PERCEIVED RISK AND DESIRED PROTECTION: TOWARD A COMPREHENSIVE UNDERSTANDING OF DATA SENSITIVITY

Yasunori Fukuta, Kiyoshi Murata, Yohko Orito

Meiji University (Japan), Meiji University (Japan), Ehime University (Japan)

yasufkt@meiji.ac.jp; kmurata@meiji.ac.jp; orito.yohko.mm@ehime-u.ac.jp

ABSTRACT

This study clarifies the characteristics of the perceived risk of personal data release and the required protection as a preliminary step toward a comprehensive understanding of data sensitivity, and has long been regarded as the key parameter distinguishing data that should be protected from data that should be utilised. Few studies have empirically considered the cognitive characteristics of data sensitivity. It is essential to consider both perceived risk and desired protection when seeking a comprehensive understanding of data sensitivity. Thus, we quantitatively examined the characteristics of both components using four types of personal data (two of which are considered sensitive under Japanese law). The authors surveyed 420 Japanese subjects and analysed the results using the Friedman test and the Mann-Whitney U-test. The perceived risks and desired protections differed significantly among the four types of data, and legally defined data sensitivities did not always explain the observed differences. The extent of interest in personal data increased the perceived risk and the desire for protection. The effects of various personal factors including gender and the tendency to self-protect were relatively weak, so further analysis is required. We discuss the remaining issues and future research directions.

KEYWORDS: non-parametrical analysis, personal data, protection request, risk perception, sensitive data.

1. INTRODUCTION

We are becoming increasingly dependent on personalised online services. Every individual is part of a vast service ecosystem involving personal data collection, distribution, and storage. The more personal data is disclosed, the greater the benefit to the individual: disclosure improves quality of life as the entire ecosystem responds. However, this enhanced service convenience and quality may be accompanied by negative trade-offs including invasion of privacy, unfair discrimination, and fraud. It is difficult to maximise advantages while minimising disadvantages. The protection of certain types of personal data is the essence of good data management, and personal data sensitivity has long been regarded as key in this context. Scholars have been investigating this issue for at least 40 years. Turn and Ware (1976) used personal data sensitivity as a classification axis in a pioneering discussion of how such data should be categorised. The 2012 EU Data Protection Directive (later replaced by the 2018 General Data Protection Regulation) distinguished some forms of personal data, the disclosure of which would seriously affect fundamental human rights, from other types of personal data (European Commission 2012), and strict conditions are imposed on handling those particularly personal – 'sensitive' – data. Japan's Act on the Protection of Personal Information (APPI; revised in 2015) defines some sensitive data as yo-hairyo (special care is required). Specifically, this includes data related to race, religion, social status, medical history, any criminal record, whether an individual was a victim of crime, and anything else prescribed by cabinet order as requiring special care to preclude discrimination, prejudice, or another disadvantage (paragraph 3, Article 2, Japan APPI 2015). Thus, the concept of data sensitivity is widely used when managing personal data and respecting privacy, but few studies have empirically examined the cognitive characteristics thereof. Here, as a first step toward a comprehensive understanding of data subtleties, data sensitivity is quantitatively analysed in terms of perceived risks and required protections.

2. LITERATURE REVIEW

2.1. Personal data sensitivity

No clear consensus had been reached on a definition of personal data sensitivity. Most studies have conceptualised such sensitivity in terms of the potential negative consequences suffered by an individual if personal data were inappropriately collected, distributed, stored, or used. Turn and Ware (1976) suggested that personal data would become sensitive "when its uncontrolled dissemination may have adverse effects on the individual concerned and on his activities" (p. 303), noting that adverse effects ranged from mild annoyance to serious physical and mental harm. Others have defined personal data sensitivity or 'sensitive data' in terms of various adverse effects such as privacy risks (Sapuppo, 2012), personal identification (Malheiros, Oreibusch and Sasse, 2013), and unfair discrimination and prejudice (Japan APPI, 2015). The higher the estimated probability of negative consequences, the higher the sensitivity. Therefore, personal data sensitivity is the extent to which an individual does not want anyone to use or disclose data because of a perceived risk of negative consequences. Ackerman, Cranor and Reagle (1999) considered someone who is 'comfortable' with disclosure to be the opposite of someone who is 'sensitive' about data disclosure. Sapuppo (2012) defined data sensitivity as an unwillingness to share. Thus, data sensitivity is a cognitive and/or affective feeling. However, the person evaluating data sensitivity is not necessarily the person who 'owns' the data. Two methods have been used to evaluate sensitivity (Fule and Roddick, 2004; Al-Fedaghi, 2012). The first involves having well-trained and highly experienced experts scientifically and comprehensively explore the status quo and define what is socially acceptable; this is the principal approach used for legislation (PPC Japan, 2016). The other method involves assessing the perceptions of data subjects via quantitative or qualitative methods (e.g., surveys and in-depth interviews); data from many subjects can be aggregated to define sensitivity. Although these methods differ (Fukuta et al., 2017), they both apply the concept of data sensitivity, and use of both methods may be essential when seeking a comprehensive understanding of such sensitivity.

2.2. Conceptual structure and research tasks

Our conceptual review of personal data sensitivity revealed how such sensitivity may be structured. Sensitivity reflects the risk perceived by a data subject when his/her personal data are collected, distributed, stored, and used. The extent of perceived risk determines the subject’s attitude about data utilisation and disclosure. If a high risk is perceived, the subject wishes to keep the data secret and/or assigns high priority to data protection. It is thus important to explore how data subjects perceive risk. Subjects may request that sensitive data be protected. As mentioned above, data sensitivity is an unwillingness to allow anyone to access or use the data. Thus, data-handling entities and lawmakers will receive requests to prioritise protection over utilisation. Such requests are very important because they connect data sensitivity to data management and privacy protection. Finally, data sensitivity should be discussed at a collective, not an individual, level. Our conceptual review revealed that data sensitivity is intrinsically subjective, i.e. an attitude about personal data disclosure and use. However, given the roles played by others in data management and privacy protection, it is best to analyse the topic collectively regardless of whether sensitivity is based on expertise or a ‘general feeling.’ In recent years, online services have begun to automatically measure data sensitivity (e.g. Fule and Roddick, 2004). However, most research to date has not focused on customised data disclosure or use, and has instead focused on data sensitivities at the level of a nation, a demographic population, or a cultural group.

Here, we empirically examine the personal data sensitivities of ordinary Japanese people in terms of risk perception and protection requirements. To collect this basic material, which is required for a comprehensive understanding of sensitivity, we explore differences in perceived risk and protection requirements among four types of personal data: political orientation (e.g. party membership and political beliefs); health status (e.g. mental and physical medical histories; diagnosis and treatment records); economic/financial status (e.g. deposits, real estate holdings, and debt); and consumption (shopping history and service records). The former two types of data are defined as ‘sensitive’ by the APPI and the latter two are not. Differences in the perceived risks and protection requirements among these types of data reveal some of the cognitive characteristics of data sensitivity. We also explore the effects of personal factors (gender, interest in personal data, and a self-protective tendency) based on our expectation that data sensitivity is multi-layered.

2.3. Measurements and data collection

Perceived risk was measured using a two-component model; the risk was the product of its subjective probability and the perceived magnitude of damage if the risk eventuated (Mitchell, 1999). We took a multi-dimensional view of perceived risk. Although various risks are assumed during personal data use (Solove, 2008), the perceived risks here included only public surveillance, discrimination/prejudice, commercial use, embarrassment, and exposure to criminals. All respondents were asked to evaluate the probability of risk and the extent of possible harm if the risk eventuated. Each subjective probability was scored from 0 to 100 and converted to a 10-point interval scale (e.g. 0–9% was converted to 1). The extent of harm was measured using a six-point scale (1: “does not harm me at all” to 6: “harms me greatly”). The perceived risk score was the extent of harm multiplied by the subjective probability of risk eventuation.

The required protection level was measured using a single-item method employing a six-point scale. The question was: “To what extent do you require data-handling entities (e.g. businesses and public institutions) to rigorously manage the following personal data?” (scores ranged from 1: “no need at all for rigorous management” to 6: “absolutely must be managed rigorously”). The required protection levels were analysed using a six-point scale: “How much legal regulation do you require when the following personal data are handled?" (1: "no need for any legal regulation at all" to 6: "absolutely must be legally regulated"). We evaluated the sensitivities of political orientation, economic/financial status, health status, and consumption. For example, all respondents were asked to rate the subjective probability of public surveillance (and four other risks), the extents of harm if the risks eventuated, and the required protection levels for the four types of data. We compared the means of the perceived risk scores for the four data types. We employed a related-sample Friedman analysis of variance (ANOVA) to determine whether risk levels differed among the four data types. Use of this non-parametric ANOVA is preferable to use of a parametric test because the distributions of the perceived risk scores were extremely distorted, lying far from a normal distribution. The effects of gender, interest, and self-protective behaviour were analysed by comparing the mean scores by gender (male or female), awareness (high or low interest), and self-protective status (high or low). Interest level was assessed based on responses to two questions using a six-point scale: "In daily life, how much do you care about handling of your personal data?" (1: "I do not care about it at all" to 6: "I care very much") and "To what extent are you interested in news and articles on personal data and privacy?" (from 1: "Not interested at all" to 6: "Very interested"). The extent of self-protection was measured based on responses to two questions using a four-point scale: "Have you ever changed the privacy settings of your PC or smartphone?" and "How often do you read privacy policies when downloading software and apps?" (1: "Never" to 4: "Frequently"). We grouped respondents by quantiles in terms of interest and self-protection levels. For example, a respondent whose mean interest score was below the first quantile point was categorised as "low interest" and a respondent whose mean score was above the third quantile point was categorised as "high interest." The Mann-Whitney U-test was employed to explore whether significant differences in mean ranks were evident between pairs of groups. Effect sizes were calculated with the aid of U statistics. The Mann-Whitney test compares non-parametric means; the test power is usually less than that of the t-test, which compares parametric means. However, as noted above, non-parametric tests were more appropriate because the data were not normally distributed.

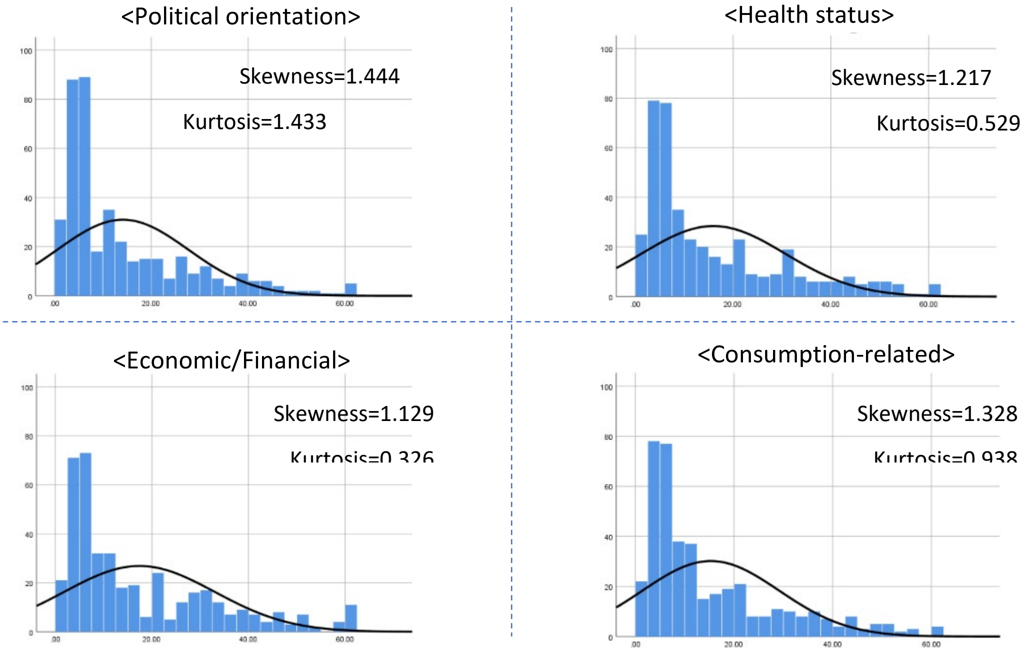
The questionnaire survey was conducted in March 2019. All 420 respondents were Japanese; we enrolled 42 males and 42 females in each of their 20s, 30s, 40s, 50s, and 60s. The questionnaire featured three sections: the first explored respondent attributes and general attitudes about privacy and personal data; the second explored the perceived risks associated with the release of four types of personal data; and the third explored the required protection levels.

3. RESULTS AND DISCUSSION

3.1. Distribution of perceived risk scores

Many previous researches on perceived risk of data release have used a single measure asking respondents about level of risk directly, and the acquired data has been applied to parametric methods such as structural equation modeling (Mitchell, 1999). However, some researchers insisted that a two-component model of risk measurement had several advantages of reliability and validity of over other types of risk measurement models (e.g., Gemünden, 1985; Mitchell, 1999). Therefore, this study adopted the two-component model for risk measurement. The histograms of perceived risk scores, shown in Figure 1, indicated that the distribution of perceived risk scores had a significant positive skew for each data type. Furthermore, the result of the Kolmogorov-Smirnov test showed that the distribution deviated significantly from normal one for each data type [Political orientation: D(420)=0.186, p=0.000; Health status: D(420)=0.179, p=0.000; Economic/Financial: D(420)=0.174, p=0.000; Consumption: D(420)=0.184, p=0.000]. Based on these results, nonparametric methods were used in the following section in order to analyse effects of data type and several personal factors on perceived risk and required protection levels.

Figure 1. Distribution of perceived risk for each data type.



3.2. Differences in risk perceptions and protection requirements among personal data types

Friedman's test revealed significant differences in the perceived risk levels of release of the four types of data [chi-square (3)=30.364, p=0.000]. We performed pairwise comparisons to locate the differences (Figure 2). The perceived risk of economic/financial data release (mean rank=2.75) was the highest and the perceived risk of political orientation data release (mean rank =2.31) was the lowest (corresponding figures for health and consumption data were 2.51 and 2.43). The p-values for each pair (adjusted using the Bonferroni correction for multiple tests) revealed a significant difference at the 5% level in the economic/financial (ranked 1st)-health status (2nd) pair (p=0.043, z=2.686, p=0.131); the economic/financial (1st)-consumption data (3rd) pair (p=0.002, z=3.595, r=0.175); and the economic/financial (1st)-political orientation (4th) pair (p=0.000, z=4.998, r=0.244), as indicated by the thick solid arrows in Figure 2. Thus, release of personal economic/financial data was perceived as significantly riskier than release of any other data, the perceived release risks of which did not differ. We similarly explored differences in protection requirements, which can be classified into two types: data-handling entities must rigorously manage data; and lawmakers must provide legal protection. In terms of data-handling entities, Friedman's test revealed significant differences in handling requirements among the four data types [chi-square (3)=228.205, p=0.000] and also in the requirements for legal protection [chi-square (3)= 203.312, p=0.000]. We performed pairwise comparisons to locate the differences [Figure 3a, data-handling entities; Figure 3b, legal protection].

Figure 2. Pairwise comparison of perceived risks.

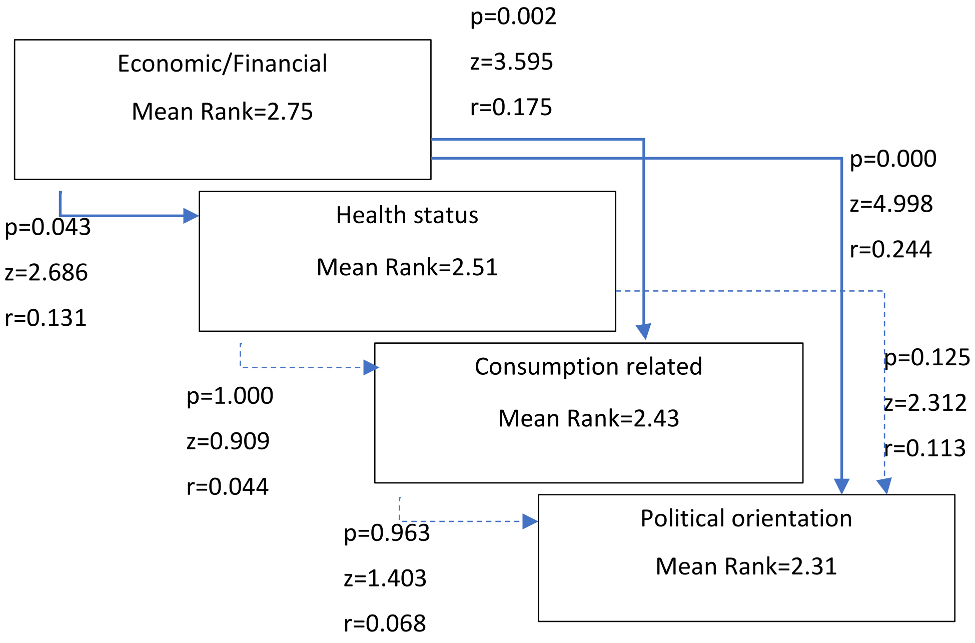
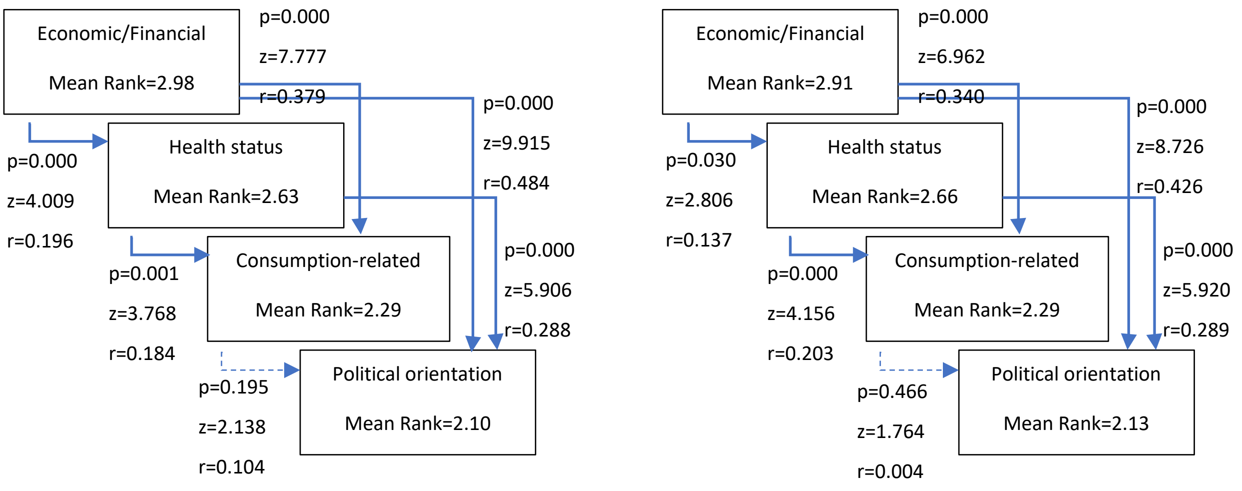


Figure 3. Pairwise comparison of protection requests.

-a. Requests for managerial protection

-b. Requests for legal protection



The results were similar. Economic/financial data required the highest level of protection (mean 2.98 for rigorous management and 2.91 for legal protection), followed by health status (means 2.63 and 2.66), consumption data (means 2.29 and 2.29), and political orientation (means 2.10 and 2.13). The p-values adjusted using the Bonferroni correction revealed that the data pairs differing significantly were identical for both forms of protection. Specifically, the required protection for economic/financial data was significantly higher than that for (second-placed) health data (p=0.000, z=4.009, r=0.196 for managerial protection and p=0.030, z=2.806, r=0.137 for legal protection); and the required protection for health data was significantly higher than that for (third-placed) consumption data (p=0.001 z=3.768, r=0.184 and p=0.000, z=4.156, r=0.203). We found no significant difference in the required protection for consumption data and (fourth-placed) political orientation data (p=0.195, z=2.138, r=0.104 and p=0.466, z=1.764, z=0.004). Thus, for both types of protection, respondents desired stronger protection of economic/financial and health data than consumption and political orientation data; the required protection for economic/financial data was particularly high.

These results suggest that perceived risk and protection requirements vary similarly by data type. Friedman’s test also revealed that these parameters were ordered: respondents were most sensitive to economic/financial personal data release, consistent with the results of a previous study showing that ordinary Japanese people were most sensitive to economic data among 13 personal data categories (Fukuta et al., 2017). The results also suggest that the perceived risks and protection requirements are not unidimensional. In previous studies on privacy and transaction risks, the perceived risks of personal data release encompassed all such data (Glover and Benbasat 2010). However, given the mean ranks and the results of pairwise comparisons, the perceived risk of release of economic/financial data and the required protection clearly differed from those of political orientation release. The effect sizes (r) for this pair were 0.244 (perceived risk), 0.484 (managerial protection), and 0.426 (legal protection). According to Cohen (1992), r=0.1 indicates a small effect, r=0.3 a medium effect, and r=0.5 a large effect. Thus, for both types of protection, the differences verged on large. The levels varied markedly, so unidimensionality was not in play. Finally, the results suggest that the levels of perceived risks and protection requirements were not always consistent with expertise-based analyses. Health and political orientation data are sensitive in the legal sense, but the data indicate that these are considered less important than economic/financial data, which are not legally protected. Moreover, clear differences were evident in terms of protection requirements, with effect sizes of about 0.3 (medium) for both. It remains unclear whether the gaps are caused by differences between evaluations that are expertise-based and those based on ‘feelings’ or by the unexpectedly weak relationship between perceived risk/protection requirements and data sensitivity. It may be necessary to redefine, or develop a new taxonomy of, sensitive personal data to replace the legal definition.

3.3. Effects of personal factors on risk perceptions and protection requirements

Previous research has revealed factors influencing perception of, and behaviours associated with, privacy and personal data disclosure. Barth and Jong (2017) systematically reviewed the privacy paradox and provided a comprehensive list of parameters that affect perceived risk of disclosure. The list includes general privacy concerns, the need for institutional trust, situational characteristics, the affective state, and perceived benefits of disclosure including economic rewards and convenience. Malhotra, Kim, and Agarwal (2004) developed a conceptual model in which several factors served as covariates of risk perception. They assumed that sex, age, and Internet experience confounded the relationships between risk perception, on the one hand, and its antecedents and consequences, on the other. The following discussion explores the effects of gender, interest in personal data, and self-protective tendencies on perceived risks and protection requirements.

1) Effects of gender: It is widely accepted that gender affects risk perception and behaviour associated with personal data disclosure (Gustafson 1998). Most studies have found that females perceive more risks than males (Siegrist 2000). The Knowledgeable Support, Institutional Trust, and Safety Concern Hypotheses have been developed in efforts to explain this difference (Siegrist 2000, Hitchicock 2001). Hypotheses focus on the effects of traditional gender roles (Freudenburg and Davidson 2007). Societal role expectations define 'good' (standard) ways of thinking and behaving, creating gender differences in terms of risk perception. In other words, gender per se may not affect risk perception; gender may interact with sociocultural factors. In general, Japanese society tends to resist changes in social norms, and traditional gender roles remain stronger than in the West. The effects of gender on risk perceptions and the protection requirements for all data types were analysed with the aid of the Mann-Whitney U-test (Table 1). In terms of risk perception, mean female ranks were higher than those of males for all data types, and all differences were significant at the 1% level (political data: U=26818, z=3.834, p=0.000; economic data: U=27424, z=4.321, p=0.000; health data: U=26699.5, z=3.738, p=0.000; consumption data: U=26877, z=3.881, p=0.000). Female respondents perceived greater risks of personal data disclosure than males. Effect sizes were all about 0.2, so gender explained 4% (the effect size squared) of the total dependent variable variance (the rank order of risk); these effects can be considered small to medium. However, the effects of gender on protection requirements were mixed. Table 1 shows that the mean ranks of female groups were all higher than those of male groups. However, no significant gender effect at the 5% level was evident for three of the eight pairs: political-managerial (U=22311.5, z=0.216, p=0.829), political-legal (U=22513.5, z=0.381, p=0.703), and consumption-legal (U=24233.5, z=1.808, p=0.071). Although the remaining five cases exhibited significant effects, the effect sizes were only about 0.1. As mentioned above, this means that gender explains only about 1% (very little) of the total rank order variance in protection requirements. Therefore, the effects of gender on protection requirements varied, but even when significant, they were small.

Table 1. Result of Mann- Whitney's U test: effect of gender.

Perceived risk	Political orientation		Economic/Financial		Health status		Consumption-related	
	Male	Female	Male	Female	Male	Female	Male	Female
Mean Rank	187.80	233.20	184.91	236.09	188.36	232.64	187.51	233.49
Sample size	210	210	210	210	210	210	210	210
Mann-Whitney U	26818		27424		26699.5		26877	
St'd Test Statistic	3.834		4.321		3.738		3.881	
Asymptotic Sig. (2-sided)	0.000		0.000		0.000		0.000	
Effect Size	0.18		0.211		0.182		0.189	
Request for managerial protection	Political orientation		Economic/Financial		Health status		Consumption-related	
	Male	Female	Male	Female	Male	Female	Male	Female
Mean Rank	209.25	211.75	194.11	226.89	199.06	221.94	197.79	223.21
Sample size	210	210	210	210	210	210	210	210
Mann-Whitney U	22311.5		25491		24451.5		24720	
St'd Test Statistic	0.216		3.026		2.017		2.215	
Asymptotic Sig. (2-sided)	n.s. (0.829)		0.002		0.044		0.027	
Effect Size	0.011		0.148		0.098		0.108	

Request for legal protection	Political orientation		Economic/Financial		Health status		Consumption-related	
	Male	Female	Male	Female	Male	Female	Male	Female
Mean Rank	208.29	212.71	194.08	226.92	196.81	224.19	200.10	220.90
Sample size	210	210	210	210	210	210	210	210
Mann-Whitney U	22513.5		25497.5		24924.5		24233.5	
St'd Test Statistic	0.381		2.959		2.409		1.808	
Asymptotic Sig. (2-sided)	n.s. (0.703)		0.003		0.016		n.s. (0.071)	
Effect Size	0.019		0.144		0.118		0.088	

2) Effects of interest in personal data handling: The level of interest in what personal data are collected and how data are handled varies, and differences in interest levels critically affect information processing. In a broad sense, it has been widely accepted that the level of interest in a thing determines the attention paid to, and the intention to learn about, the thing (Klapper 1960). According to one information processing model (the Bettman Model) of consumer behavioural research, a high level of interest incentivises integration and memorisation of relevant internal and external information (Peter and Olson, 2010). The Elaboration Likelihood Model has been used in research about communication and advertising; it suggests that people who have a high interest in, and considerable knowledge of, a certain object, tend to process information principally via a central route that imposes a large cognitive burden (e.g. the need to understand text in an advertisement) (Cacioppo et al., 1986). Together, these models suggest that highly motivated central information processing, triggered by a high degree of interest in personal data, may establish a lifelong belief in the negative outcomes of personal data disclosure. Therefore, the level of interest positively influences perceived risk and protection requirements.

We used the Mann-Whitney U-test to explore the effects of interest on perceived risk of personal data disclosure and protection requirements. Table 2 shows that the mean rank of the high interest group exceeded that of the low interest group in all 12 pairs. Furthermore, the p-values for all pairs indicated that interest significantly affected risk perceptions on disclosure of, and protection requirements for, each type of data at the 1% level. Thus, the level of interest positively affects risk perception and protection requirements, regardless of the type of personal data. Effect sizes ranged from 0.3–0.4. Based on the Cohen estimation, the effects of interest on risk perception and protection requirements were medium or greater; the level of interest thus explained 10–15% of the total variance in perceived risk and protection requirements. The effect sizes of interest clearly exceeded those of gender in all 12 cells, particularly in terms of protection requirements.

Table 2. Result of Mann-Whitney's U test: effect of interest.

Perceived risk	Political orientation		Economic/ Financial		Health status		Consumption-related	
	High	Low	High	Low	High	Low	High	Low
Mean Rank	152.09	102.50	152.63	101.77	150.01	105.31	151.74	102.97
Sample size	150	111	150	111	150	111	150	111
Mann-Whitney U	11488		11570		11176.5		11436	
St'd Test Statistic	5.247		5.383		4.730		5.160	
Asymptotic Sig. (2-sided)	0.000		0.000		0.000		0.000	
Effect Size	0.325		0.333		0.293		0.319	

Request for managerial protection	Political orientation		Economic/ Financial		Health status		Consumption-related	
	High	Low	High	Low	High	Low	High	Low
Mean Rank	150.12	105.16	154.36	99.44	153.28	100.89	151.96	102.68
Sample size	150	111	150	111	150	111	150	111
Mann-Whitney U	11193.5		11828.5		11667.5		11468.5	
St'd Test Statistic	4.891		6.379		5.812		5.392	
Asymptotic Sig. (2-sided)	0.000		0.000		0.000		0.000	
Effect Size	0.303		0.395		0.360		0.334	
Request for legal protection	Political orientation		Economic/ Financial		Health status		Consumption-related	
	High	Low	High	Low	High	Low	High	Low
Mean Rank	150.84	104.19	155.26	98.22	155.38	98.06	154.56	99.16
Sample size	150	111	150	111	150	111	150	111
Mann-Whitney U	11301		11963.5		11981.5		11859.5	
St'd Test Statistic	5.057		6.506		6.316		6.031	
Asymptotic Sig. (2-sided)	0.000		0.000		0.000		0.000	
Effect Size	0.313		0.403		0.391		0.373	

3) Effects of self-protective behaviour: Data subjects can protect their own data independent of data-handling entities and governments. For example, smartphone network settings can be changed to self-protect data. Reading the privacy policies of products and services is another form of self-protection. The need to protect personal data triggers both self-protection and protection requests to other entities, because the perceived risk of personal data use is high. Thus, self-protection might be positively related to both perceived risk and protection requirements. Conversely, self-protection might complement or substitute for protection by other entities, reducing the perceived risk. Under such circumstances, self- protection would exhibit negative relationships with both perceived risks and the need for protection by others. The Mann-Whitney U-test yielded mixed results (Table 3). The mean rank of the high self-protection group was greater than that of the low self-protection group for all 12 pairs. However, for some pairs, no significant effect of self-protection at the 5% level was evident. Specifically, there was no significant effect on the perceived risk of health data release (U=9629.5, z=1.881, p=0.060) or legal protection of data on political orientation (U=9599, z=1.876, p=0.061). The paired effect sizes varied widely from about 0.1–0.3, and many were below 0.2. Compared to the effects of interest, the effects of self-protection on perceived risk and protection requirements were generally low.

Table 3. Result of Mann- Whitney's U test: effect of self-protection.

Perceived risk	Political orientation		Economic/Financial		Health status		Consumption-related	
	High	Low	High	Low	High	Low	High	Low
Mean Rank	143.84	116.21	140.23	120.68	139.41	121.70	143.22	116.97
Sample size	145	117	145	117	145	117	145	117
Mann-Whitney U	10271.5		9749		9629.5		10182.5	
St'd Test Statistic	2.934		2.077		1.881		2.788	
Asymptotic Sig. (2-sided)	0.003		0.038		n.s. (0.060)		0.005	
Effect Size	0.181		0.128		0.116		0.172	

Request for managerial protection	Political orientation		Economic/Financial		Health status		Consumption-related	
	High	Low	High	Low	High	Low	High	Low
Mean Rank	143.36	116.80	140.80	119.97	142.99	117.26	150.97	107.37
Sample size	145	117	145	117	145	117	145	117
Mann-Whitney U	10202.5		9831.5		10148		11306	
St'd Test Statistic	2.896		2.43		2.869		4.793	
Asymptotic Sig. (2-sided)	0.004		0.015		0.004		0.000	
Effect Size	0.179		0.150		0.177		0.296	

Request for legal protection	Political orientation		Economic/Financial		Health status		Consumption-related	
	High	Low	High	Low	High	Low	High	Low
Mean Rank	139.20	121.96	144.34	115.59	146.33	113.12	148.72	110.16
Sample size	145	117	145	117	145	117	145	117
Mann-Whitney U	9599		10344.5		10633		10979.5	
St'd Test Statistic	1.876		3.292		3.686		4.218	
Asymptotic Sig. (2-sided)	n.s. (0.061)		0.001		0.000		0.000	
Effect Size	0.116		0.203		0.228		0.261	

4) The effects of personal factors: Of the three personal factors examined, the effects of interest were the strongest and the most stable. Statistically significant (positive) effects were evident for all 12 pairs, and classified as medium (over 0.3) in 11. The higher the extent of interest in personal data, the higher the level of perceived risk and need for managerial and legal protection, regardless of type of data. The effects of gender were more complicated. In terms of effects on perceived risk, statistically significant effects were evident for all data types. Females tended to perceive higher risks than male. However, the effects on protection requirements were mixed: sometimes significant and sometimes not. The effect size of gender was less than that of interest; the reason for this is unclear, so more research is needed. We measured only direct effects of gender, and as noted above, it may be useful to examine the interactions between gender and personal and sociocultural factors. Similar mixed results and relatively low effect sizes were also observed for self-protective tendencies. If self-protection is viewed as a need for personal data protection, self-protection would be expected to co-vary with perceived risks and protection requirements. Conversely, if self-protection has a complementary or other relationship with data protection performed by other entities, self-protection would be negatively related to perceived risks and protection requirements. Given the small positive direct effect, the former hypothesis may be more suitable, but both hypotheses may be correct; the small effect size may reflect offsetting of positive and negative effects.

4. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

Data sensitivity is the extent to which data subjects do not want anyone to know or use their personal data because of a perceived risk of negative consequences. Such unwillingness is reflected in data protection requirements. It is essential to understand the characteristics of risk perception and protection requirements; these are the essence of data sensitivity. We quantitatively evaluated these features by focusing on four types of personal data. Comparisons of the perceived risks and protection requirements revealed that these differed by data type. Expertise-based data sensitivity does not explain such variation. Thus, a personal data taxonomy with a definition of sensitive data that differs from the legal definition is required if such sensitivity is to be comprehensively understood. We examined the effects of several personal factors on perceived risk and protection requirements. We found that the level of interest in personal data exerted significant positive effects, but the effects of gender and self-protection were vaguer and weaker. In terms of gender, interaction with sociocultural factors may be in play. More study of complementary or other relationships between protection requirements and self-protective behaviour may be needed to clarify why the effects of self-protective behaviour were so weak.

We sought to clarify the characteristics of the perceived risks of personal data use and the associated protection requirements. This is the first step toward a comprehensive understanding of data sensitivity. Several more steps are required. First, we plan to complement the present work using both qualitative data and data mining. Although our quantitative results demonstrated that perceived risks differed significantly by the extent of a subject's interest in his/her personal data, additional qualitative differences may also be in play between high and low interest groups. Second, the relationships among perceived risk, protection requirements, and other aspects of data sensitivity require more attention. If a triadic relationship is in play among perceived risk, protection requirements, and unwillingness to expose personal data, clarification of this would greatly aid a comprehensive understanding of data sensitivity. Finally, data sensitivity varies socio-culturally, so international comparisons are required. Expertise-based definitions of data sensitivity (such as those employed in privacy laws) have been compared among nations, but data sensitivities among ordinary people must also be compared to appropriately balance globalisation and localisation. Overall, a comprehensive understanding of data sensitivity is required to develop appropriate data management and privacy protection systems.

ACKNOWLEDGEMENTS

This study was supported by the JSPS KAKENHI Grant-in-Aid Scientific Research (C) 17K03879 and 17K04013, and the Kurata Grants subsidised by the Hitachi Global Foundation.

REFERENCES

Ackerman, M. S., Cranor, L. F., & Reagle, J. (1999). Privacy in e-commerce: Examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM Conference on Electronic Commerce*,1–8.

Al-Fedaghi, S. (2007). How sensitive is your personal information? In *Proceedings of the 2007 ACM Aymposium on Applied Computing,* 165–169.

Barth, S., & De Jong, M. D. (2017). The privacy paradox–Investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review. *Telematics and informatics*, *34*(7), 1038-1058.

Cacioppo, J. T., Petty, R. E., Kao, C. F., & Rodriguez, R. (1986). Central and peripheral routes to persuasion: An individual difference perspective. *Journal of Personality and Social Psychology*, 51(5), 1032–1043.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.

European Commission (2012). Proposal for a regulation of the European parliament and of the Council- on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). 11 April 2012. Retrieved from https://ec.europa.eu/transparency/regdoc/rep/1/2012/EN/1-2012-11-EN-4-1.PDF

Freudenburg, W.R. and Davidson D.J. (2007). Nuclear families and nuclear risks: The effects of gender, geography, and progeny on attitudes toward a nuclear waste facility. *Rural Sociology*, 72(2), 215-243.

Fukuta, Y., Murata,K., Adams, A.A., Orito, Y., & Palma, A. M. L. (2017). Personal data sensitivity in Japan: An exploratory study. *ORBIT Journal*, 1(2), Retrieved from https://doi.org/10.29297/orbit.v1i2.40

Fule, P., & Roddick, J. F. (2004). Detecting privacy and ethical sensitivity in data mining results. In *Proceedings of the 27th Australasian Conference on Computer Science-Volume 26*, 159–166.

Gemünden, H. G. (1985). Perceived risk and information search. A systematic meta-analysis of the empirical evidence. *International Journal of Research in Marketing*, *2*(2), 79-100.

Glover, S., & Benbasat, I. (2010). A comprehensive model of perceived risk of e-commerce transactions. *International Journal of Electronic Commerce*, 15(2), 47–78.

Gustafsod, P. E. (1998). Gender Differences in risk perception: Theoretical and methodological erspectives. *Risk analysis*, *18*(6), 805-811.

Hitchcock, J. L. (2001). Gender differences in risk perception: broadening the contexts. *Risk*, *12*, 179-204.

Japan APPI (2015). *Act on the Protection of Personal Information* (revised in 2015) Retrieved from http://www.japaneselawtranslation.go.jp/law/detail/?id=2781&vm=&re=

Klapper, J.T. (1960), *Effects of Mass Communication*, Free Press.

Malheiros, M., Preibusch, S., & Sasse, M. A. (2013). “Fairly truthful”: The impact of perceived effort, fairness, relevance, and sensitivity on personal data disclosure. In *International Conference on Trust and Trustworthy Computing*, 250–266.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information systems research*, *15*(4), 336-355.

Mitchell, V. W. (1999). Consumer perceived risk: Conceptualisations and models. *European Journal of Marketing*. 33(1/2), 163–195.

Peter, J.P. and J.C. Olson (2009), *Consumer Behavior and Marketing Strategy*, McGraw-Hill Higher Education.

PPC Japan (2016). *The Proceedings of the 19th Personal Information Protection Commission of Japan*. 30 September 2016. Retrieved from https://www.ppc.go.jp/files/pdf/280930_giziroku.pdf

Sapuppo, A. (2012). Privacy analysis in mobile social networks: The influential factors for disclosure of personal data. *IJWMC*, *5*(4), 315–326.

Siegrist, M. (2000). The influence of trust and perceptions of risks and benefits on the acceptance of gene technology. *Risk analysis*, *20*(2), 195-204.

Solove, D.J. (2008). *Understanding Privacy*, Harvard University Press.

Turn, R., & Ware, W. H. (1976). Privacy and security issues in information systems. *IEEE Transactions on Computers*, (12), 1353–1361.