The Value of Non-Personally Identifiable Information to Consumers of Online Services: Evidence from a Discrete Choice Experiment

Garrett Glasgow[a,*]
NERA Economic Consulting

Sarah Butler[b]
NERA Economic Consulting

January 7, 2016

ABSTRACT

We estimate the value of non-personally identifying information to consumers of online services through a discrete choice experiment based on hypothetical streaming video services. Non-personally identifying information for online services is typically information on the ways in and times at which customers use the service, and is distinct from personally identifying information such as email addresses or telephone numbers. For most of our survey respondents, we find no evidence that they were willing to pay to avoid sharing their non-personally identifying information with third parties. A smaller group of respondents never selected a service that shared information with third parties.

Running Title: The Value of Non-Personally Identifiable Information
Keywords: privacy; experiment; mixed logit; willingness to pay
JEL classification: C25; C93; D12
Acknowledgments: We thank Kenneth Train for many helpful comments.
Word Count: 1988

[a]Senior Consultant, NERA Economic Consulting, 4 Embarcadero Center, Suite 400, San Francisco, CA 94111. garrett.glasgow@nera.com. Telephone: +1 415 291 1029. Fax: +1 415 291 1020
[*]Corresponding Author
[b]Vice President, NERA Economic Consulting, 4 Embarcadero Center, Suite 400, San Francisco, CA 94111. sarah.butler@nera.com. Telephone: +1 415 291 1022. Fax: +1 415 291 1020

1 Introduction

There is a growing interest in estimating the value that individuals place on their personal information. Most of this literature is framed in terms of privacy, and focuses on the value of *personally identifiable information* (PII), such as email addresses, telephone numbers, and home addresses. However, most online services also collect *non-personally identifiable information* (NPII) from individuals, such as the ways in and times at which individuals use the service. For example, a streaming video service might collect personal information about viewing habits and preferences, such as which movies a customer has watched and what times a customer usually uses the services. While this non-personally identifiable information is often combined with PII, it need not be, and companies often collect anonymized usage data for purposes of improving existing service and developing new products.

Do individuals place value on their NPII? The value of PII is often assumed to spring from the control of that information, which prevents financial loss through improper use or price discrimination (Acquisti and Varian 2005). There are no clear financial implications for sharing anonymized NPII. However, individuals may still value control of this information due to fear that this information might be 'deanonymized' (Ohm 2010), an endowment effect (Acquisti et al. 2013), or a general wariness about sharing any information online (Goldfarb and Tucker 2012).

2 The Survey Experiment

We estimate the value of NPII through a discrete choice experiment that examines hypothetical streaming video services. This experiment was administered online to 300 survey respondents who stated they were considering purchasing or renewing a streaming video service in the next year. The sample was demographically balanced, with equal numbers of men and women, and equal numbers in each of the age groups 18-34, 35-55, and over 55.

Each survey respondent was asked to evaluate a set of hypothetical streaming video service packages (e.g., online providers of movies and television shows) with different 'bundles' of attributes. One attribute of these hypothetical services was the privacy policy that governed the sharing of the individual's NPII and PII.

The text description of the privacy policies was limited to approximately the same length as the descriptions of the levels of the other attributes to avoid indicating to respondents that information sharing was an important feature of the study. Interested respondents could click on a hyperlink to obtain more detailed descriptions.

A policy of not sharing any data was described as: 'Some services do not share any information with third parties, except when necessary to provide services directly related to the streaming video service, such as credit card billing.'

Sharing NPII was described as: 'Some services share your usage information with third parties engaged in research, marketing, or advertising.' The phrase 'usage information' was hyperlinked. Clicking on the link created a popup textbox that read: 'Third parties use usage information to adjust program offerings, create advertising for different types of audiences, and develop new products. While this is information about you, it is not connected to information that would allow someone to contact you directly.'

Sharing PII was described as: 'Some services share your personally identifying information with third parties engaged in research, marketing, or advertising.' The phrase 'personally identifying information' was hyperlinked. Clicking on the link created a popup textbox that read: 'Third parties use personally identifying information to offer you new products, discounts on services related to your interests, and to create advertising that is specific to you. This information would allow advertisers and other companies to contact you directly to advertise additional services and products.'

In addition to the privacy policies for our hypothetical streaming video services, survey respondents were also presented with information on various possibilities for catalog size, content availability, commercials, and price. The attributes and the levels of each attribute used in the discrete choice experiment were as follows:

Table 1 here

Hypothetical streaming video services were created using an orthogonal fractional factorial design. Each survey respondent faced 11 different choice scenarios in total, each consisting of 4 hypothetical streaming video services along with a 'none of these' option. With the exception of price, the levels of the alternatives were coded as dummy variables relative to the baseline category for each choice alternative for the analysis below. Another dummy variable indicated the 'none of these' choice alternative.

3 Model Specification

We estimate respondent choice using a mixed logit model (Train 2009). The coefficients on catalog content, speed of delivery, and commercials are all specified as independent normal distributions, while the coefficient on the no service option is specified as fixed. The coefficient on price follows a lognormal distribution in order to ensure that the calculated distribution of willingness to pay (WTP) will have finite moments (Daly et al. 2011).

A total of 40 survey respondents never selected a streaming video service that shared data. The implied coefficients on the data sharing variables for these individuals are infinitely negative, and specifying normally distributed random coefficients on these variables led to unrealistically large estimates of WTP to avoid sharing data (on the order of tens of thousands of dollars).

Rather than eliminate these individuals from the analysis, our solution to this problem was to specify the random coefficients on the data sharing variables as a mixture of a normal distribution and a point mass at -10. Individuals with coefficient values at this point mass would have a near-zero probability of selecting a streaming video service that shares data.

We estimate this model by defining two sets of coefficients. β_A follows the joint distribution described above, with normal distributions on the data sharing variables. β_B is identical to β_A, but sets the coefficients on the data sharing variables to -10. We can then estimate the probability that respondent n chooses hypothetical streaming video service i in choice scenario t as a weighted average of two mixed logits:

$$L_n(i,t|\beta_n,S) = (1-S)\frac{\mathrm{e}^{X_{nit}\beta_{nA}}}{\sum_{j=1}^{J}\mathrm{e}^{X_{nit}\beta_{nA}}} + S\frac{\mathrm{e}^{X_{nit}\beta_{nB}}}{\sum_{j=1}^{J}\mathrm{e}^{X_{nit}\beta_{nB}}}$$

where S is a fixed coefficient to be estimated that gives the fraction of the population that never selects a streaming video service that shares information. Note that this representation of our model is just for analytic convenience – this is still a standard mixed logit model, with distributions on the data sharing variables as described above.

4 Results

The coefficients on the data sharing variables apply only to those respondents who selected at least one hypothetical streaming service that shares data. For these individuals the mean and standard deviation on the coefficients for sharing NPII are not statistically significant, indicating that we cannot rule out the possibility that these individuals are indifferent about sharing NPII with third parties. In contrast, the mean and standard deviation of the coefficient for sharing both NPII and PII are statistically significant. The coefficient indicating the fraction of respondents who never selected a streaming video service that shares data is roughly 0.13, approximately matching the fraction in our sample ($40/300 \approx 0.13$). The results for the other attributes are all sensible, with respondents showing a preference for services that have more content, faster content availability, and no commercials. The coefficient on 'no streaming service' is negative, which is unsurprising as these respondents were individuals who had or were considering obtaining a streaming video service.

Table 2 here

For each feature of these hypothetical streaming video services, we calculate the distribution of the WTP by taking 10,000 draws from the normally-distributed random coefficient for the feature, and dividing each of those draws by one of 10,000 random draws from the lognormally-distributed price coefficient. These calculations show a mean WTP of $0.85 per month to avoid sharing NPII, although this is based on a statistically insignificant coefficient. The mean WTP to avoid sharing NPII and PII together is $5.52 per month, which is similar to that placed on more content ($6.62), faster content availability ($8.08), and not having to view commercials ($6.41).

Note that these WTP calculations are only for the share of the population that was not at the point mass of -10 on the coefficients on the data sharing variables. We cannot calculate the WTP at this point mass since the coefficient value assigned to this point mass is arbitrary. For individuals at that point mass, all we can say is that the WTP to avoid sharing data exceeds the cost differences in our survey design.

5 Discussion

For the large majority of our survey respondents, we did not find evidence that they were willing to pay to avoid sharing their NPII with third parties.

There are several possible interpretations for the 40 individuals who never selected a hypothetical streaming video service that shared personal information with third parties. Some may have been willing to select services that share personal information with third parties, but the bundles of attributes offered to them in the survey did not induce them to make this tradeoff. Others may be 'privacy fundamentalists' who place high importance on protecting their private information (Woodruff et al. 2014). Finally, some may be engaged in protest behavior, refusing to select hypothetical services that shared information in order to express support for online privacy. Our survey may have made the issue of data sharing salient to these individuals, leading to more pro-privacy responses (Tsai et al. 2011). The fact that 10 out of these 40 respondents report subscribing to a streaming video service, and state they are aware their streaming service is collecting NPII and PII, lends some support to this last interpretation.

6 Acknowledgments

We thank Kenneth Train for many helpful comments.

7 References

Acquisti, Alessandro, and Hal Varian. 2005. 'Conditioning Prices on Purchase History.' *Marketing Science* 24 (3): 367-381.

Acquisti, Alessandro, Leslie K. John, and George Lowenstein. 2013. 'What is privacy worth?' *The Journal of Legal Studies* 42 (2): 249-274.

Daly, Andrew, Stephane Hess, and Kenneth E. Train. 2012. 'Assuring Finite Moments for Willingness to Pay in Random Coefficient Models.' *Transportation* 39 (1): 19-31.

Goldfarb, Avi and Catherine Tucker. 2012. 'Shifts in Privacy Concerns.' *American Economic Review: Papers and Proceedings* 102 (3): 349-353.

Ohm, Paul. 2010. 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization.' *UCLA Law Review* 57 (6): 1701-1777.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation.* 2nd ed. New York: Cambridge University Press.

Tsai, Janice Y., Serge Egelman, Lorrie Cranor, and Alessandro Acquisti. 2011. 'The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study.' *Information Systems Research* 22 (2): 254-268.

Woodruff, Allison, Vasyl Pihur, Sunny Consolvo, Lauren Schmidt, Laura Brandimarte, and Alessandro Acquisti. 2014. 'Would a Privacy Fundamentalist Sell Their DNA for $1000 ... if Nothing Bad Happened as a Result? The Westin Categories, Behavioral Intentions, and Consequences.' In *Proceedings of the Tenth Symposium On Usable Privacy and Security: SOUPS '14, USENIX*, 1-19.

Table 1: Attributes and Levels in Discrete Choice Experiment

Attribute	**Levels**
Privacy Policy	NPII Shared ('Share NPII') NPII and PII Shared('Share NPII and PII') No Information Sharing (baseline category)
Catalog Size	10,000 movies, 5,000 TV episodes ('More content') 2,000 movies, 13,000 TV episodes ('More TV/Fewer Movies') 5,000 movies, 2,500 TV episodes (baseline category)
Speed of Content Availability	TV episodes next day, movies in 3 months ('Fast Content Availability') TV episodes in 3 months, movies in 6 months (baseline category)
Commercials Shown Between Content	Yes ('Commercials') No (baseline category)
Price per Month	\$6.99 \$8.99 \$10.99 \$12.99

Table 2: Factors Influencing Choice of Streaming Video Service

Variable	**Fixed/Mean Coeffs.**	**σ of Coeffs.**
Share NPII	-0.0556	0.1439
	(0.0587)	(0.1628)
Share NPII and PII	-0.3664*	0.7084*
	(0.0776)	(0.0899)
Fraction Non-Sharing	0.1270*	
	(0.0199)	
Price per Month	-2.0950*	1.1181*
	(0.1139)	(0.0890)
More Content	0.4353*	0.7041*
	(0.0710)	(0.0844)
More TV/Fewer Movies	0.0094	0.7349*
	(0.0751)	(0.0981)
Commercials	-0.4183*	0.8694*
	(0.0706)	(0.0815)
Fast Content Availability	0.5335*	0.7378*
	(0.0659)	(0.0755)
No Streaming Video Service	-2.9936*	
	(0.1618)	

Note: * $p < 0.001$ (two-tailed). Standard errors in parentheses. All randomly distributed coefficients follow normal distributions except price, which is a lognormal distribution, and the data sharing variables, which are normal distributions with a point mass at -10. Random coefficients were estimated using 500 quasi-random draws.