Investigating Loan Applicants' Perceptions of Alternative Data Items and the Effect of Incentives on Disclosure

Charlene Jennett, Sacha Brostoff, Miguel Malheiros, M. Angela Sasse
Dept. of Computer Science
UCL (University College London)
London, WC1E 6BT, UK

c.jennett, s.brostoff, m.malheiros, a.sasse@cs.ucl.ac.uk

ABSTRACT
Lenders use information about loan applicants to predict whether a person is a good or bad credit risk; however borrowers express reservations about disclosing their personal information. In this paper we describe the design of a study in which we try to identify which data items have bigger privacy costs for individuals and whether it is possible to adjust lenders' data collection procedures in order to improve the privacy of the borrowers while maintaining or improving the accuracy of risk assessment methods. We aim to explore whether consumers could be equitably encouraged to give different information than they do presently, by offering incentives for disclosures. These incentives are: an uncertain long term financial gain; a certain short term financial gain. We also explore an inequitable manipulation using peer pressure. The advantages and disadvantages of this methodological approach are also discussed.

Categories and Subject Descriptors
H.1.2 [**Models and Principles**]: User / machine systems – *human factors, human information processing*

General Terms
Experimentation; Human Factors.

Keywords
Loans; Privacy; Incentives.

1. INTRODUCTION
In credit scoring an automated system uses a "score card" to classify the applicant as a "good" or "bad" risk based on their application form and their previous credit history; the "good" customers being those that are deemed more likely to repay their debt. The approach is limited to a set number of questions, often with categorical answers, and applicants' reliability is estimated upon how well they perform in respect to these set questions, with performance being how closely they match the ideal good borrowers' profile. There is evidence to suggest that automated systems are better predictors of which applicants would be "good" or "bad" customers than human decision makers [8]. However a limitation of credit scoring is that it can lead to financial exclusion for certain social groups, e.g. the unemployed, the low paid, the elderly, young people not in education / employment / training, black and ethnic minorities, migrants, and people with disabilities [2]. The automated approach is most accurate with applicants who score very high or low, but is relatively inaccurate for applicants scoring near the cut-off value between 'good and 'bad' [5]. Thus applicants from the more marginalized social groups may not match the description while still being good borrowers. Having a

wider set of questions could increase the predictive power of score cards, increasing a person's chances of showing their ability to repay. For example, regular payment of utility bills is information that could be useful but is currently not used by some lenders. Conversely, extra information about an applicant could also reveal to the lender that a person that might have previously been classed as a "good" customer is actually a "bad" customer. This outcome could be advantageous to the applicant as well as the lender, as both counterparties would be protected from irresponsible lending.

2. ALTERNATIVE DATA ITEMS
A common idea in the financial services community is that additional information about borrowers would be beneficial to lenders: more data means better statistical modeling and fairer credit scoring [5]. There are risks that the existing data items are sub-optimal – the information may deliver inaccurate predictions for groups that are currently financially excluded, and more generally may be too sensitive to consumers for the value delivered in risk management. The potential risks of additional and excessive data collection are not mooted either: triggering consumer privacy-protective behaviours that reduce data quality and consequent business decisions, increased data management, security measures and breach costs, and so on.

For the current study we generated a range of items with the intention that they would have a wide range of acceptability to respondents, and that they would have either intuitive face validity, or that they have been mentioned in past literature. We intuitively divided the items into seven main categories, naming them: the ability to make regular repayments (utility bills, tax, debt); one's reliability for maintaining assets and health (e.g. insurance claims); one's work ethic; character references (e.g. recommendations from others); indicators of personality (self-esteem, materialism, self-discipline); level of education (personal capability, financial capability); and relationships with others (ability to commit to partners, current stability, family stability during childhood). Several of these measures are not currently feasible because of practical difficulties in collecting and verifying the data, and others due to ethical and legal data protection issues. However, laws and norms change, and we wish to explore the current privacy space with respect to the financial services – i.e. what other data items are people currently willing to trade for financial outcomes?

3. DATA DISCLOSURE CONCERNS
General and specific factors contribute to peoples' data disclosure concerns. The perceived outcome of disclosure matters when data might reflect negatively on the data subject. For example, a person

might be reluctant to reveal he is a smoker as it could result in a higher life insurance premium [9]. Another fear is the possibility of one's privacy being invaded if the information was disclosed to or obtained by unauthorised persons. Contextual factors also play a role in privacy preferences. For instance, when shopping at a supermarket one's purchases are on display to others; however when shopping online, if a person finds out that the retailer is storing their purchase information this might be viewed as an invasion of privacy [6]. Similarly, regarding loan applications, one might expect consumers to question "Is it appropriate to give this information in this context?"

We assume that any new data items that could replace several existing data items while maintaining or increasing a score card's predictive power would be personal, and that people would be reluctant to disclose such data to lenders, even if they produce a net increase in their privacy. If the new data items *could* be successfully collected and verified, how could lenders convince consumers to give them permission to use them for credit scoring?

4. THE USE OF INCENTIVES

Acquisti [1] suggests that "soft paternalism" might be the answer. Researchers might need to design systems that "nudge" individuals into being more open to making certain choices. The way a request is "framed" affects how individuals respond to it. For instance, a 2002 Jupiter research study found that 82% of online shoppers were willing to give personal data to new shopping sites in exchange for the chance to win £100 [4]. Even modest incentives, such as "loyalty points" or a reduction in bank fees, can generate changes in consumer behaviour [3].

Social influence also impacts consumer behaviour. For example, Thaler and Sunstein [7] describe how Montana successfully adopted a large-scale educational campaign with the message "Most (70 percent) of Montana teens are tobacco free". This strategy "nudged" college students to change their behaviour, leading to statistically significant decreases in smoking.

5. STUDY DESIGN

To investigate the impact of both monetary and social incentives, four different versions of a loan information request were created:

- **No Incentive.** You're asked if you would voluntarily be willing to provide extra information about yourself.

- **Long Term Gain – Uncertain.** As part of the scenario you are told that if you provide extra information about yourself and you are accepted for the loan, your interest rate will be lowered. The more information you give, the lower your interest rate could potentially be.

- **Short Term Gain – Certain.** If you provide extra information about yourself you will receive payment, regardless of whether or not you are accepted for the loan. The more information you give, the more payment you will receive (up to £50).

- **Peer Pressure.** You are asked if you would voluntarily be willing to provide extra information about yourself, and are told that 90% of other respondents willingly gave the lender extra information about themselves.

When presented with information categories and asked to what degree of detail they would be willing to provide that information to a lender, we predict that participants will give more information when given an incentive compared to when no incentive is given. Concerning the alternative data items, we aim to answer the following questions: (1) Which types of information are people most reluctant or willing to share? (2) Which of these incentives "nudge" participants to disclose more of the new data items?

Currently we are conducting a survey study investigating the extent that people rate themselves as "comfortable" in disclosing each of the alternative data items to lenders (also existing loan application form items). The effect of incentives, and whether this leads to greater disclosure, will be investigated in the next study.

6. ADVANTAGES AND DISADVANTAGES

By investigating how people respond to a loan request with no incentive versus incentives, we believe that this methodology will allow us to gain a greater insight into people's privacy-related attitudes around particular data items in a financial services context. Furthermore, this research is novel as the use of incentives has not previously been investigated in the context of loan requests. A disadvantage of this methodology, however, is that participants are asked to imagine the scenario - when faced with an actual loan possibility, a person might behave differently. Studies involving actual monetary risks could be used, if suitable simulations of the real world financial risk could be designed.

7. REFERENCES

[1] Acquisti, A. (2009). Nudging Privacy: The Behavioral Economics of Personal Information. *IEEE Security and Privacy*, 7 (6), 82-85.

[2] Datta, K. (2009). Risky migrants? Low-paid migrant workers coping with financial exclusion in London. *European Urban and Regional Studies, 16 (4)*, 331-344.

[3] Elliot, A. (2009). Fair banking: The Road to Redemption for UK Banks, pp. 37-39. CSFI publications: London, UK.

[4] Grossklags, J. & Acquisti, A. (2007). When 25 cents is too much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information. In *Proceedings of the Sixth Workshop on the Economics of Information Security*, Pittsburgh, PA.

[5] Hand, D. (2010). Personal Communication.

[6] O.Hara, K. & Shadbolt, N. (2008). *The Spy in the Coffee Machine: The End of Privacy as We Know It*, pp.47-80, 210-231. Oneworld Publications: Oxford.

[7] Thaler, R. H. & Sunstein, C. R. (2008).*Nudge: Improving Decisions About Health, Wealth and Happiness*., pp. 57-78. Penguin Books: London.

[8] Thomas, L. C. (2000). A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers. *International Journal of Forecasting, 16*, 149-172.

[9] Varian, H. R. (1996). Economic aspects of personal privacy. In *U.S. Dept. of Commerce, Privacy and Self-Regulation in the Information Age, 1997*.