Jacques Bughin

is a director with McKinsey & Company's Brussels office. Since joining the firm in 1992, he has worked on more than 600 clients' projects and has been based in the Amsterdam, London, Montreal, New York and Toronto offices. He is a core leader of McKinsey's telecom/media/high-tech practice, as well as Marketing and Strategy practices. He co-leads the McKinsey Technology Initiative, looking at how technology trends are creating major opportunities to adapt businesses. He received both a Master's *summa cum laude* and a doctoral degree in Economics. He has lectured at many universities and is also a fellow of ECORE (a think-tank on economic policy in Belgium), of Applied Economics at KUL University, and of the Aspen Institute. He has co-authored several books, including *Managing Media Companies*, *Harnessing Creativity* (Wiley and Sons), as well as being published in many leading business and academic journals — for example, *Management Science*, *McKinsey Quarterly* and the *Journal of Direct, Data and Digital Marketing Practice*.

Keywords: online, conjoint, privacy

Jacques Bughin
Director, McKinsey & Company
Avenue Louise, 480
B-1050 Brussels, Belgium
Telephone: +32 2 645 42 11

Digital user segmentation and privacy concerns

Jacques Bughin

Received (in revised form): 3rd August 2011

Abstract

How valuable, and, at the same time, how respective of privacy is the web? This paper uses conjoint techniques to assess how online users make the trade-off between use of free online applications and the desire to protect their privacy. While a large number of web users are indeed greatly concerned by cyber-privacy issues, I find that the value users are willing to pay to ensure their privacy protection is only several euro per month — an order of magnitude *below the value* users extract from using currently mostly ad-based, free online applications such as Facebook and Google. I further show that value for privacy is not homogenous in the online user population. I identify seven segments with different trade-offs between the value of cyber-privacy protection and use of online applications. *Journal of Direct, Data and Digital Marketing Practice* (2011) **13**, 156–165. doi:10.1057/dddmp.2011.28

Introduction

‘In the Google-Facebook race, privacy will determine the winner’ (<http://tech.fortune.cnn.com/2011/07/12/google-facebook-race-privacy/>, accessed 12 July 2011)

There is no doubt that concern for privacy has become an important matter on the internet.^{1,2} A Google web search regarding online privacy generated no less than 700 million entries on July 17, 2011. Assessing the evolution of web search queries on *Google Insights* on the same date shows that the terms ‘Google privacy’ and ‘Facebook privacy’ have quickly increased in importance — even if searches linked to online privacy, in total, have *declined* relative to all searches worldwide.

But beyond the obvious, the picture remains blurred: if one expected privacy concerns to be important, why then would users not take the appropriate actions to secure themselves against privacy leakage risks? Today, online users still rarely use the opt-out service that prevents third parties from using their data.³ Similar behaviour was noted by Hahn and Layne-Farrar⁴ regarding direct marketing purposes. Acquisti

This paper reflects the author's own private view, and does not engage the institutions he is affiliated with. The data for this article were collected during a project commissioned by IAB on the value of advertising-based applications.

and Gross⁵ also found that stated concerns about sharing information are (at best) a weak predictor of someone joining or not joining an online social network such as Facebook.

As Chellappa and Sin⁶ concede, the question is therefore not whether online users are concerned about privacy leakage risk. The real question is how online users weigh the value of protection against privacy leakage risk against the value they receive from using the web as freely as possible. This study follows seminal work by Hann *et al.*^{7,8} and uses conjoint analysis to estimate trade-offs made by users between the benefits of using the web and the risk of leaking cyber-privacy.

Of significance, work by Hann *et al.*^{7,8} and others has concentrated on e-commerce websites, where ‘trade-off’ is defined as the convenience of using e-commerce versus the risk of third-party information use. I chose here to concentrate *on all other application domains outside of e-commerce*, where the value of the web is likely driven by the fact that they are mostly ad-funded and free to users, but users face the risk of revealing too much information, which could be used without consent, for targeted advertising or for any other use. The applications I test include all major online domains, from online social networks and wikis to online search, user-generated video and others. The scope here is *privacy issues linked to the unpermitted use of personal data and web-surfing behaviour information, including targeted advertising*. By concentrating on all applications aside from e-commerce, we will be able to contrast how people make trade-offs on the web for the majority of their online time usage. Other distinctive aspects of the research are that it covers four countries (the US and the three largest economies in Western Europe — France, Germany and the UK) and is timely (survey conducted in 2010). In addition, it develops a comprehensive clustering of users’ trade-off valuation. While other research concentrates solely on clustering along the axis of privacy value, I segment user behaviour along both axes of usage and privacy valuations.

The three main findings are as follows. *First*, the conjoint analysis demonstrates a large spread in the value of cyber-privacy, though with a rather low average valued at about 4 euro/month per user (or about 8 euro per household) for the four countries analysed.⁹ This value is also significantly lower than the value users derive from using free, ad-funded, online services. The use value weights 38 euro per household per month. *Secondly*, the value of privacy is positively — albeit very weakly — correlated with the *stated concern* of users regarding the risk of abuse of their information. This finding confirms the thesis that what really matters is the *relative* value of protection versus the free use of online applications. *Thirdly*, I find up to seven segments regarding how online users make trade-offs between usage and privacy protection. This implies for marketers that a ‘one size fits all’ approach to privacy and online services is an ill-advised strategy and that more optimal solutions would require segmentation tactics.

Panel of 750 internet users in US, Germany France and UK

Conjoint analysis analyzing trade off between on-line usage benefits and privacy

Data and methods

Sampling

The research was based on an online survey administered in the spring of 2010 with a random sample of 750 online broadband users. Some 75 per cent were used in the final analysis after cleaning out incomplete surveys or inconsistent results.¹⁰ The survey originated as a request by the Internet Advertising Bureau (IAB)¹¹ to estimate the value of online applications. I refer those interested to the IAB report for further details on the data collection. I concentrate here on four countries — the US, Germany, France and the UK.

The value of internet services is estimated from a list of 16 applications/services clustered into three main domains: *Communication*: email, instant messaging, internet phone and social networks; *Entertainment*: gaming, gambling, virtual worlds, music/video, advanced upload services, podcasts and content reading; and *Information Services*: search/comparison shopping, maps, directories/yellow pages, blogs and wikis.

The questionnaire includes socio-demographic elements, internet usage, stated services interest and willingness-to-pay, as well as a list of conjoint attributes aimed at assessing the value trade-off between use of services and privacy leakage risk. I compared the sample with the internet population of each country using five key variables. Regarding access, I looked at (a) mix of access technology, (b) access performance (download speed) and (c) monthly access spent; for services, I looked at (d) time spent online and (e) usage rate of services such as search, social networks and online video. Based on these variables, the sample is statistically representative of the population. *By mid-2010*, the typical sampled online household was using a DSL line with average speed of 4Mbps at a price of 31 euro per month. Users were also accessing the web 23 days a month and were spending about two hours a day accessing online applications, including the web. Two global *premium* brands (ie brands that users are willing to pay for in order to continue accessing, even if there are free alternatives) stood out in the four countries — not surprisingly to many, these two brands are Google (and some of its offspring, such as YouTube) and Facebook.

Statistical methodology

Methods used to estimate value of usage and privacy are usually based on opinion polls/stated preferences. However, it is well known that directly stated preferences are usually a biased indication of true behaviour, especially when it concerns free services (eg because it leads to free-riding issues). Further, I have highlighted that stated concerns typically weakly correlate with true values when it relates to online privacy. Conjoint analysis alleviates these problems. The technique involves presenting respondents with a set of trade-offs to make between service/product benefits and related costs (privacy, paid service etc.), for which they rank their preferences. For each product

Conjoint		Attributes	Levels
		Communication	<ul style="list-style-type: none"> No access to communications and social network services Access only to basic services (mailing, instant messenger) Access to all services (mailing, instant messenger, social networks, internet phone, ...)
		Entertainment	<ul style="list-style-type: none"> No access to entertainment Access to basic music and UGCvideo Access to all services (music, video, games, gambling, web TV, ...)
		Information services	<ul style="list-style-type: none"> No access to information services Access to search engines (search engine, directories) Access to all services (search engine, directories, comparison shopping engine, web mapping, yellow pages, wikis, blogs, ...)
		Advertising	<ul style="list-style-type: none"> No advertising at all Advertising limited to rich media ads (video, audio) Advertising limited to pop-up banners Heavy rich media ads (video, audio) Heavy pop-up banners
		Information required	<ul style="list-style-type: none"> No personal info required Limited personal info required (name, age, usage, ...) Full info required (name, age, usage, credit card, ...)
		Monthly price	<ul style="list-style-type: none"> Range of prices tested: 0, 3, 6, 9, 18 EUR per month

Figure 1: Conjoint attributes and levels

and service benefit, the ranking contribution is the part-worth times the level of that product/service benefit usage. Therefore, the part-worth measures the marginal utility of the product/service in the ranking of the conjoint stimuli. This method, while common practice in marketing science, has been only rarely used in assessing privacy issues. Exceptions are Hann *et al.*⁷ and Kondo *et al.*¹²

One critical element in the design of conjoint research is to have a limited set of domains/attributes (fewer than ten at least and more like five to seven) for which to test.¹³ I considered six attributes in this study. The first three are the three ‘domains’ of online applications referred to above, that is, communication services (instrumented by the variable denoted, COMM), entertainment services (ENTER) and information services (WEBS). The next two attributes are ‘nuisance attributes’, that is, the extent of receiving targeted advertising (ADV) and improper use of user data (PRIVACY). The last one is PRICE. I was cautious to define price as *the price paid on top* of everything paid to use online, for example, access. Valuation of using services is uncorrelated with access price in this sample.

For each of these attributes, I created a set of levels, usually three (Figure 1). As the combination leads to a large set of conjoint stimuli, I reduced the set of stimuli to about 30 possible combinations. For example, there is no point in simulating a bundle of communication services, such as e-mail and instant messaging, only for 25 euro per month, per user. Finally, the currency reference was that of the user’s country, at the euro/currency rate as of 1 April 2010.

The statistical model used is a random probit model, in which I observe the rank preference of each respondent (‘the dependent variable’) against various combinations of the attributes (COMM, ENTER, WEBS, ADV, PRIVACY, PRICE) (‘the independent

variables'). Each attribute has a random individual component (except PRICE), reflecting that each individual may have specific preferences around the average core value of each attribute. From the law of large numbers, a normal distribution is hypothesized and maximum likelihood techniques are used to derive the best efficient estimates of the coefficients vector, B , of the model, linking attributes to rank preferences. Formally, for the j -th individual and the i -th level of attribute: the empirical model to be estimated by maximum likelihood is defined as,

$$\text{Arg max } Y_{ij} = B' \cdot X_j + e_{ij} \quad (1)$$

where

$$Y_{ij} = \text{Pr}(I_{ij}) / (1 - \text{Pr}(I_{ij})) \quad (2)$$

$$X_{ij} = [\text{COMM}_{ij}, \text{ENTER}_{ij}, \text{WEBS}_{ij}, \text{ADV}_{ij}, \text{PRIVACY}_{ij}, \text{PRICE}_j] \quad (3)$$

and e_{ij} is specific to the j -th individual and I ($=0.1$) is indicative of the conjoint stimulus being ranked as best choice by the j -th individual. (Detailed results are available upon request.)

Table 1 presents the derivation of the willingness-to-pay per service, at the *sample mean*. The goodness of fit is large (pseudo- R^2 of more than 0.7) and each coefficient of the attributes, as well as derived willingness-to-pay, is statistically significantly different from zero. The major insights from the maximum likelihood estimates are as follows.

- (1) The average willingness-to-pay for online services is 23 euro per month, per user (and up to 38 euro per household). This is higher than what a broadband home pays for its monthly connection costs in my sample and up to four times what service providers collect from exposing users to advertisers online. Hence, the surplus value collected by broadband users is quite large.

Table 1: Estimation results

Variables	Coefficient	Implied willing-to-pay
COMM	0.210	10.4*
ENTER	0.128	6.4*
WEBS	0.141	7.0*
ADV	-0.054	2.8*
PRIVACY	-0.032	1.6*
PRICE	-0.020	N/A

* $P < 0.01$; ** $P < 0.05$.

- (2) On the other hand, the value of privacy is only 4 euro per month, per user (or roughly 8 euro per month, per household). This is about one-fifth the value of usage and might explain why, despite online concern, usage of online applications is high and growing quickly.
- (3) Online privacy value is roughly equally split between limited targeted advertising and respect for data privacy.
- (4) The value curves are plotted in Figures 2 and 3. They demonstrate a significant spread around the mean — for example, with some users not willing to pay anything for cyber-privacy and a few willing to pay up to 100 euro per month, per user. This large variance of value calls for user segmentation in order to isolate clusters of more homogenous behaviours (see later section on user segmentation in the article).¹⁴

Figure 2: Usage value curve
Note: Average — 23.3 EUR/month

Figure 3: Cyber-privacy value curve
Note: Average — 4.2 EUR/month

- (5) Finally, remember that I also asked people to directly state their willingness-to-pay for usage and privacy protection. I found positive (albeit not statistically significant) correlations (respectively, 14 per cent and 6 per cent) between stated and conjoint-based value for usage and privacy. In general, I found that low-value people were overstating their value and vice versa, as well as overstating their privacy concerns in most cases. This is in line with other work findings (eg West³) and reinforces the superiority of the conjoint method.

Online user segmentation

7 digital segments

The shape of the value curves and the statistical evidence for the random logit model demonstrate the fact that individuals systematically differ in their trade-off between value of cyber-privacy and value of online usage. I used clustering to assess the prevalence of various cohesive segments in the online population.¹⁵

While other studies demonstrate three clusters, for example, Westin¹⁶ and Hann *et al.*,⁷ these studies did not integrate the diversity of applications online as this one does. Further, Westin¹⁶ used opinion surveys, which I confirm to be rather ‘noisy’ when it comes to their true representation of online privacy. I found up to seven clusters in my sample (Figure 4).

There is a large segment (26 per cent of users) matching the average value of both usage and privacy: about 25 euro per month, per user from using online, for a value of privacy of 3.5 euro per month, per user. I label this segment the ‘Average value users’. Around this segment, we have a large chunk of low-usage value users (59 per cent of user population) versus a more narrow set — 15 per cent — of

		<div><div></div><div>Size</div><div>Percent</div></div>							<div><div></div><div>Low</div><div>High</div></div>
		Willingness to pay				Pain points			Total
	Segment	Communication	Entertainment	Web services	Total user value	Advertising	Privacy	=	Surplus
“High user Value” segments	Premium information <div>2</div>	21.0	17.5	67.6	= 106.1	-1.8	-1.2	=	103.1
	Premium entertainment <div>1</div>	18.9	66.0	18.2	= 103.1	-1.7	-0.4	=	101.0
	Premium communication <div>4</div>	69.2	15.0	18.3	= 102.5	-1.5	-1.2	=	99.9
	Comprehensive internet <div>7</div>	29.1	12.0	20.3	= 61.4	-2.9	-1.8	=	56.7
	Traditionalists <div>26</div>	22.0	16.0	19.3	= 55.3	-29.5	-9.1	=	16.7
	Average value <div>26</div>	11.6	4.6	9.1	= 25.3	-2.3	-1.1	=	21.9
“Low user value” segments	Fair deal <div>19</div>	3.9	2.2	3.9	= 10.0	-5.5	-4.5	=	0
	Low value <div>2</div>	2.7	1.4	2.4	= 6.5	-1.0	-0.4	=	5.0
		10.2	5.0	8.1	23.3	-2.6	-1.6		19.1

Figure 4: Segmentation

internet users who gain very high surplus value (80 euro per month, per user) from using online applications.

Low-usage value segments

Regarding the low usage value, one witnesses two sub-segments, with one — which I label the '*Fair-deal users*' (19 per cent of user population) — placing a high value on cyber-privacy, up to double the respondent average. This is a segment for which the value of protection is as high as the value of usage, bringing net surplus to virtually zero for this group of online users. This segment is typically more male, married and less educated.

The second segment, '*Low-value users*' (40 per cent of the population), has the lowest willingness-to-pay for protection as well as for usage value. This segment is over-weighted in terms of age and is also the least urban.

High-usage value segments

About 14 per cent of internet users retrieve net value above the '*Average-value users*' segment. Four niche 'domain premium' segments are available in the data.

The first domain premium segment is called the '*Premium entertainment users*', and is composed of 1 per cent of users generating significant value by consuming entertainment services (approximately 65 per cent of their total surplus versus approximately 20 per cent market average). This niche segment focuses on using TV/videos and music, and is characterized by a higher percentage of single urban internet users (73 per cent live in urban areas versus 59 per cent for the full-sample average; 45 per cent versus 31 per cent are single).

The '*Premium information services users*' is the second niche segment, composed of 2 per cent of users. They get large value from using information web services (approximately 65 per cent of their total surplus), especially search and map-direction services. In general, this segment is characterized by more highly educated users and users who tend to be married with children. Finally, the last niche segment, labelled '*Premium communication users*', is composed of 4 per cent of users who place a strong value on communication, particularly email, social networks, and instant messaging (approximately 70 per cent of the total surplus value). This segment is characterized by a propensity towards young, urban, female internet users. Finally, I have one segment representing 7 per cent of users, which I label the '*Comprehensive Internet users*'. This segment generates value from intensively using *all* types of services (entertainment, information and communication) and is slightly more female and single than the user population average.

Privacy segment

Finally, the seventh segment stands for only 1 per cent of the population, also generates large value from using all online services, but places major value on the respect of their privacy — that is, close

**'No size fits all'
approach to digital
privacy and advertising
exposure**

to 40 euro per month, per user. I call this segment the 'Cyber-privacy fundamentalists'. They tend to have the lowest rate of marriage and have low income. We can contrast this with Westin¹⁶ and Dinev and Hart.¹⁷ Westin found a much larger segment of fundamentalists, in the range of 15–20 per cent of the user population. However, his segmentation is based on stated concern and not (more accurately) on the value that users place on cyber-privacy. In my results, I really have two 'privacy value' segments, but the largest one (19 per cent) is as much claiming low-usage value as it is calling for cyber-privacy.

Conclusions

This paper has used conjoint analysis to assess the trade-offs made by online users between using online services and the risk of privacy leakage. What makes the paper distinctive is its focus on all domains of online (outside of e-commerce) for four large Western economies (the US, France, Germany and the UK) and its clustering of trade-off behaviour. Key findings are that there is large variance among online users regarding these trade-offs, with up to seven user segments prevailing. Further, the value that users are willing to pay in order to protect themselves against privacy leakage is low relative to the value users get from using the existing breadth of online applications such as Facebook and Google.

In general, the paper also shows that reaction to privacy and advertising is quite diverse among internet users, as is the valuation of ad-based free services. Marketeers would be well advised not to apply an overly simplistic, 'one size fits all' approach when it comes to advertising exposure, and to the guarantee given to users regarding data privacy.

References and Notes

1. GBDe. (2003) 'Future of the internet, cyber security', available at http://www.gbde.org/ig/cs/Cyber_Security_Nov03.pdf.
2. Market Research Limited. (2009), available at <http://www.marketresearchmedia.com/2009/05/25/us-federal-cybersecurity-market-forecast-2010-2015/>.
3. West, R. (2008) 'The psychology of security', *Communications of the ACM*, Vol. 51 (April), No. 4, pp. 34–40.
4. Hahn, R. and Layne-Farrar, A. (2001) 'The benefits and costs of online privacy legislation', Working paper 01-14, American Enterprise Institute-Brookings Joint Center for Regulatory Studies, October 2001.
5. Acquisti, A. and Gross, R. (2006) 'Imagined communities: Awareness, information sharing, privacy on the Facebook', *PET*, pp. 1–22.
6. Chellappa, R.K. and Sin, R.G. (2005) 'Personalization versus privacy: An empirical examination of the online consumers' dilemma', *Information Technology and Management*, Vol. 6, pp. 181–202.
7. Hann, I.-H., Hui, K.-L., Lee, T.S. and Png, I.P.L. (2002) 'Online information privacy: Measuring cost-benefit trade-off', Paper presented at the 23rd International Conference on Information systems.
8. Hann, I.-H., Hui, K.-L., Lee, T.S. and Png, I.P.L. (2003) 'The value of online privacy: An empirical estimation', Working paper, Marshall School of Business, University of Southern California.

9. To put this value in perspective, it is lower than price points typically charged for paid security software online, for example, Eset, McAfee and others.
10. I formally scanned the distribution of answers. The answers lying in the extreme parts of the distribution (more than two standard deviations of the distribution) were reviewed and inconsistent answers were then removed, for example, users spending five hours a day on the internet, but only spending on search, for example.
11. Internet Advertising Bureau (IAB). (2010) 'Consumers driving the digital uptake', white paper.
12. Kondo, M., Nakamura, A. and Mitomo, H. (2009) 'Quantifying the benefits of the internet and its applications', *Keio Communication Review*, Vol. 31, pp. 37–50.
13. Green, P. and Srinivasan, V. (1990) 'Conjoint analysis in market research: new developments and directions', *Journal of Marketing*, Vol. 54, pp. 3–19.
14. While not reported here, I similarly found that the standard deviations of the parameter attributes are all significant, confirming that a random model was appropriate and that user value distribution is large.
15. Hierarchical clustering was applied using average between-group linkage with (dis)similarity measured by the squared Euclidean distance to derive cohesive groups. A hierarchical method is recommended when one has no prior information on the number of clusters or cluster centers (see Hann *et al.*⁸).
16. Westin, A. (2001), Testimony before U.S. House of Representatives, Committee on Energy and Commerce, Subcommittee on Commerce, Trade, and Consumer Protection, Hearing on 'Opinion Surveys: What consumers Have To Say About Information Privacy', 8 May.
17. Dinev, T. and Hart, P. (2003) 'Privacy concerns and internet use — A model of trade-off factors', Presentation to the Academy of Management Meeting, Seattle, USA.