

19 June 2014

The Value of Personal Information to Consumers of Online Services: Evidence from a Discrete Choice Experiment

By **Sarah Butler** and
Garrett Glasgow

Recent years have seen rapid growth in privacy-related litigation. While the most high-profile cases have involved data security breaches at large retailers¹, a growing number of cases involve data collection and sharing done by legitimate companies—that is, companies that collect user data and sell this information on to third party advertisers and vendors. For example, in one recent case, comScore settled the class action case challenging its data use and data sharing rather than face potentially damaging litigation.²

Despite the growth of legal suits involving data collection and data privacy, many areas of the law and economics surrounding these matters remain undetermined. One of the central economic issues in a data privacy case is the alleged harm to consumers caused by the breach of privacy. Unless the private or personal data have been used in an unauthorized transaction, the framework for establishing legal and economic harm in this area is unclear, even in cases involving the exposure of personal financial information such as credit card numbers and security PINs.

While there are academic papers evaluating the extent to which consumers will allow or prohibit uses of their personal information, empirical data that quantify the value of personal data are rare. There is even less information on how much consumers would be willing to pay to keep their data private or how much they would want to be compensated for the use of their information. Our study fills this gap. In this paper, we use survey and modeling-based techniques to quantify the value that consumers place on their personal information.

“Privacy Matters.”

Below are some examples of recent data privacy cases with rulings in both Plaintiffs’ and Defendants’ favor. A number of the recent decisions relate to whether or not the type of information collected is personally identifying or private.

- **In re: Hulu Privacy Litigation:** Judge Beeler in the Northern District of California denied (in part) the summary judgment motion of Hulu, an online provider of video content. At issue in this case is the nature of the information that was shared with third parties, whether this information constitutes information which renders a user personally identifiable, and whether this information was shared “knowingly” in violation of the VPPA.³
- **In Re: Zynga Privacy Litigation:** The 9th Circuit Court of Appeals affirmed the dismissal of a complaint filed against Facebook and Zynga for violations of the Electronic Communications Privacy Act (ECPA). In this matter the panel of judges agreed with the lower court’s assertion that the “referrer headers” do not constitute “communications” under the act.⁴

Other recent examples of cases with rulings that relate to the nature and comprehension of privacy policies include:

- **In the matter of Snapchat, Inc., a corporation:** The FTC recently settled with the mobile messaging service Snapchat. In the consent order, the FTC indicated that Snapchat misrepresented the functionality and security of its photo sharing application, specifically the extent to which images could be deleted and whether or not users could detect when an image had been captured or saved in some manner.⁵
- **In re: LinkedIn User Privacy Litigation:** In his 28 March 2014 decision, Judge E. Davila allowed Plaintiff’s putative class action against LinkedIn to proceed. In this case, the named Plaintiff alleges she read and relied on the LinkedIn privacy policy prior to her purchase of the LinkedIn premium account and that this policy did not accurately reflect the site’s data privacy procedures. Plaintiff further alleges that LinkedIn misrepresented its compliance with industry standards for privacy protection.
- **In re: Google Inc. Gmail Litigation:** Class certification was denied (and was refused review at the Court of Appeals) in a case against Google for its scanning and use of information contained in users’ emails. In this matter before Judge Lucy Koh, the court agreed with Google that the nature of scanning disclosures and the extent to which consumers consented would require individualized inquiries and thus the matter was not appropriate for certification.⁶

Using Survey Methods to Estimate the Value of Personal Information

To estimate the value that consumers place on their personal information, we designed an experiment using an established survey method called conjoint analysis, which involves giving survey respondents hypothetical choices from which to choose their preferred option. Specifically, our survey respondents were asked to evaluate a set of hypothetical streaming video service packages (e.g., online providers of movies and television shows) where the options differed along a variety of dimensions, including the extent to which the provider kept the user’s personal data private.⁷ Although the survey gave respondents a set of hypothetical choices to evaluate, there is some realism here—streaming video service providers have the opportunity to collect different types of information about consumers. Moreover, we could explore the value placed on different types of personal information. For example, providers of streaming video services could potentially collect and share with third parties both *usage information* (such as personal information about viewing habits and preferences) and *personally identifying information* (such as email addresses). Below we examine the value of each of these types of personal information.

The survey was designed as a discrete choice experiment—a specialized type of survey which asks consumers to evaluate products that are “bundles” of attributes and make decisions about the most preferred product based on its set of attributes. Discrete choice experiments have a long and respected history in market research and academics, and can help researchers estimate how consumers value a particular attribute by observing how they make choices about products as a whole. The survey was administered to an online panel of qualified survey respondents. Qualified respondents were adults who stated they were considering purchasing or renewing a streaming video service in the next year.⁸

After qualifying for the survey, respondents were presented with descriptions of five attributes of online streaming video services they might wish to consider when choosing a provider—the provider’s privacy policy, catalogue size, content availability, commercials, and price.

The levels of these five attributes were:

- Privacy Policy:
 - Information is collected but not shared (baseline category)
 - Usage information is shared with 3rd parties (“share usage”)
 - Usage and personal information are shared with 3rd parties (“share usage and personal”)

- Catalogue Size:
 - 10,000 movies and 5,000 TV episodes (“mostly movies”)
 - 2,000 movies and 13,000 TV episodes (“mostly television”)
 - 5,000 movies and 2,500 TV episodes (baseline category)

- Speed of Content Availability:
 - TV episodes next day, movies in 3 months (“fast content”)
 - TV episodes in 3 months, movies in 6 months (baseline category)

- Commercials Shown Between Content:
 - Yes
 - No

- Price per Month:
 - \$6.99
 - \$8.99
 - \$10.99
 - \$12.99

With respect to the provider’s privacy policy, respondents were told “Different services have different policies regarding how information about you and your viewing habits are used.” The three levels of the privacy policy attribute were 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.”

- “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.”

- “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.”

After reviewing information about the different possible attributes, respondents were then asked to make choices from the hypothetical service packages shown. Each choice scenario shown to respondents presented four different streaming video services, each with a distinct combination of the attributes previously described. Respondents were asked to select the service they would be most likely to purchase, or to state they would not purchase any of the four services presented. Each respondent was presented with 11 different choice scenarios in total.

This choice data was then analyzed using a statistical technique called regression analysis. The underlying model of consumer demand is called a logit model, which is appropriate for an analysis of data that represent consumer choices. By estimating a logit model, we can determine the role that each attribute plays in determining the probability that a respondent will select a particular streaming video service package.

Table 1: **Logit Analysis of Streaming Video Service Choices**

| Attribute | Coefficient | Standard Error | Z |
|---|-------------|----------------|---------|
| Usage Data are Shared | -0.365 | 0.048 | -7.617 |
| Usage and Personal Data are Shared | -0.563 | 0.050 | -11.182 |
| Price | -0.094 | 0.009 | -10.092 |
| Mostly Television Content | 0.024 | 0.053 | 0.456 |
| Mostly Movie Content | 0.384 | 0.049 | 7.780 |
| Commercials are Shown | -0.262 | 0.040 | -6.526 |
| Faster Content is Available | 0.410 | 0.041 | 10.092 |
| No Streaming Video Service ⁹ | -1.262 | 0.113 | -8.821 |

The Value of Personal Information

Table 1 shows the results of our survey. The attributes of the choices shown to the survey respondents are in the first column. The second column shows the importance—positive or negative—of each attribute. Positive coefficients indicate that the attribute makes the streaming video service more desirable, while negative coefficients indicate that the attribute makes the streaming video service less desirable. As shown in Table 1, the results show that respondents prefer services with larger catalogues of movies, dislike commercials, and prefer faster access to content. These are obviously sensible findings.

With respect to privacy policy, the negative coefficients on the two data privacy policy variables—“usage data shared” and “usage and personal data are shared”—reveal that consumers are more likely to choose a streaming video package that did not share their data and less likely to choose a streaming video package that shared their data. Thus, the results of our model suggest that the survey respondents value their personal data.

We can calculate consumers’ willingness to pay (WTP) for particular attributes by dividing the coefficient for each attribute by the price coefficient. For attributes with negative values, we would interpret the WTP as the money a consumer would give to avoid having the attribute or the discount a consumer would require in order to accept the attribute. Our results suggest a WTP of \$3.90 per month to not share usage information with third parties. The WTP to not share both usage and personally identifying information is \$6.01 per month. Given that the highest priced service in our survey was \$12.99, this WTP suggests that consumers care a great deal about privacy and would be willing to pay a substantial fee to avoid sharing their information with third parties.

Given this, it is perhaps surprising that the majority of our respondents were unaware that streaming video services collected usage and personally identifying information and could potentially share this information with third parties. Overall, only 40.3% of our respondents were aware that usage data are collected by streaming video services, and only 35.3% were aware that personally identifying data are collected by streaming video services. Awareness was higher among those respondents that already had a subscription to a streaming video service, but even so, more than half of these consumers are unaware that personal and usage data are collected.¹⁰

What Does this Mean for Privacy Litigation?

These results indicate that consumers do value their information and that even information that cannot be used to personally identify a specific individual has value. Moreover, consumers want to be compensated for the use of their data. Perhaps unsurprisingly, consumers value their personal information more than they value anonymous data on their movie and television viewing habits. What do these results imply for how we may want to think about the value of personal information in the context of privacy litigation?

First, these results demonstrate the potential for personal information to be valued using a scientific and rigorous empirical approach.

Second, the key issues in privacy litigation are complex and likely will involve tradeoffs between the value consumers place on their personal information and any improvements in service that they may get if their information were used by the provider. For example, the sharing of information may allow advertisements and entertainment suggestions to be more targeted and useful.

Third, there are likely to be differences across individuals in how much they value their personal information. For example, as with many other online innovations, consumer awareness of the details surrounding streaming video services is still relatively undeveloped.¹¹ While our research did not find a statistically significant difference in the amount of compensation desired by consumers who knew data were collected and those who did not, there was variation between the two groups in their willingness to pay to prevent data sharing.¹² Further study may reveal additional differences in the valuation of personal information based on differences in consumers' understanding of data collection and data sharing.

Fourth, an increasingly important issue in privacy litigation is likely to be the extent to which consumers are "wilfully ignorant" of data collection and sharing policies or the extent to which the language used to describe such policies is clear and not misleading. This is an empirical question, and in other types of litigation matters, notably false and misleading advertising matters, survey and statistical methods have addressed these types of questions.

Our results here demonstrate that survey and modelling techniques provide an empirical method for evaluating the tradeoffs consumers must make when thinking about the use of their data. The study demonstrates that consumers value their information and place different valuations depending on whether the data are anonymous or are personal data. Finally, our data suggest that the valuation of privacy is likely to be somewhat dependent on the characteristics of the consumer and the extent to which the individual was aware of and understood the privacy policy.

Notes

- ¹ See the sidebar in this paper. Also see S. Romanosky, D. Hoffman and A. Aquisiti (2012) "Empirical Analysis of Data Breach Litigation," iConference Proceedings, pp.124-137.
- ² See settlement in *Harris and Dustan v. comScore* (dated 30 May 2014).
- ³ "As to the Facebook disclosures, there are material issues of fact about whether the disclosure of the video name was tied to an identified Facebook user such that it was a prohibited disclosure under the VPPA. In addition, the record is not developed enough for the court to determine as a matter of law whether Hulu knowingly disclosed information or whether Hulu users consented to the disclosures." See *In re: Hulu Privacy Litigation: Order Granting in Part and Denying in Part Hulu's Motion for Summary Judgment (comScore and Facebook)*, dated 28 April 2012, p. 2.
- ⁴ Broadly speaking, referer headers are the address of the webpage that immediately preceded the prior page and referred or linked to the next page. "The referer header information that Facebook and Zynga transmitted to third parties included the user's Facebook ID and the address of the webpage from which the user's HTTP request to view another webpage was sent. This information does not meet the definition of "contents," because these pieces of information are not the "substance, purport, or meaning" of a communication." See *In Re: Zynga Privacy Litigation: Appeals from the United States District Court for the Northern District of California James Ware, District Judge, Presiding* dated 8 May 2014.
- ⁵ "It is ordered that respondent and its officers, agents, representatives, and employees, directly or indirectly, shall not misrepresent in any manner, expressly or by implication, in or affecting commerce, the extent to which respondent or its products or services maintain and protect the privacy, security, or confidentiality of any covered information, including but not limited to: (1) the extent to which a message is deleted after being viewed by the recipient; (2) the extent to which respondent or its products or services are capable of detecting or notifying the sender when a recipient has captured a screenshot of, or otherwise saved, a message; (3) the categories of covered information collected; or (4) the steps taken to protect against misuse or unauthorized disclosure of covered information." *In the matter of Snapchat, Inc., a corporation: Agreement Containing Consent Order*, File No. 132 3078.
- ⁶ "In sum, the Court finds that with respect to the Education Class, the substantial individual questions regarding the nature of each Google Apps Administrator's disclosures are likely to lead to individual questions regarding express consent that will predominate over common questions. The Court need not determine whether class-wide express consent questions will predominate over individual questions with respect to the Minor Class, Cable One Class, and the Non-Gmail User Classes because, as discussed below, the Court finds that individualized questions regarding implied consent will overwhelm any common issues regarding these Classes." See Order Denying Plaintiffs' Motion for Class Certification *In re: Google Inc. Gmail Litigation*, dated March 18, 2014.
- ⁷ While there are privacy related cases pending against video streaming services (as an example, see *In Re: Hulu Privacy Litigation N.D. California*), this study does not draw or replicate facts from any particular case. Any study designed for litigation would need to address the specific details of the particular matter.
- ⁸ 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. A total of 300 respondents completed the survey. A total of 65.3% of respondents already had at least one streaming video service. Among the respondents with an existing service, 24.7% had a subscription to Amazon Prime, 19.3% had a subscription to Hulu Plus, and 54.7% had a subscription to Netflix (percentages can add up to over 100% as respondents can subscribe to more than one streaming service).
- ⁹ This attribute represents the ability for a respondent to decide that he/she would prefer "no streaming service" over all of the other options shown in a particular choice set.
- ¹⁰ Among current subscribers to streaming video services, 43.9% were aware of the collection of usage data and 41.8% were aware of the collection of personally identifying information. Among those not currently subscribing to a streaming video service the numbers were 35.7% and 23.1%, respectively.
- ¹¹ Members of NERA's Survey and Sampling practice worked on a number of the pivotal litigation matters involving the sale of search engine keywords by third parties to competitors. These cases involved investigations into consumer understandings and perceptions of search and online advertising. As consumers' understanding of search engines grows and advertisers identify themselves more clearly, the number of keyword cases has decreased dramatically.
- ¹² There also appear to be differences in the understanding of data collection by gender and age. These issues are explored in a more detailed, forthcoming paper.

About NERA

NERA Economic Consulting (www.nera.com) is a global firm of experts dedicated to applying economic, finance, and quantitative principles to complex business and legal challenges. For over half a century, NERA's economists have been creating strategies, studies, reports, expert testimony, and policy recommendations for government authorities and the world's leading law firms and corporations. We bring academic rigor, objectivity, and real world industry experience to bear on issues arising from competition, regulation, public policy, strategy, finance, and litigation.

NERA's clients value our ability to apply and communicate state-of-the-art approaches clearly and convincingly, our commitment to deliver unbiased findings, and our reputation for quality and independence. Our clients rely on the integrity and skills of our unparalleled team of economists and other experts backed by the resources and reliability of one of the world's largest economic consultancies. With its main office in New York City, NERA serves clients from more than 25 offices across North America, Europe, and Asia Pacific.

Contacts

For further information and questions, please contact:

Sarah Butler

Vice President
+1 415 291 1022
sarah.butler@nera.com

Dr. Garrett Glasgow

Senior Consultant
+1 415 291 1000
garrett.glasgow@nera.com