

Source-side Valuation of Personal Data: Results from an Expert-Survey

Vikash Gautam

Koan Advisory, New Delhi, India

Email: vikash.igidr@gmail.com

Abstract: This paper discusses two aspects of the source-side valuation of the personal data using an expert-survey that pertains to its time value and monetary value. The sought responses in the survey are with reference to a sequence of monthly income levels and make a distinction between sensitive and non-sensitive personal data-types. The results suggest a considerable and increasing time value of personal data that may empower more efficient search and transaction over time. Additionally, loss aversion pertaining to personal data usage increases with income levels, and is larger than pure monetary prospects.

Keywords: Personal Data; Expert-Survey; Valuation

JEL Classification Number: D90

1. Introduction

The globalised nature of economic activities and the advances in computations have considerably altered the way one looks at the data(or information). A considerable emphasis is on the personal data of individuals for customised provision of goods and services at lower transaction costs on one hand; and financial frauds, privacy risks and identity thefts on the other. This paper, using a focused-group survey, attempts to discuss two aspects of individuals valuation of their personal data that pertain to time and monetary value. The attempt is directional and it aims to generate a larger debate on the given issue.

2. Survey Design

The specificity in associating a value to the personal data, as knowledge and perception on the corresponding legal and institutional developments, individual and social identities, and accrual of benefits from personal data sharing, prompted us to conduct an expert-survey, rather than an open survey. Additionally, data is intrinsic to all aspects of individuals' being, which for an evaluation purpose, entails a perverse mapping of several heterogenous dimensions to a unidimensional plain. Focusing on an expert-survey, where the experts as in the present case are professionals in the services sector in two metros of India (Mumbai and Delhi), controls for several sources of heterogeneity. In pursuing the

expert-survey, however, we are cautious of the fact that the results so derived have limited applicability, and therefore, this research is merely directional and it attempts to offer a framework for more substantive investigations.

The expert-survey consists of 36 respondents of which 28 work in the private sector while the remaining work in the public sector; and their monthly incomes range between INR 10,000 to above INR 150,000. To resolve ambiguity that may arise due to differences in the expertise areas and the exposure to the issues related to the personal data, we adequately explained the survey questionnaire to the respondents. The survey makes a distinction between the sensitive and the non-sensitive personal data as per the Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules of India, 2011. The sensitive personal data consists of all personal data relating to passwords; financial information; physical, psychological and mental health conditions; sexual orientations; medical records and history; and biometric information. Non-sensitive personal data includes all personal data which is not sensitive, and sensitive data that is freely available or accessible in public domain or furnished under the Rights to Information Act of 2005 or any other law for the time being in force. Although these broad groupings of personal data-types offer an important distinction, the within-group differences between the data-types may also be considerable. To take care of this aspect, we instruct the respondents to consider only those personal data-heads, in each group, that have financial implications. This makes our framework tractable.

The survey consists of two key aspects of the sensitive and the non-sensitive personal data-types that pertain to their time and monetary value. The reference for responses is a progressive sequence of monthly income levels from INR 10,000 to INR 500,000. Some of these monthly income levels approximately correspond to the current monthly income of the respondent, while others correspond to a hypothetical monthly income level. Specifically, at each level of monthly income and for each of the sensitive and the non-sensitive personal data-types, we ask the respondents about the savings due to the personal data-based customisations, and requisite compensatory amount in case of a mala fide use of their personal data.

In the survey design, a hypothetical bias may systematically influence the responses. Hypothetical bias is the difference between what a person indicates of doing in the survey with reference to a hypothetical state and what she actually does (Loomis, 2014). In discussing the mechanisms leading to the hypothetical bias, Zauberman et al. (2006) suggest that in evaluative judgments, respondents initially form a hypothesis and then integrate the information in the context of that hypothesis. This 'anchoring- and

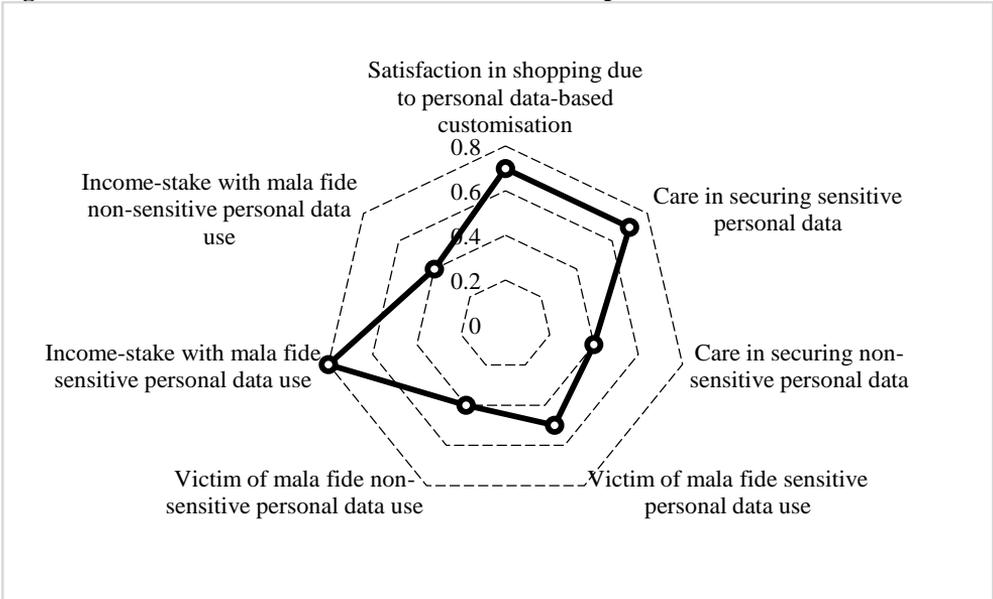
adjustment-type of mental processing' may mechanically influence the responses at each income level. However, whether this influence leads to inflated, deflated or insignificant changes to the actual value, will depend on the perceptions around the anchoring income level in the set of monthly income references. These perceptions may be based on the inferences from a similar issue in the current market or in an alternate market. For example, Harrison (2006) finds right-truncated willingness to pay (WTP) in many respondents based on their assessment of the trading price of a similar good. In the current context, given the value of personal data is a more intensely discussed subject in the advanced countries, where individuals have higher stake in terms of foregone income, we expect the stated responses (by Indian respondents) to be left-truncated.

To ascertain that the advanced countries' frames are indeed an intrinsic issue in the survey responses, we engaged in a set of pre- and post-survey informal deliberations with the respondents on almost all aspects of the personal data in the questionnaire. Such 'cheap talks' are noted to have considerable influence in deducing the frames and influences on respondents (Cummings and Taylor, 1999). We notice that our priors, i.e., advanced countries' frames, go through in most of the cases. Correspondingly, we notice that many respondents state monetary amounts that are implausible high fraction of their reference monthly income. To take care of these biases due to the reference-frames, we report the medians as the representative responses as suggested by Loomis (2014).

3. Survey Results

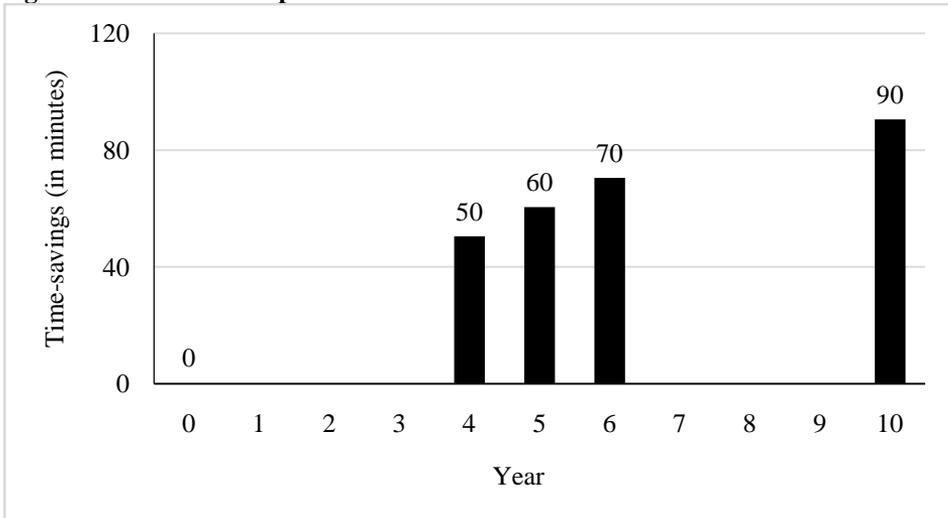
Among the respondents, eight of them reveal that they have been a victim of mala fide sensitive personal data use while 13 of them reveal that they have been a victim of mala fide non-sensitive personal data use. Moreover, on average, the respondents spend four hours of their weekly time on social media and use five paid software freely through torrents or other similar file-hosting platforms. Figure 1 reports respondents' beliefs on personal-data related attributes (both between zero and one). The likelihood of income at stake with mala fide use of sensitive personal data is perceived to be the highest at 0.8, while the likelihood of being a victim of such a scenario is marginally lower at 0.7 despite the stated intensity of care also being 0.7. The corresponding likelihoods for non-sensitive personal data in all the three cases are 0.4. Interestingly, the respondents report the intensity of satisfaction in shopping due to personal data-based customisation also at a moderately higher level of 0.7. Thus, on average, the respondents seem to be mindful of both, the advantages accruing due to personal data-sharing and the corresponding risks, and their optimal preferences on data sharing is not on either of the extremes.

Figure 1: Perceived likelihoods and intensities of the personal-data related attributes



For assessing the time-value of the personal data, the sought responses intricate two distinct features – the 'experienced' time-savings over the last five years and the 'expected' time-savings over the next five years. The respondents reveal no considerable difference between the time-saving potential of the sensitive and the non-sensitive personal data, and in most cases attribute the accrued time-benefits to a mix of these personal data-types. Figure 2 presents the median values of these time-savings. For convenient comparability, time-saving values in each year are difference-normalised by the year-2013 value, which we code as year zero value. The cumulative distribution of time-savings reveals that the increase in time-savings is more for the experienced years (year 0 to year 5) than the expected years (year 6 to 10). This may be because of two reasons. First, an increasing change in time-savings may only be possible with an increasing sharing of the personal data on margin. If one has already shared low-risk personal data for customised provisions, additional personal data sharing, which is riskier, may be improbable as the consequent costs may considerably outweigh the benefits so accrued. Second, even though one may expect a consistent or increasing change in time-savings due to personal data-based customisations, such an expectation may not be degenerate. The ambiguity or uncertainty around these expectations may also result in a discounted valuation of the future benefits.

Figure 2: Time-value of personal data



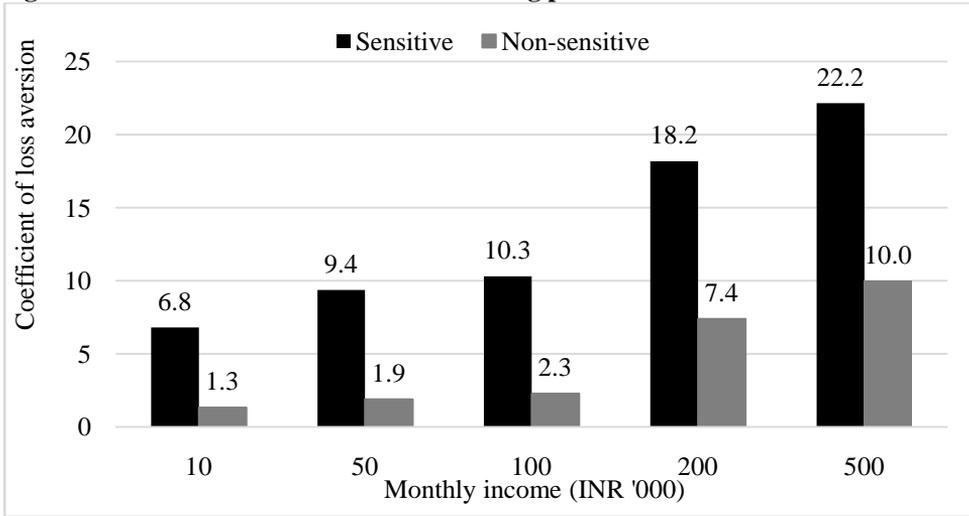
Note: Year zero on X-axis represents calendar year 2013.

The other important aspect of the survey pertains to perceived monetary gains and losses due to bona fide and mala fide use of the personal data, respectively. Two commonly used approaches exist to compute the valuation of losses relative to the gains, i.e., coefficient of loss aversion (CLA). Kahneman and Tversky (1979) suggest that for x as the underlying stake (i.e., personal data in our case) and $U(\cdot)$ as the utility function, if $-U(-x) > U(x)$ for all x , the CLA is the mean or median of $-U(-x)/U(x)$. Wakker and Tversky (1993) use a stronger definition for the CLA as the mean or median of $-U'(-x)/U'(x)$ given $-U'(-x) > U'(x)$ for all x . While the first definition projects CLA as the utility from the outcomes as losses and gains; the second definition captures individual's attitude as the degree of steepness of the utility function for losses relative to the steepness of the utility function for gains, irrespective of the degree of utility for gains and losses. Thus, given that the second definition can unambiguously separate the attitudes towards losses from utility curvature, we prefer it over the first definition.

Figure 3 presents the CLA for the sensitive and the non-sensitive personal data-types across all the reference monthly income levels. Two results are noteworthy. First, the CLA increases with the reference income levels for both the personal data-types. However, the increase is most steep in moving from the monthly reference income of INR 100 thousand to INR 200 thousand. This pattern suggests that although a change in the CLA may be intrinsically associated with the characteristics of the reference income, there may also be important differences in an individual's socio-professional attributes at different income

levels that influence the CLA. For example, individuals at higher income levels typically have several social and professional networks that may not exist at the lower income levels and that may influence the attitude towards relative losses and gains. Fehr-Duda et al. (2010), on a similar note, notices an increasing relative risk-aversion with stake size not only because of the change in the monetary payoffs, but also because of the change in the framing weights of an individual. Second, given that the CLA for monetary prospects are typically in the range of 1.5 to 5 in the literature (Abdellaoui et al., 2007), personal data, to a large extent, has a more solemn position than money for a typical individual. This result is important in motivating public policy for having a more representative stance on issues related to the personal data.

Figure 3: Coefficient of loss aversion in valuing personal data



4. Conclusion

This study discusses some aspects of the source-side valuation of the personal data using an expert-survey. A major result emerging from the current exercise pertains to a considerable and increasing time value of personal data which may empower more efficient search and transaction over time. At the same time, the potential of losses from mala fide use of the personal data, which may lead to monetary theft, identity theft, reputation loss, among others, that are enormous. Thus, it becomes important to restrict mala fide use of the personal data without hindering its bona fide use. Four aspects of this disposition are pertinent to the discourse on the personal data. First, several players are intrinsically involved in dissemination and use of personal data that include consumers,

social groups, producers, marketplace platforms and governments. However, the value obtained by each player, individually or at different levels of aggregation, is end-use-specific, which is dynamically influenced by the priors of the end-user and the state of computational analytics of the time. Second, the valuation of personal data is asymmetric in the domains of losses and gains, with losses weighing disproportionately more than the gains. Although this is true for most of the underlying assets, personal data being end-use-specific rather than source-specific add an additional layer of complexity to this asymmetry. Third, the policy maker's problem is to aggregate all private and social values associated with the usage of personal data and then to devise a mechanism that generates sufficient incentive to encourage bona fide personal data sharing and discourage mala fide usage of the personal data. Fourth, personal data are not only a sought-after asset domestically, but in a globalised world, it is internationally useful and related to several other complex problems of exchange and operation such as foreign trade and investment and global value chains. Thus, it is imperative to enforce a domestic regulation on the personal data usage that is congruent with the rest of the world.

References

- Cummings, R. and L. Taylor, 1999, Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method, *American Economic Review*, 89, 649-65.
- Fehr-Duda, H., A. Bruhin, T. Epper and R. Schubert, 2010, Rationality on the rise: Why relative risk aversion increases with stake size, *Journal of Risk and Uncertainty*, 40, 147-80.
- Harrison, G., 2006, Experimental evidence on alternative environmental valuation methods, *Environmental and Resource Economics*, 34, 125-62.
- Kahneman, D and A. Tversky, 1979, Prospect theory: An analysis of decision under risk, *Econometrica*, 47, 263-92.
- Loomis, J., 2014,2013 WAEA keynote address: Strategies for overcoming hypothetical bias in stated preference surveys, *Journal of Agricultural and Resource Economics*, 34-46.
- Wakker, P and A. Tversky, 1993, An axiomatization of cumulative prospect theory, *Journal of Risk and Uncertainty*, 7, 147-75.
- Zauberman, G, K. Diehl and D. Ariely, 2006, Hedonic versus informational evaluations: Task dependent preferences for sequences of outcomes, *Journal of Behavioral Decision Making*, 19, 191-211.