THE VALUE OF PERSONAL INFORMATION ONLINE: RESULTS FROM THREE STATED PREFERENCE DISCRETE CHOICE EXPERIMENTS IN THE UK

Dimitris Potoglou[1,2], Sunil Patil[2], Covadonga Gijón[3], Juan Palacios[3], Claudio Feijóo[3]

[1]Cardiff School of Planning and Geography, Wales, United Kingdom

[2]RAND Europe, Cambridge, United Kingdom

[3]CeDInt - Technical University of Madrid, Pozuelo de Alarcón, Spain

Abstract

This paper proposes the application of a widely used approach, known as stated preference discrete choice experiments, to estimate the value of personal information in three real-life contexts and situations. The paper develops three experiments describing hypothetical situations in which respondents considered varying aspects of their personal information (e.g. storage, sharing with third parties) when (a) purchasing online a product, (b) a service or (c) conducting pure search online. The survey was carried out with sample quotas pre-specified in order to match the profile of the Internet-user population in the UK with respect to gender, age group, geographical area of residence and personal annual income. The results from the experiment provide new insights in the value and influence of attributes of personal information when conducting online transactions. In particular, main results show that there was little interest by respondents to pay in order to introduce control over their personal data, that the extend of sharing of personal information with third parties was seen the most important aspect when choosing online retailers and search engines, and that an unspecified duration of data storage was received as badly as the data storage beyond several years for online retailers and worse than shorter durations.

Keywords: *personal information, privacy, online purchase, insurance, pure search, stated preference, discrete choice experiments, UK, search engines*

Presented at the 21st European Conference for Information Systems, June 5-8, 2013, Utrecht

1 Introduction

All types of electronic media are increasingly interconnecting people among them and with both the virtual and physical worlds. While we exchange and access information using these systems, data records are collected on who we are, where we are, what we do, and how we do it. With data storage capacity increasing and becoming more affordable, computational power increasing geometrically and improved broadband penetration and affordability, the collection and analysis of these data is opening a wealth of innovations related with personalised services and applications. In fact, while companies have always collected customer data and used them to create value, this is now realised in a larger scale and much cheaper and faster than ever before. However, while personalization of online services provide value to customers -an Internet report (Bughin, 2011) estimated this value in US and selected EU countries at €100 billion for 2010-, there are also demonstrable users' concerns about possible privacy abuse of their personal data (Cooper, 2008) as well as annoyance with the advertising interruptions (Spaulding, 2010). Indeed, with commercial and technical developments in this area relatively fragmented, more research on the economics of personal information is needed in spite of initial works by the OECD and the WEF (WEF, 2011). In particular, policymakers face considerable challenges when attempting to regulate personal data in online markets; not only are the markets complex with many new emerging stakeholders and services, but the challenges multiply as the data flows increase and as the collection of personal information in business-to-consumer transactions and the respect of consumers' preferences are two fundamentally competing goals. In addition, consumers are not aware in general of the further usages of their personal information beyond their immediate service provider and they would need to be better informed of likely market initiatives. They also feel threatened by the unbounded use of personal information by third parties irrespective of contextual integrity (Nissenbaum, 2010). Moreover, society as a whole needs information on whether or not industry is gaining from the existing information asymmetry or in what business models can they rely to achieve improved protection and/or satisfaction.

Precisely, this paper examines what is the economic value of the personal information component in different transactions and use cases based on an experimental design. It delimits analysis to ecommerce sites, recommender systems and search engines. These sectors have been chosen since they comprise a relevant part of the daily online activities of users, and they are based on well-defined transactions where personal information is exchanged. Experimental design is needed as there is no direct market evidence on how individual consumers respond to nuances in personal information usage by providers. In particular, stated preference (SP) methods allow examination of such hypothetical situations to compensate for the absence of real market behaviour.

Following this same rationale, in recent years several experimental studies have been conducted attempting to quantify individual valuations of personal data in diverse contexts. Initial research was mainly aimed at analysing privacy issues. For instance Hann et al (2002) implemented a ranking-conjoint experiment on different websites to attach a monetary value to privacy issues, such as mistakes on personal information treatment, improper access or secondary use of information, concluding that providers need to offer substantial monetary incentives to overcome individual concerns. Next strand of research focused on behavioural patterns regarding the type of information disclosed as well as the environment where the transaction took place. For example Huberman et al (2005) used reverse second-price auctions for personal data on age and weight. These authors concluded that the willingness to accept was related to self-perception factors, in particular individuals closer to the average were more inclined to reveal personal information than individuals who perceived themselves to be far from the average. A similar approach was followed by Danezis et al (2005) on location, concluding that respondents tended to consider more valuable their data for commercial than for academic usage. Cvrcek et al (2006) found that extending storage of location from one month to a year caused a twofold increase in the median bids. These initial experiments provided relevant hints at factors influencing user perspective but only looked into partial aspects of personal information from a privacy perspective, and did not follow any utility theory to arrive at economic valuations. The other main type of existing practical research on the valuation of personal information is based on laboratory settings where personal information is a key part of an economic transaction on a real good or service. Two relevant examples were carried out by Jentzsch et al (2012) and Beresford et al (2010). Both works propose respondents to choose between two online retailers with different approaches to personal information, and both reach similar conclusions about consumers willing to pay to the “privacy-friendly retailer”. These experimental settings provide further insights into the processes and motivations embedded in the valuation of personal information, but lack a comprehensive perspective on all the attributes –and their valuation- attached to transactions linked with personal information.

Departing from this previous literature, this paper aims at widening the scope of existing results on the current status of the perceived value in the use of consumers’ personal information in online transactions, establishing the specific influence of individual attributes in the valuation of personal information. For this, the experiment described in the paper covers three frequent and relevant usage scenarios, a broader and more granular number of attributes than previous works, and uses a representative sample of Internet users in the UK to reach conclusions as general and valid as possible. Although the experiment includes information on these variables, correlation with online behaviour and influence of socio-demographics were postponed for further research.

The paper is organized as follows. After the background information and brief review of this section, the next section describes in detail the methodology of the stated-preference-discrete-choice experiment used in the survey. Section 3 explains the design of the experiment, and section 4 the survey implementation and the preliminary data analysis. From there, the econometric analysis and some of the main results are presented. The paper closes with the discussion of results.

2 The stated-preference-discrete-choice-experiment methodology

The stated-preference-discrete-choice-experiment (SPDCE) is a multi-attribute survey-based approach for eliciting consumer's choices for non-market goods, services or situations in a hypothetical setting (Louviere et al. 2000). Their main purpose of conducting is to identify the independent influence of attributes in the choices made by a sample of survey participants and their valuation of these attributes.

The attractiveness of the SPDCE method lies in its capacity to account for multi-attribute issues, explore non-existing alternatives, and largely avoid the problem of multi-collinearity, a common issue when modelling observed (actual) individual behaviour (Hensher et al. 2005). Throughout almost 30 years of research, the SPDCE approach has found wide applicability in variety of subject areas including transport (e.g. Iraguen and Ortuzar 2004), environmental valuation (e.g. Birol et al. 2006), healthcare (e.g. Ryan et al. 2001) and marketing (e.g. Allenby et al. 2004). SPDCE involves presenting respondents with sets of two or more hypothetical alternatives and asking them to choose the one they would prefer the most. The different alternatives in a choice situation are defined as 'packages' comprised of a set of relevant attributes (characteristics) constructed by researchers in a preparatory design stage of the survey. Attributes take a range of values (levels) to form these alternatives. Qualitative analysis including literature reviews, focus-groups and cognitive testing, is particularly appropriate in defining the relevant attributes and levels to be used in the experiment (Kløjgaard et al. 2012). The combinations of attribute levels to form the sets of alternative options are constructed using principles of statistical experimental design, including optimal and efficient designs (Hensher et al. 2005, Bliemer and Rose 2009, Huber and Swerina 1996).

Using choice-based experiments ('pick-one' task) allows the analyst to both design the experiments (if efficient designs are used)[1] and conduct subsequent analysis using discrete-choice analysis which is grounded on a rigorous theory, the Random Utility Theory (RUT) (Louviere and Woodworth 1983). Under RUT, for each alternative-option i an individual n assigns a utility U_{in}, which contains an observable (deterministic) part V_{in} and a random (unobservable) part ε_{in} (McFadden 1974):

[1] Recent advancements in the design of SPDCE recommend the generation of alternatives using efficiency criteria (reduction of the asymptotic variance-covariance standard errors) rather than orthogonality across the attributes of the alternatives. Efficient designs will generally results in designs that either improve the reliability of the parameters estimated from SPDCE data at a fixed sample size or reduce the sample size required to produce a fixed level of reliability in the parameter estimates with a given experimental design (Huber and Zwerina, 1996; adapted from Bliemer and Rose, 2009).

$$U_{in} = V_{in} + \varepsilon_{in} = \sum_i \beta_i X_{in} + \sum_p \gamma_p Z_{pn} + \varepsilon_{in} \quad [1]$$

The observable part of the utility V_{in} is a linear-in parameters function of attribute levels (characteristics) (X_{in}) describing the alternative and individual characteristics (Z_{pn}), and β_i and γ_p are coefficient estimates for each attribute level X and coefficients representing the (potential) influence of personal characteristics in the choosing alternative i, respectively.

Under RUT, it is assumed that a respondent n will consider the available option described by attribute levels X and will choose the alternative with the highest utility. Given that the above formulation of utility includes a stochastic component, it is only possible to describe the probability of choosing alternative i over another alternative k as:

$$Prob\ (i\ is\ chosen) = prob\ \{V_i + \varepsilon_i > V_k + \varepsilon_k;\ \forall k \in C\} = prob\{V_i - V_k > \varepsilon_k - \varepsilon_i) \quad [2]$$

where C is the set of all possible alternatives. Assuming a type I extreme value distribution for the error terms and independence between the alternative options, the probability of choosing alternative i takes the form of the conditional logit model (McFadden 1974)[2]:

$$Prob\{i\ is\ chosen\} = \frac{\exp(\mu V_i)}{\sum_{j \in C}(exp\mu V_j)} \quad [3]$$

where μ is the scale parameter, which for any single sample is assumed to be equal to 1.

Collecting the choices of survey respondents across the different sets of alternatives allow the estimation of β and γ parameters and the estimation of the probability that alternative i will be chosen among the set of alternatives presented to the respondents. Furthermore, results can be used to derive estimates of consumers' valuation for different aspects of a non-market good or service – i.e., the amount of money they are willing to pay (or willing to accept) to obtain some benefit (or avoid some cost or situation) from a specific action (Louviere et al. 2000).

The above theoretical framework and prior empirical evidence to support the use of SPDCE for elicitation of choices over a set of alternatives composed of different levels is regarded as promising and appropriate approach in understanding individuals' valuations for their personal information and a contribution to the literature in this field. This study is aimed at testing this assertion by developing three discrete choice experiments as described in the following sections.

3 Design of SPDCE to estimate the value of personal information

This study focused on three hypothetical scenarios in which respondents' valuation for their personal information was examined: purchase of a product online (Experiment 1), (b) purchase of a service online (Experiment 2) and (c) conducting pure search using a search engine (Experiment 3).

The design of the SPDCE questionnaire followed three stages (Bliemer and Rose 2009): (1) qualitative research, (2) model specification and (3) experimental design. As part of the first stage, we conducted a literature review (see previous section) and consulted with experts in order to define the choice context, the attributes and attribute levels that would describe the scenarios. The attributes and levels used to describe the alternative options in each of the experiments are listed in Tables 1 and 2.

Attribute	Levels
Cost per transaction against security costs	(1) Discount £4.00 (2) Discount £2.00 (3) No charge (4) £2.00 (5) £4.00

[2] Different assumptions about the distribution of the error terms give rise to different modelling structures (e.g. probit, mixed logit)

Additional information saved and linked to your account	(1) Only email (2) Purchase history and email (3) Purchase history, browsing and navigation history and email (4) Purchase history, browsing, navigation history, email and additional personal details
Permission of sharing this additional information with third parties	(1) No (2) Yes
Time your personal information is stored for	(1) 1 year (2) 2 years (3) 5 years (4) Without an explicit temporal limit
Availability of product or service at a conventional store/outlet (Only available in the Conventional store/outlet alternative)	(1) This item can also be easily purchased in your neighbourhood at a conventional retailer (2) This item can also be purchased from a conventional retailer, but it would require from you to make a special effort because of day/hour of purchase, distance to reach the merchant, etc.) (3) This item is not available to purchase from a conventional retailer in your neighbourhood
Additional services offered by the product provider	*Product scenario only (Experiment 1)* (1) None (2) Faster checkout (one-click order) (3) Detailed reviews of products/seller (4) Priority shipping of product at the same price
Additional services offered by the service provider	*Insurance and service scenario only (Experiment 2)* (1) None (2) Faster checkout (one-click order) (3) Legal advice on the phone (4) Detailed reviews of products/seller

Table 1. Attributes and levels in the purchase of product (Experiment 1) and services (Experiment 2)

In Experiments 1 and 2, respondents were asked to imagine that they were about to repeat a recent online purchase of a product and service, respectively, and were offered a choice of three online retailers with varying levels of requirements, treatment and storage of their personal information. These three options included a cost per transaction, negatively correlated with personal data requirements asked by the retailer. The main objective of this design was to make respondents face situations in which they had to make trade-offs between privacy and costs. To complete the choice set respondents were also presented with the possibility of opting-out the experiment and purchasing the good or service from a conventional retailer. An example of a choice situation is shown in Figure 1.

Attribute	Levels	
Monthly charge of using the search engine account	(1) Free (2) £0.50 (3) £1.00 (4) £1.50 (5) £2.00	
IP address (nearby location) stored?	(1) No (2) Yes	[Yes: present additional benefit = search listings highlight results closer to your area or popular in your area]
Search history stored?	(1) No (2) Yes	[Yes: present additional benefit = search listings highlight results which may be more personalised]
Search history linked to your email or IP address?	(1) No (2) Yes	[Yes: present additional benefit = you may receive promotional offers related to your search]
Duration of storage of search history	(1) Not applicable (2) 1 year (3) 2 years (4) No explicit temporal limit	
Advertisement displayed on the search webpage	(1) No (2) Yes	
Additional features associated with the search	(1) None (2) Search listings highlight results closer to your area or popular in your area (3) Search listings highlight results which maybe more personalised (4) You may receive promotional offers related to your search	
Treatment of personal information related to your	(1) Nothing is shared with third parties [only presented with non-zero monthly charge] (2) Search history and/or IP address are shared with third parties	

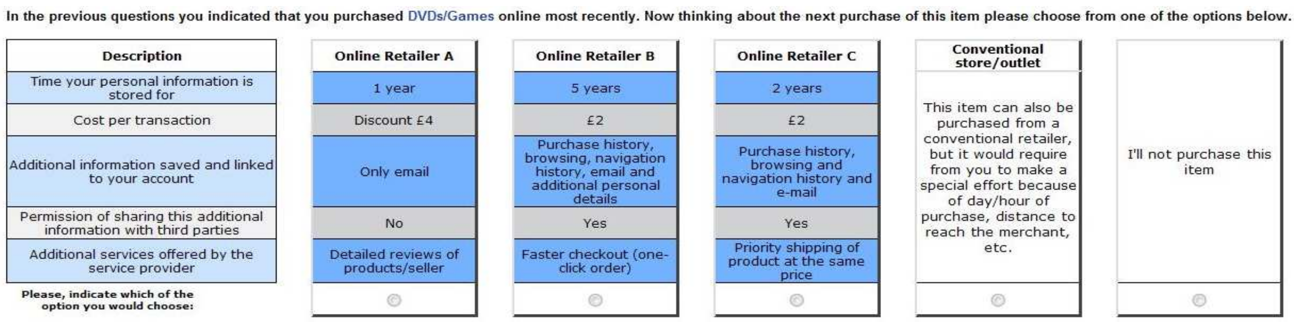
account with the search provider	(3) Email address is shared with third parties (4) Telephone number, and Email address shared with third parties (5) Telephone number, Email address, search history and/ IP address are shared with third parties

Table 2. Attributes and levels of attributes in pure search (Experiment 3)

For Experiment 3, the scheme was similar: respondents were presented with a choice of two search engines with varying levels of requirements, treatment and storage of their personal information. Some of the options involved a monthly charge that would be used against the cost for collection, management, storage and processing of users' personal information so that they could obtain a better experience and targeted service. As in Experiments 1 and 2, respondents could opt-out to select none of the alternatives offered. An example of a choice situation in Experiment 3 is shown in Figure 2.

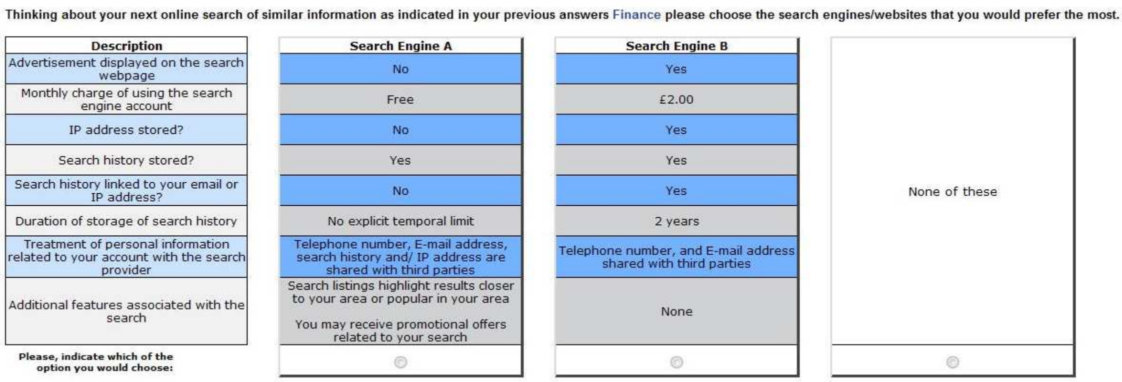
The multinomial choice model (MNL) was selected as the most suitable choice model to describe the choice among different options involving varying levels of attributes. In Experiments 1 and 2, the MNL model consisted of five utility functions, one for each of three online retailers, one for the conventional retailer and one for the opt-out alternative which was set equal to zero. Similarly, in Experiment 3, two observed utility-functions described the choice between two different search engines and one utility, fixed at zero, was specified for the opt-out option.

Based on the specification of the above MNL models, the hypothetical choice situations presented to participants were based on the generation of D-optimal design matrices assuming zero priors for unlabelled alternatives (Bliemer and Rose 2009). The design matrix in all experiments included 60 different choice situations, which were further blocked into 12 blocks so that each respondent was presented with 5 choice situations for each of the three experiments. The experimental design matrices were generated using the software Ngene (Ngene 2010).


In the previous questions you indicated that you purchased DVDs/Games online most recently. Now thinking about the next purchase of this item please choose from one of the options below.

Description	Online Retailer A	Online Retailer B	Online Retailer C	Conventional store/outlet	
Time your personal information is stored for	1 year	5 years	2 years	This item can also be purchased from a conventional retailer, but it would require from you to make a special effort because of day/hour of purchase, distance to reach the merchant, etc.	I'll not purchase this item
Cost per transaction	Discount £4	£2	£2		
Additional information saved and linked to your account	Only email	Purchase history, browsing, navigation history, email and additional personal details	Purchase history, browsing and navigation history and e-mail		
Permission of sharing this additional information with third parties	No	Yes	Yes		
Additional services offered by the service provider	Detailed reviews of products/seller	Faster checkout (one-click order)	Priority shipping of product at the same price		
Please, indicate which of the option you would choose:					

Figure 1. An example of a choice situation in Experiment 1


Thinking about your next online search of similar information as indicated in your previous answers Finance please choose the search engines/websites that you would prefer the most.

Description	Search Engine A	Search Engine B	
Advertisement displayed on the search webpage	No	Yes	None of these
Monthly charge of using the search engine account	Free	£2.00	
IP address stored?	No	Yes	
Search history stored?	Yes	Yes	
Search history linked to your email or IP address?	No	Yes	
Duration of storage of search history	No explicit temporal limit	2 years	
Treatment of personal information related to your account with the search provider	Telephone number, E-mail address, search history and/ IP address are shared with third parties	Telephone number, and E-mail address shared with third parties	
Additional features associated with the search	Search listings highlight results closer to your area or popular in your area You may receive promotional offers related to your search	None	
Please, indicate which of the option you would choose:			

Figure 2. An example of a choice situation for pure search (Experiment 3)

4 Survey implementation and preliminary data analysis

The data collection was conducted with participants who were registered with the Internet Panel of 'Research Now' (http://www.researchnow.co.uk), a market research agency with the largest panel of Internet users in the UK. The main survey was conducted 8-10 August 2012. Prior its official release, the survey was modified in accordance with post-survey cognitive questions in a testing phase with 31 participants. A total of 517 respondents completed the survey. Descriptive statistics of the sample and comparisons with the Internet-user population in the UK are shown in Table 3.

Sample quotas were pre-specified in order to match the profile of the Internet-user population in the UK with respect to gender, age group, geographical area of residence and personal annual income, which were publicly available (Office for National Statistics 2011). Chi-square tests showed that our sample was representative of the 2001 UK Internet-user population in terms of gender ($\chi^2(1)=1.20$, $p=0.274$), age ($\chi^2(6)=5.33$, $p=0.502$) and geographic region ($\chi^2(11)=9.808$, $p=0.547$). On the other hand, the income-group proportions between our sample and the 2011 UK Internet-user population were significantly different ($\chi^2(11)=47.462$, $p=0.001$), mainly because of the large proportion of Internet-users for whom their annual personal income was unknown (20.9% vs. 9.7% in our survey).

The SPDCE data were first assessed for accuracy and consistency. Respondents who had never bought any product or service online were not shown the corresponding experiments for product and service purchase respectively. This could create a bias in spite of the representativeness of the sample as it is possible -for excluded users- that disclosing personal information was one of the reasons for not using these services online. If this were the case, the results of the experiment would show lower values of each attribute level. Also, respondents who were not able to make comparisons between the choices in the experiments were excluded from further analysis. Finally, respondents who consistently chose the same retailer – i.e., always retailer A, B or C – were excluded from further analysis as non-traders (Hess et al. 2010). Table 4 shows the number of participants whose choices were analysed.

Variable	Sample (%)	Internet users in UK (2011 Q4, %)	Variable	Sample (%)	Internet users in UK (2011 Q4, %)
Gender (female)	52.0	49.6	*Region*		
Age group			East of England	10.1	7.2
18-24	13.9	17.1	East Midlands	7.2	9.5
25-34	21.5	19.6	London	12.8	13.3
35-44	19.3	19.5	North East	3.7	4.0
45-54	18.4	18.8	North West	11.6	11.0
55-64	15.9	14.0	Northern Ireland	2.3	2.5
65-74	7.9	7.9	Scotland	8.5	8.3
75 and over	3.1	3.2	South East	13.7	14.1
			South West	9.3	8.7
Annual individual income			Wales	4.5	4.7
Less than £10,399	27.8	20.9	West Midlands	8.3	8.3
£10,400 - £15,599	14.1	15.2	Yorkshire / Humberside	8.1	8.4
£15,600 - £20,799	12.6	15.9			
£20,800 - £25,999	9.3	12.9	*Occupational status*		
£26,000 - £31,199	6.6	10.4	Working full time	41.0	
£31,200 - £36,399	6.6	7.3	Working part time	17.2	
£36,400 - £41,599	4.1	4.6	Student	7.2	
£41,600 - £46,799	2.5	3.8	Retired	16.1	
£46,800 - £51,999	2.7	2.7	Not in paid work because of long term illness or disability	7.0	
£52,000 - £77,999	2.9	4.1	Seeking work	5.8	
£78,000 - £103,999	1.2	1.8	Other	5.8	
£104,000 or higher	0.0	0.3			
Not reported	9.7	20.9			

Table 3. Sample characteristics vs. the 2011 UK online-user population

Question	Experiment 1	Experiment 2	Experiment 3
Number of participants who had never not bought any product or service on the Internet	15	69	0
Number of participants not able to make comparisons in the experiment	42	44	43
Non traders (participants who always choose the same retailer/search engine across the 5 choices)	28	37	6

Total number of observations available for modelling	432	367	468

Table 4. Number of respondents excluded from the discrete choice analysis

5 Econometric approach and results

We used error-component-multinomial-logit (mixed logit) models to analyse the SPDCE data in order to account for the correlation between the 5 observations that came from the same respondent in each experiment. The specification of the utility *U* of a participant *j* choosing an online retailer *i* in a choice exercise *t* in Experiments 1 and 2 was as follows:

$$U_{ijt} = constant_i + \beta_1 Cost + \beta_2 Additional\ Inf. + \beta_3 Inf.Sharing + \beta_4 Storage\ Time + \beta_5\ Additional\ Services + \zeta_j + \varepsilon_{ijt} \quad [4]$$

In Experiment 3, the utility of a participant *j* choosing search engine *i* in a choice exercise *t* was as follows:

$$U_{ijt} = const_i + \beta_6 Monthly\ Charge + \beta_7 IP\ address\ storage + \beta_8 Search\ history\ storage + \beta_9 Search\ history - email\ link + \beta_{10}\ Advertisement + \beta_{11} Treatment\ of\ Per.Inf. + \zeta_j + \varepsilon_{ijt} \quad [5]$$

where ζ was the error component following the normal distribution with mean zero and standard deviation σ_ζ, which varied across alternative retailers *i* and respondents *j* and accounted for the correlations between observations obtained from the same respondent. The error component ε followed the Gumbell distribution with mean zero and accounted for differences between respondents *i*, alternatives *j* and choice exercises *t*. The parameters β_1-β_{11} and the constants were estimated using the software BIOGEME (Bierlaire 2003). These models were estimated maximizing the simulated likelihood calculated using 500 MLHS draws for the error components (Hess et al. 2006). All attributes except *Cost* and *Monthly Charge* were dummy coded.

The estimation results in Experiments 1 and 2 and Experiment 3 are presented in Tables 5 and 6, respectively. In Experiments 1 and 2, respondents were less likely to choose an option that involved *storage and linkage of additional information* other than their email address, which was the reference level. As requirements for additional information to be saved and linked to an individual's account increased respondents were increasingly against those options. However, there was no significant difference when additional personal details were stored along with purchase history, browsing, navigation history and email. Also, respondents were not in favour of *sharing their personal information with third parties*. Similarly, they were less likely to choose online retailers who would store respondents' personal information for 5 years or without specifying a temporal limit, relative to the reference level of 1 year. Interestingly, there was no significant difference between storing respondents' personal information for 1 and 2 years. Also, respondents valued equally options which offered storage of personal information for 5 years and options which offered storage of personal information without temporal limit. On the other hand, they were more likely to select online retailers who would offer some additional benefit such as priority shipping, faster checkout or detailed reviews of the product and seller. Finally, respondents were less likely to purchase the product from a conventional vendor or service provider relative to online retailers or vendors located in the respondents' neighbourhood.

Attribute	Experiment 1 Product purchase online Coefficient (t-test)	Experiment 2 Service purchase online Coefficient (t-test)
Cost per transaction against security costs	-0.149 (-12.0)	-0.147 (-9.9)
Additional information saved and linked to your account		
Only email	Reference	
Purchase history and email	-0.250 (-3.1)	-0.350 (-4.1)
Purchase history, browsing, navigation history and email		
Purchase history, browsing, navigation history, additional personal details and email	-0.560 (-6.9)	-0.733 (-8.4)
Permission of sharing this additional information with 3rd parties 1 if Yes; 0 if No)	-0.840 (-9.8)	-1.07 (-10.3)
Time your personal information is stored for		

1year	Reference	
2 years	0.0	0.0
5 years Without an explicit temporal limit	-0.433 (-6.4)	-0.565 (-7.6)
Additional services for product purchase		
None	Reference	
Priority shipping of product at the same price Faster checkout (one-click order) Detailed reviews of products/seller	0.478 (6.1)	N/A
Faster checkout (one-click order) Legal advice on the phone Detailed reviews of products/seller	N/A	0.340 (4.3)
Availability of product or service at a conventional store/outlet		
This item can also be easily purchased in your neighbourhood at a conventional retailer	Reference	
This item can also be purchased from a conventional retailer, but it would require from you to make a special effort (because of day/hour of purchase, distance to reach the merchant, etc.)	-0.692 (-4.0)	-0.897 (-4.7)
Standard deviation σ_ζ	0.817 (13.7)	0.766 (10.7)
No. of observations	2160	1835
No. of individuals	432	367
Log-likelihood, constants only, L(c)	-2924.8	-2308.5
Log-likelihood, constants only, L(final)	-2828.5	-2272.3
Rho-square	0.134	0.152

Table 5. Estimation results in Experiments 1 and 2

Concerning Experiment 3, the estimated parameters show that respondents were more likely to avoid online retailers in which their *IP address would be stored* or their *search history would be stored and linked to their IP address*. The latter was marginally significant. Similarly, they were more likely to choose options where *their information was not shared with third parties*. Among the different levels of personal information, respondents were more sensitive when the information to be shared included their telephone numbers, email address, search history and IP address than situations when email address and search history and/or IP address were presented separately. Given that respondents were not in favour of storage of their location and search history by the internet service provider, the linkage between search history and their email or IP address only had marginal influence on their choice for search engine. Finally, display of advertisements during search was not statistically significant.

Attribute	Experiment 3. Pure search Coefficient (t-test)
Monthly charge of using the search engine account	-1.71 (-10.1)
IP address (nearby location) stored? (1 if Yes, 0 if No)	-0.375 (-2.5)
Search history stored? (1 if Yes, 0 if No)	-0.366 (-2.1)
Search history linked to your email or IP address? (1 if Yes, 0 if No)	-0.325 (-2.0)
Advertisement displayed on the search webpage (1 if Yes, 0 if No)	-0.235 (-1.6)
Treatment of personal information related to your account with the search	
Nothing is shared with third parties	Reference
Email address is shared with third parties Search history and/or IP address are shared with third parties	-1.03 (-5.9)
Telephone number, email address, search history and IP address are shared with third parties	-1.66 (-8.2)
Standard deviation σ_ζ	-1.507 (-20.0)
No. of observations	2340
No. of individuals	468
Log-likelihood, constants only, L(c)	-1873.6
Log-likelihood, constants only, L(final)	-1621.8
Rho-square	0.263

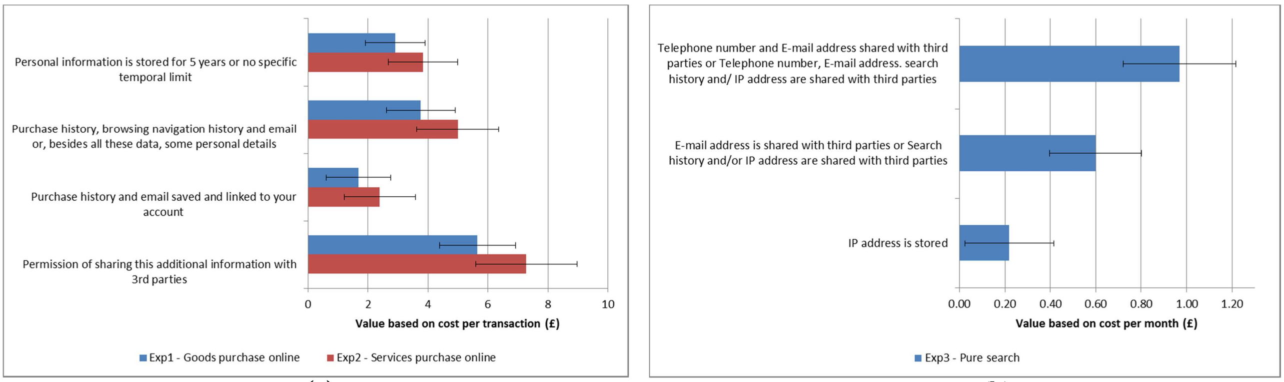
Table 6. Estimation results in Experiment 3

6 Value of personal information

The SPDCE is consistent with utility maximisation and demand theory (Louviere et al. 2000). Once parameter estimates are estimated it is possible to estimate valuations about different attributes, such the willingness to pay[3] (WtP) or the willingness to accept[4] (WtA) for changes in the level of a given attribute (Hensher et al. 2005). In the case of the WtP/WtA[5] regarding personal information, this can be calculated as being equal to:

$$WtP = -\beta_{\cos t}^{-1} \ln\left\{ \frac{\sum_i \exp\left(V_i^1\right)}{\sum_i \exp\left(V_i^0\right)} \right\} \qquad [5]$$

where V_i^0 represents the marginal utility of the base level (e.g. additional information saved and linked to your account: Only email) and V_i^1 represents the marginal utility of another level of the same attribute (e.g. additional information saved and linked to your account: Purchase history and email). β_{cost} is the coefficient of the cost per transaction in Experiments 1 and 2 and the monthly charge for using the search engine in Experiment 3, noted as β_{cost}, gives the marginal utility of price.



(a) (b)

Figure 3: (a) Valuation of personal information when purchasing goods and services and (b) valuation of personal information in pure search experiment and 95% confidence intervals for statistically significant parameter ratios

In a simple linear relationship, each attribute in the utility expression and price are associated with one coefficient each. In that case, equation [5] can be simplified for any individual to the ratio of two utility parameters and provide an estimate of WtP/WtA:

$$WtP = -1\left(\frac{\beta_{purchase\ history\ \&\ email}}{\beta_{price}} \right) \qquad [6]$$

The results of the above computations are presented in Figure 3. On average, respondents placed statistically-significant valuations of their personal information including storage of their information for more than 5 years when purchasing good and services at £2.91 and £3.84 per transaction, respectively. Storage of purchase history for goods and services was valued on average between £1.68 (purchase history and email for product purchase) and £4.99 (purchase and browsing history, email and personal details for service purchase). The highest valuations, £5.65 for product purchase and £7.28 for service purchase, were placed on sharing of personal data with third parties (Figure 3a).

---

[3] Willingness to pay is the maximum amount of money an individual would pay in exchange for getting the good or service object of study.

[4] Willingness to accept is the minimum amount of money an individual would receive in exchange for giving up an endowed object.

[5] In spite of neoclassical economic theory postulating that both measures are identical, there is empirical evidence that shows divergence between WtA and WtP values. In the experiments presented in the paper, there is a composite of both figures as respondents were asked both to pay and to receive discounts. Values obtained are expected to be closer to the value of WtA as this value is usually found to be much higher than WTP. There have been some pieces of research which have tried to find out the sources of this disparity. However, so far there is no consensus among researchers regarding the reasons for this gap. A complete review of WTP/WTA studies can be found in Horowitz & McConnell (2002).

Concerning, the choice of search engine for conducting pure search, respondents valued their IP address at 22p per month, storage of their search history at 21p per month and the linkage between their search history with their geographical location (IP address) and email at 19p month. The highest valuations were for sharing the above information with third parties and ranged between 60p and 97p.

7 Discussion

This paper proposed the application of a widely used approach, known as stated preference discrete choice experiments, to estimate the value of personal information in real-life contexts and situations. The aim of this proposition of to move away for opinion-poll type of questions, which can only offer abstract and frequently vague evidence concernings citizens' importance and valuation of their personal information. In this paper, we developed three discrete choice experiments describing hypothetical situations in which respondents considered varying aspects of their personal information (e.g. storage, sharing with third parties) when purchasing a product, service or conducting pure search online. More than 90% of the participants were able to make comparisons across all three experiments. This finding indicated that the choice tasks undertaken were congitively accessible for the majority of respondents. In particular, in each experiment a number of scenarios were presented to respondents with specific attributes and including a monetary cost attribute for the estimation of individuals' WTP. The values of prices have been kept low to be credible and realistic to minimise the possibility for strategic behaviour. Users could choose among various alternatives and a "choose neither". With the inclusion of this alternative, is it possible to compare more realistically the behaviour of users, confronting the conventional and online worlds and acknowledging that just online options do not explain completely all consumer choices in a real-life situation.

Findings appear to confirm the privacy paradox[6]. On one hand, participants are worried about the use of their data and they certainly value their privacy (see below). On the other, there was little interest by respondents to pay in order to introduce control over their personal data. This finding offers an indication that simple privacy enhanced technogies paid on behalf of consumers might not constitude a viable option, and that a better approach to reconcile user perceptions on the usage of personal information in online transactions is still needed. Admittedly, privacy-enhancing technologies could be welfare enhancing for consumers and society as a whole, although a complete model including this analysis is still missing. The findings in the survey amount to the possibility that privacy-enhancing technologies may lead to non-zero sum market outcomes as it has also started to be explicitly discussed in economic research (Acquisti, 2008). Another avenue for further research from these results would be to educate consumers in how to make intelligent use of the tools within their reach. Having said that, not having options on privacy-enhancement is very different, particularly from a policy perspective, than choosing – judiciously or not[7] – not to exercise them.

The extend of sharing of personal information with third parties was seen the most important aspect when choosing online retailers and search engines. Therefore it is questionnable whether the freemium business models based on this appoach would be viable. It also signals that consumers do differentiate between the bounded use of personal information that takes place within the providers business objectives and the largely unkown usage by third parties. This is an area of current intense policy and commercial debate and these results could contribute to effectively explain that this distintion about usages matters signficantly to consumers. These results follow Nissenbaum (2010), who states that users' concerns originate not from the potential lack of restrictions over the flow of personal information, but from the distress about maintaining the context integrity of personal information while it flows across systems and services.

Finally, an unspecified duration of data storage was received as badly as the data storage beyond 5 years for online retailers and worst than shorter durations. In case of pure search however, the duration of data storage did not matter to users possibly because it can be thought to include less personal information (details of person's address, payment card information etc.). This is an intriguing finding,

[6] Privacy paradox: discrepancy between privacy concerns and actual privacy settings (Barnes, 2006).
[7] An experiment carried out by Acquisti & Grossklags (2005) provided evidence on the difference between individual decision making and rational behaviour. The authors concluded that in some Internet scenarios most individuals are not able to make rational decisions because of lack of information, the so called "bounded rationality" effect.

which might have further implications for policy and with further evidence might reflect a contradictory insight in the right to be forgotten in this context.

References

Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. 3 IEEE Security Privacy Magazine 26–33 IEEE. doi:10.1109/MSP.2005.22

Allenby, G. M., Shively, T. S., Yang, S. and Garratt, M. (2004) A choice model for packaged goods: Dealing with discrete quantities and quantity discounts. Marketing Science, 23(95-108).

Beresford, A. R., Preibusch, S., & Kübler, D. (2010). Unwillingness to Pay for Privacy: A Field Experiment. IZA Discussion Paper No 5017, 117(5017), 25–27. doi:10.1016/j.econlet.2012.04.077

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models, Switzerland.

Birol, E., Karousakis, K. and Koundouri, P. (2006). Using a choice experiment to account for preference heterogeneity in wetland attributes: The case of Cheimaditida wetland in Greece. Ecological Economics, 60, 145-156.

Bliemer, M. and Rose, J. M. (2009). 'Designing Stated Choice Experiments: State of the Art' in Kitamura, R., Yoshii, T. and Yamamoto, T., eds., The Expanding Sphere of Travel Behaviour Research: Selected Papers from the 11th International Conference on Travel Behaviour Research, Bingley, United Kingdom: Emerald Group Publishing Limited, 499-538.

Bughin, J. (2011). The Web's €100 billion surplus. Retrieved from http://www.mckinseyquarterly.com/The_Webs_100_billion_euro_surplus_2724

Cooper, A. (2008). A survey of query log privacy-enhancing techniques from a policy perspective. ACM Trans. Web, 2(4), 1–27. doi:10.1145/1409220.1409222

Cvrcek, D., Kumpost, M., Matyas, V., & Danezis, G. (2006). A study on the value of location privacy. Proceedings of the 5th ACM workshop on Privacy in electronic society WPES 06, 109. doi:10.1145/1179601.1179621

Danezis, G., Lewis, S., & Anderson, R. (2005). How Much is Location Privacy Worth ? Fourth Workshop on the Economics of Information Security, 78, 739–748. doi:10.1111/j.1747-0285.2011.01230.x

Hann, I. H., Hui, K. L., Lee, T. S., & Png, I. P. L. (2002). Online information privacy: Measuring the cost-benefit trade-off. TwentyThird International Conference on Information Systems (pp. 1–10). Citeseer.

Hensher, D. A., Rose, J. M. and Greene, W. H. (2005). Applied Choice Analysis - A Primer, New York: Cambridge Univ..

Hess, S., Rose, J. M. and Polak, J. (2010). Non-trading, lexicographic and inconsistent behaviour in stated choice data. Transportation Research Part D: Transport and Environment, 15(7), 405-417.

Hess, S., Train, K. and Polak, J. (2006). On the use of modified Latin hypercube sampling (MLHS) method in the estimation of mixed logit model for vehicle choice. Transportation Research Part B, 40(2), 147-163.

Huber, J. and Swerina, K. (1996). The importance of utility balance in efficient choice designs. Journal of Marketing Research, 33(3), 307-317.

Huberman, B. A., Adar, E., & Fine, L. R. (2005). Valuating privacy. 3 IEEE Security Privacy Magazine 22–25doi:10.1109/MSP.2005.137

Iraguen, P. and Ortuzar, J. (2004). Willingness-to-pay for reducing fatal accident risk in urban areas: an Internet-based Web page stated preference survey. Accident Analysis and Prevention, 36, 513-524.

Kløjgaard, M. E., Bech, M. and Søgaard, R. (2012). Designing a stated choice experiment: The value of qualitative process. Journal of Choice Modelling, 5(2), 1-18.

Louviere, J. and Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregated data. Journal of Marketing Research, 20, 350-367.

Louviere, J. J., Hensher, D. A. and Swait, J. D. (2000). Stated Choice Methods: Analysis and Application, Cambridge Press.

McFadden, D. (1974). 'Conditional logit analysis of qualitative choice behaviour' in Zerembka, P., ed. Frontiers in Econometrics, New York: Academic Press, 105-142.

Ngene (2010) Ngene 1.0.2 User Manual and Reference Guide: The cutting edge in experimental design, ChoiceMetrics, Sydney.

Nissenbaum, H. (2010). Privacy in context. Technology, policy and the integrity of social life. Stanford: Standford U. Press.

Office for National Statistics (2011) 'Internet Access - Households and inviduals', available: http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-226727 [accessed Nov. 2012].

Ryan, M., Bate, A., Eastmond, C. J. and Ludbrook, A. (2001) Use of discrete choice experiment to elicit preferences. Quality in Health Care, 10(Supplement 1), i55-i60.