, and differences in individual strategies as to whether to share or protect personal data. Three main strategies, namely constant, maximum and additive are consistent with previous literature: people often do not simply add risk evaluations of different items. Our findings highlight individual differences in data sharing tactics that provide future challenges for framing communication messages to raise people's awareness of sharing their personal data in general and specifically synergetic pieces of personal data.

6. ACKNOLEDGEMENTS

This work was funded by the RCUK Horizon Digital Economy Research Hub grant, EP/G065802/1.

7. REFERENCES

Dawson, Ian GJ, Johnnie EV Johnson, and Michelle A. Luke. "Do people believe combined hazards can present synergistic risks?." *Risk Analysis* 32.5 (2012): 801-815.

Figure 1. Percentage of participants binned by how much they were ready to pay (£20, red bars; from 50 pence to £15, grey bars; nothing, white bars) to secure different types of personal data, Study 1. Types of data: "Bank", bank account statements; "D/com", digital communication history; "Health", health records, "Soc/net", social networking profiles and activities; "Loc", geospacial location; "Bill", household bills; "O/purch", online purchase history; "L/card", loyalty cards information; "O/ads", online ads click history; "iB", Internet browsing history; "iS", Internet search history; "Dem", demographic information. "High", "medium" and "low" refer to different categories of personal data in terms of privacy.

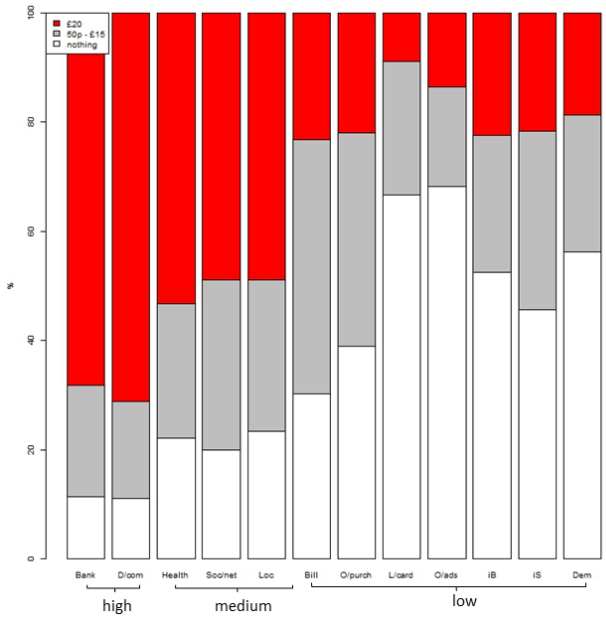

Figure 2. Average amount individuals were ready to pay for each type of personal data, Study 2. Error bars represent standard error of the mean. "High", "medium" and "low" refer to different categories of personal data in terms of privacy. Types of data: "Bank", bank account statements; "Soc/net", social networking profiles and activities; "Loc", geospacial location; "Electricity", electricity bills; "Broadband", broadband bills, "Mobile", mobile phone bills, "L/card", loyalty cards information; "iB", Internet browsing history.

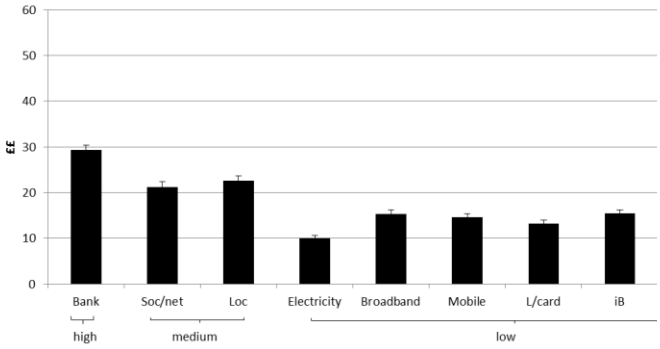

Figure 3. Average amount (in pounds) participants were ready to pay to protect bank statement data combined with other types of personal data. Dotted bars indicate the sum of amounts participants were ready to pay when judging about the items separately, black bars indicate the amount participants were ready to pay when making a judgement about a pair of items, Study 2. Error bars represent standard error of the mean. Types of data are notated similar to Figure 2.

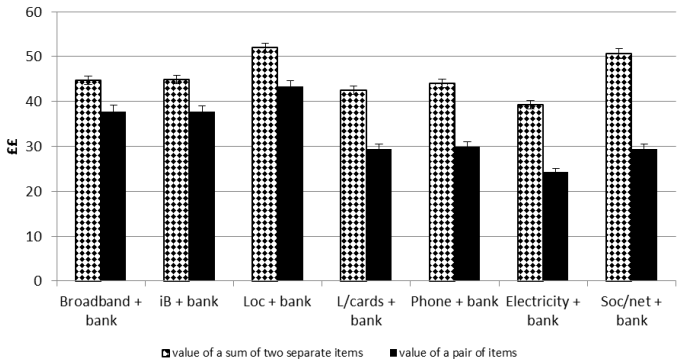

Figure 4. Proportion of individuals using different strategies (random, constant, maximum and additive) to assess risks of sharing multiple types of data, Study 2. See description of the strategies in the body of the text.