

Privacy and Disclosure: Customer Attitudes Towards the Use of Big Data by Retail Companies

May Merhej Sayegh¹ and Karine Aoun Barakat¹

¹*Université Saint Joseph, Beirut, Lebanon*
may.merhejsayegh@usj.edu.lb, karine.aoun@usj.edu.lb

Abstract

Data generation, collection and analysis by retail companies is not a new phenomenon, however the speed at which it is being done, the variety and reliability of sources used, and the volumes treated has reached a new level and heralded the advent of Big Data. The collection and use of this data poses a variety of considerations and questionings. Are customers willing to disclose their personal information to companies? What do they expect in return? What is their attitude towards their personal information being traded and exchanged? Based on a customer survey, the article points to a dissociation between data and money from the customer viewpoint, leading us to consider consumers as being “data altruists”. It also shows a need for consumers to protect their privacy using several methods including selective information disclosure and seeking additional clarifications. Furthermore, customers engaging in information sharing with companies do so without resorting to information falsification.

Keywords: *Big Data, Privacy, Retail*

1. Introduction

The collection of Big Data and its manipulation has become a vital tool for companies looking to optimize their activities, know more about their customers, and gain an advantage on their competitors. The newness and rapid evolution of this practice paired with the absence of clear and adapted legislations poses numerous considerations from customers in terms of information privacy, confidentiality and exchange.

This has led stakeholders to consider a multitude of questions: How can Big Data are used while respecting individual rights and liberties, such as the right to privacy and confidentiality? What are the boundaries companies set themselves and are they aligned with customer expectations? To what extent are consumers willing to cooperate with companies using Big Data and what behaviors are considered to be unacceptable?

Through the course of this article and after proceeding with an exploratory review of the literature in order to define Big Data and understand the attitudinal considerations it triggers from the customer side with the help of the CPM theory, a qualitative methodology is adopted to gain company insight on how Big Data is being used. This is followed by a consumer survey that is aimed at understanding the ways in which companies can exploit Big Data while taking into account consumers' attitudes in terms of privacy protection, information disclosure and exchange, and motivation to give information.

2. Defining Big Data

Despite references dating back to the mid-nineties, the term Big Data only became widespread as recently as in 2011. Since then, its definitions have witnessed a rapid

Received (December 16, 2017), Review Result (February 25, 2018), Accepted (March 11, 2018)

evolution. According to Gartner, Inc. Big Data can be defined as: “high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.” [1]. Likewise, TechAmerica Foundation defines Big Data as being: “a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of the information” [2].

Whilst technology has served as the enabler of big data services, the wider interest in Big Data has been driven by thoughts of the potential commercial value that it may bring. Mainly, due to the ability of generating value from data in new ways which were not previously possible [3].

For retailers, Big Data techniques have been used to increase the productivity of corporate teams and forecast the evolution of markets [4], [5]. Big Data has proven to be a powerful tool for improving the efficiency of marketing such as for pricing strategies and advertising campaigns, as well as the management of product inventory and supply chains [6]. It has also changed how much companies know about their customers, the way in which they interact with them, and enhanced their ability to propose products and services that are more adapted to their needs.

However, the collection of Big Data has the potential to worsen consumer privacy concerns [7]. The increasing variety of data sources and contexts means that very often, the consumer is not even aware that data collection is taking place. It also leads companies, in some cases, to collect and analyze data which will not be used and thus entails the decrease of efficiency within the company.

3. Big Data and Consumer Interactions

Prior researchers have suggested that Big Data has brought a broad range of social and ethical issues and questions due to the underdeveloped regulations and regulatory infrastructure, which may give rise to consumer exploitation by businesses [3], [8], and [9]. This has been aggravated by the fact that collecting unstructured data is now more economically viable, and shifts in consumer usage of technology have made a wider range of data available, therefore giving an incentive for companies to collect as much data as possible. Yet, just because consumers are willing to provide data, this does not mean that its use is free from privacy implications [10]. There are also significant interrogations about truth, control, and power in Big Data studies: researchers have the tools and the access, while individuals as a whole do not. Their data were created in highly context-sensitive spaces, and it is entirely possible that some users would not give permission for their data to be used elsewhere. Many are not aware of the multiplicity of agents and algorithms currently gathering and storing their data for future use. Furthermore, users are not necessarily aware of all the multiple uses, profits, and other gains that come from information they have given [11]. This raises several important questions: How do customers protect their privacy? What type of information are they willing to disclose? What motivates them to give information to companies? And what is their attitude towards information exchange with third parties?

3.1. Privacy Protection

Consumer information privacy may be defined as an individual’s ability to determine when, how, and to what extent personal information such as personal profile information, search history, purchase behavior data...is disseminated to others [12]. Consequently, consumers’ privacy fears stem from their potential loss of control over their personal information.

This is addressed through the Communication Privacy Management (CPM) theory [13], which studies the tension between disclosure and privacy by examining how and

why people decide to reveal or conceal private information across various relational contexts. At the core of the CPM theory is the idea that disclosure has both benefits and risks, and thus that people must balance their competing needs for privacy and for disclosure. Disclosure renders people vulnerable to exploitation by others because information changes from being privately owned to being co-owned [13]. As such, disclosure always involves some degree of risk. This risk, according to the CPM theory, leads to the erection of boundaries around what information is considered public and private, thus allowing consumers to control who has access to the information and motivating them to set expectations for co-ownership of information once it is disclosed to others [13].

These boundaries include: withholding information about themselves, limiting exposure by falsifying information, and coordinating privacy rules with retailers by seeking clarifications prior to disclosure thus assessing the danger of disclosure and avoiding turbulence [14]. We can therefore consider as a first hypothesis the following:

H1: Consumers will withhold information from retailers, falsify information or seek additional clarifications prior to giving information as a privacy protection rule.

3.2. Selective Information Disclosure

Not all personal information carries with it the same degree of risk. As Petronio states, “private information changes in degrees of risk based on perceived repercussions for revealing and concealing” [15]. This suggests that some information is more readily disclosed or withheld according to the perceived consequentiality of disclosure. It further suggests that individuals develop privacy rules such that as information is perceived to be more risky to reveal, it is more likely to be withheld. Studies on interpersonal communication show this to be true, for example, in certain situations people use topic avoidance protection rules to help them decide whether to avoid discussing particularly risky topics in order to protect themselves [16].

Hence, we will test using the following hypothesis whether consumers differentiate between various types of information whereby they decide to disclose some and conceal others.

H2: More sensitive information will be withheld to a greater extent by consumers than will less sensitive information.

3.3. Motivation to give Information

A commonly observed behavior is explained in what is known as the “privacy paradox” in which people commonly report that they are greatly concerned about privacy, yet they divulge their sensitive personal information rather freely [17], [18].

In this case, customers decide to make a trade-off between the loss of privacy and the benefits brought by disclosing information. For example, a shopping website analyzes a customer’s demographic information and browsing history, in order to offer personalized product recommendations to the user and improve the customer experience. Driven by some benefits, such as a personalized service or monetary incentives, the customer may be willing to provide his sensitive data to a trustworthy data collector [19]. We can thus consider our third hypothesis to be the following:

H3: Consumers are willing to give information in exchange for benefits or monetary incentives.

3.4. Third Party Information Exchange

In this context as well, we can talk about confidentiality which can address the issue of how to keep information exchanged in the relationship between consumer-company from

being disclosed or sold to third parties [20]. Unauthorized or inadvertent disclosures of data gained as part of an intimate relationship are considered to be breaches of confidentiality [21], which ultimately lead to a break down in the relationship between the company and its customers. We can therefore consider our fourth hypothesis as follows:

H4: Consumers are not willing to accept that companies disclose or sell their information to third parties.

4. Research Methodology

As the goal of our research is to collect information that will allow us to examine what are consumers' attitudes towards the use of Big Data by retail companies and its consequences notably in terms of information privacy, confidentiality and exchange, an inductive approach was followed. This consisted of conducting an exploratory research aimed at producing an understanding of the facts by relying on the perspective gained from existing theories. Then the quantitative methodology was adopted and a consumer survey covering 150 customers of the retail sector was conducted.

The quantitative research tool used in the study is the questionnaire. The questionnaire is considered to be a well-established tool within social science research for acquiring information on participant social characteristics, present and past behavior, standards of behavior or attitudes and beliefs and reasons for action with respect to the topic under investigation [22]. In our case, data collected from the survey and analyzed using SPSS allowed us to test our hypotheses.

To add insight to our study, and prior to launching our customer survey, three interviews were conducted with marketing managers working with Big Data within retail companies. The interview guide was prepared by relying on the literature. Each interview lasted between 45 minutes and 1 hour. During the interviews, professionals were asked to talk about data collection and analysis within their company, their perceptions of privacy, confidentiality and the limits they put while using data. Interviews were registered and re-transcribed [23].

Based on the data collected, the relevant categories were identified using open coding which allows the detection of the interesting elements in the data. Then, we used axial coding through which we were able to link the data together to address the selective coding of the chosen items from our theoretical construct. The results were presented in a table showing the main themes and variables as well as the persons met by indicating the key ideas, verbatim and also by encoding some variables [24].

5. Big Data Privacy from a Company Viewpoint

Following three interviews conducted with professionals within the retail industry, information was collected that allowed us to highlight the level of respect and breaches of privacy in collecting and analyzing data, incentives given in exchange for information as well as practices related to data transfer to third parties.

The table below presents the analysis of the conducted interviews.

Table 1. Interview Analysis

Verbatim	Theme	Theoretical References
"Ethical considerations are taken into account ... customers are informed and give their approval before information about them is collected"	Privacy Protection	[21], [25]
"A prerequisite to joining our loyalty program is the disclosure of certain information such as customer name, email address, mobile phone	Motivation to give information	[19]

number, home address...” “ We use the collected customer data to propose products that are more adapted to customers’ tastes”		
“Customers are not clearly told what the data is being used for ...”, “The privacy policy is nearly inexistent, and customers are not informed about their rights and what is being done with their data.” “Customers have come to trust us with their personal information” “Customers are almost always willing to give information and no complaints have been registered about the way in which customer data is being used”	Trust between customers and companies	[9]
“The company does not hand over the data to third parties” “The company acts on behalf of third parties who wish to send targeted advertisements to our customers”	Third Party Information Exchange	[10], [11], [9]
“The ethical boundaries set by the company include the sale of company data or the use of data for personal benefits, the excessive use of targeted advertisements ...”	Ethics	[10]

Company professionals consider themselves to be guardians of customer data, and fulfill the ethical duty of confidentiality - the obligation of the organization to safeguard the information and not to transfer it to a third party- as they are aware that it is essential for nurturing the trust relationship between the company and its customers. However, they tend to neglect certain aspects of customer privacy including customer rights, the way in which customer information is being used.

6. Customer Survey Findings

A survey was then conducted online involving 150 respondent chosen randomly amongst the customers of a large retail group in Lebanon during the month of May 2017. Questions included in the survey were based on the 5-point Likert scale (1 being strongly disagree and 5 being strongly agree), and collected data was analyzed using SPSS. In terms of privacy protection, consumers are in large part willing to give information to companies, the information given is accurate. Most often they seek additional clarifications by for example reading the privacy policy of the company.

Table 2. Findings Privacy Protection and Information Disclosure

Questions \ Values	Willing to give Information to Companies	Give Accurate Information	Seek Additional Clarifications from Companies
Mean	3.54	3.59	3.36
Median	3.59 ^a	3.64 ^a	3.43 ^a
Mode	4	4	4
Std. Deviation	.824	.914	1.271
Variance	.680	.835	1.614
Skewness	-.567	-.207	-.228

Therefore, we can say that H1 is partially validated as contrary to what was thought, consumers are in large part willing to give personal information about them that is accurate, but they do seek additional clarifications prior to giving this information.

With regards to selective information disclosure, our hypothesis H2 holds true as consumers appear to share information selectively. Based on the observed frequencies, they are more willing to share common information such as their name (92%); email address (90%), birthrate (71%), and mobile phone number (57%). And are more protective of information related to their home address (26%), home phone number (11%), salary and financial information (3%), name of friends (6%) and any other information which is less frequently asked for by companies or considered as more sensitive.

Table 3. Findings Motivation to Give Information

Questions \ Values	Give Information to Join a Loyalty Program	Give Information in Exchange for Free Products or Monetary Incentives	Give Information to Improve the Company's Products and Services	Give Information to Stay Informed about the Company's Latest News
Mean	3.63	2.93	3.87	3.01
Median	3.72 ^a	2.93 ^a	4	3.09 ^a
Mode	4	3	4	4
Std. Deviation	.878	1.041	0.917	1.156
Variance	.771	1.083	0.841	1.336
Skewness	-.837	-.010	-0.683	-.198

When it comes to understanding what motivates consumers to give information to companies, it is clear to see that monetary incentives or free products is not what customers look for first. On the other hand, customers are willing to surrender information on condition it is used to improve a company's products and services or to join the company's loyalty program and become part of the company's inner circle. Our hypothesis H3 is partially true, given that consumers give information in exchange for benefits such as joining a company's loyalty program or staying informed about the company's latest news, but do not expect monetary incentives or rewards in return.

Table 4. Findings Third Party Information Exchange

Questions \ Values	Allow Companies to Share Customer Information with other Companies	Allow Companies to Share Customer Information with Research Companies	Allow Companies to Sell Customer Information to other Companies	Allow Social Networks to Share Customer Information with Companies
Mean	1.95	2.21	1.56	2.19
Median	1.81 ^a	2.08 ^a	1.46 ^a	2.13 ^a
Mode	2	1 ^b	1	2
Std. Deviation	.940	1.070	.790	.953
Variance	.883	1.145	.624	.909
Skewness	.796	.410	1.456	.498

The sharing of consumer information between companies is not considered to be a desirable practice; however, consumers become slightly more accepting in the case where

information is shared with research companies. Selling personal information is strongly disagreed with by customers. Hence, the H4 hypothesis is confirmed.

Two main themes appear as a result of the survey. The first major finding is the dissociation between data and money from the customer viewpoint. Consumers are willing to give information and do not expect anything in return and likewise expect companies not to exchange this information and make money from it. This is aligned with the conclusions of Kohane and Altman 2005 [24] who found that monetizing the sharing process produced an ambivalent result, leading to the possibility that many consumers could be thought of as “data altruists”.

Secondly, consumers appear to be transparent with companies when giving information, they protect their privacy by practicing selective information disclosure or asking for additional clarifications prior to handing over personal information. Consumers are trusting of companies and will reward companies towards which they have positive sentiments with access to additional information, and punish those mishandling their information by denying them access to it in the future.

7. Conclusion

With the appearance of Big Data, and in the absence of legislations or regulatory bodies, companies are increasingly faced with the obligation managing their need for customer information and customers’ need for privacy. Our article shows that customers balance between a willingness to share information and the need to protect their privacy. The former is motivated by the desire to become part of the company’s inner circle and contribute positively to the development of its products and services. While privacy protection is primarily achieved through selective information disclosure, consumers willingly hand over accurate information to companies in a context of trust which what has prevented any misdoings so far. Therefore, companies are pushed to nurture this trust by being constantly aware of the positions taken by their customers, and aligning themselves accordingly.

The limitations of the article are linked to the newness of the topic notably the limited literature surrounding Big Data and privacy which made the exploratory phase of our research difficult. Furthermore, given the time constraints, our quantitative research was restricted to a survey of 150 respondents, and in this sense an increased and more varied sample size and composition would have been more desirable, as well as more detailed data analysis.

Building on this article, academics can consider conducting the same research within a different context or for a different sector. Widening the scope is used to include issues such as passive data collection, data ownership, and surveillance. Exploring the concepts of data monetization and data control is performed to try understanding possible links between these two themes and customer perceptions of companies and brands.

References

- [1] Gartner IT Glossary (n.d.), Retrieved From <http://www.gartner.com/it-glossary/big-data/>.
- [2] A. Gandomi and M. Haider, “Beyond the hype: Big data concepts, methods, and analytics”, *International Journal of Information Management*, vol. 35, no. 2, (2015), pp. 137-144.
- [3] D. Nunan and M. Di Domenico, “Market research & the ethics of big data”, *International Journal of Market Research*, vol. 55, no. 4, (2013), pp. 505-520.
- [4] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. Hung Byers, “Big data: The Next Frontier for Innovation, Competition, and Productivity”, *McKinsey Global Institute*, (2012).
- [5] B. Harcourt, “Against Prediction”, *University of Chicago press*, (2007).
- [6] C. P. Chen and C. Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data”, *Information Sciences*, no. 275, (2014), pp. 314-347.
- [7] E. Aguirre, D. Grewal, A. L. Roggeveen and M. Wetzels, “Personalizing online, social, and mobile communications: opportunities and challenges”, *Working Paper*, (2015).

- [8] J. Lane, V. Stodden, S. Bender and H. Nissenbaum, "Privacy, big data, and the public good: Frameworks for engagement", New York, NY: Cambridge University Press, (2014).
- [9] F. Neuhaus and T. Webmoor, "Agile ethics for massified research and visualization", *Information, Communication and Society*, vol. 15, no. 1, (2012), pp. 43-65.
- [10] D. Boyd and H. Eszter, "Facebook Privacy Settings: Who Cares?", *First Monday*, vol. 15, no. 8, (2010).
- [11] D. Boyd and A. E. Marwick, "Social Privacy in Networked Publics: Teens' Attitudes, Practices, and Strategies. A Decade in Internet Time", *Symposium on the Dynamics of the Internet and Society*. Available at SSRN: <https://ssrn.com/abstract=1925128>, (2011).
- [12] A. Nill and R. J. Aalberts, "Legal and ethical challenges of online behavioral targeting in advertising", *Journal of Current Issues and Research in Advertising*, Sunny Press, vol. 35, (2014), pp. 126-146.
- [13] M. J. Metzger, "Communication privacy management in electronic commerce", *Journal of Computer Mediated Communication*, vol. 12, no. 2, (2007), pp. 335-361.
- [14] S. Petronio, "Boundaries of Privacy: Dialectics of Disclosure", Suny Press, (2002).
- [15] W. A. Afifi and L. K. Guerrero, "Motivations underlying topic avoidance in close relationships", *Balancing the secrets of private disclosures*, (2000), pp. 165-180.
- [16] L. Xu, C. Jiang, J. Wang, J. Yuan and Y. Ren, "Information security in big data: privacy and data mining", *IEEE Access*, vol. 2, (2014), pp. 1149-1176.
- [17] E. Aguirre, D. Mahr, D. Grewel, K. D. Ruyter and M. Wetzels, "Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness", *Journal of Retailing*, vol. 91, (2015), pp. 34-59.
- [18] P. Norberg and D. Horne, "The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors", chapter 5, (2014).
- [19] Westin A. Computers, health records, and citizen rights. [accessed July 30, 2008]. http://eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED143358&ERICExtSearch_SearchType_0=no&accno=ED143358, (1976).
- [20] L. O. Gostin and J. G. Hodge, "Personal privacy and common goods: A framework for balancing under the national health information Privacy Rule", *Minnesota Law Review*, (2002), pp. 1439-1479.
- [21] M. Bulmer, "Questionnaires, Sage Benchmarks in Social Science Research Methods", Sage Publications, London, 1st edition, (2004).
- [22] H. K. Klein and D. M. Michael, "A set of principles for conducting and evaluating interpretive field studies in information systems", *MIS quarterly*, vol. 23, no. 1, (1999), pp. 67-93.
- [23] L. Garreau and R. Bandeira-De-Mello, "La théorie enracinée en pratique: vers un dépassement de la tension entre scientificité et créativité dans les recherches basées sur la théorie enracinée?", *AIMS 2010, Jun 2010, Luxembourg, Luxembourg*, (2010), pp. 1-19.
- [24] L. O. Gostin, "Health information privacy: Cornell Law Review", (1995), pp. 101-184.
- [25] I. S. Kohane and R. B. Altman, "Health-information altruists-a potentially critical resource", *The New England journal of medicine*, vol. 353, no. 19, (2005), pp. 2074-77.