How do firms ask for consumers' data permission? And how do customers react?

Caterina D'Assergio
University of Bologna
Sara Valentini
University of Bologna
Elisa Montaguti
University of Bologna

Cite as:
D'Assergio Caterina, Valentini Sara, Montaguti Elisa (2019), How do firms ask for consumers' data permission? And how do customers react?. *Proceedings of the European Marketing Academy*, 48th, (9878)

Paper presented at the 48th Annual EMAC Conference, Hamburg, May 24-27, 2019.

How do firms ask for consumers' data permission? And how do customers react?

Abstract

To regulate the growing potential violation of people's right to protect their privacy, the EU has recently introduced a new data protection regulation that offers EU citizens a shelter for their personal information by requesting companies to clearly explain how people's information is used and to get their express permission to used it. The key question for firms is how such information should be requested in order to minimize the risk of reducing or preventing access to customer data. This is the paper's focus. By examining, through LDA, 367 re-permission emails sent by firms serving EU customers before the introduction of the GDPR regulation, we are able to develop a taxonomy of the main themes used to pose these requests and identify which topics are more effective at increasing people's likelihood to provide access to their personal details. We find that the framing of the request and the presences of monetary incentives increase consumer's propensity to express permission.

Keywords: Re-permission emails, propensity to disclose personal details, Latent Dirichlet Allocation

Conference Track: Relationship Marketing

1. Introduction

Over the years, people have increasingly shared their personal information with retailers, social media platforms and companies they interacted with, in exchange for their services. Companies, in turn, have progressively used these data to profile their customers without them expressing permission. To regulate this growing potential violation of people's right to protect their privacy, the EU has recently introduced a new data protection regulation that offers EU citizens a shelter for their personal information by requesting companies to clearly explain how people's information is used. This is having several implications for firms including the need to go about requesting permission to access the data of their users and/or customers. In facts, before May 25, 2018 European and non-European companies interacting with E.U. citizens undertook a massive data re-permission request campaign whereby their prospects, users and customers where reached by an unprecedented number of letters, emails or SMS messages. Interestingly, this huge wave of simultaneous requests represents a unique opportunity for researchers to understand how companies across industry and country chose to frame a very contentious request to their potential and extant customers.

Additionally, the introduction of the new GDPR in Europe documents an increasing need for control over how firms gather, store and use personal data. But, what does this mean for firms? Should this lead to a reduction in potential and actual customers' propensity to disclose information? What can firms do to reassure their customers and users, mitigate their privacy concerns, and increase their likelihood to share personal data?

The main purpose of this study is threefold. First, we would like to map how firms communicate their requests for data and develop a taxonomy of the main topics used while taking advantage of the unique discontinuity offered by the introduction of the new GDPR in Europe. Second, we would like to understand which topics were used to a larger extent, and how such topics were combined in the same message. Third, we attempt to understand which topics are more effective at increasing people' likelihood to share their personal details and how many.

To address these points, we use a multi-method approach. First, we examine re-permission emails sent by a sample of 367 firms using Latent Dirichlet Allocation (LDA) models (Blei, Ng and Jordan, 2003). We validate this approach also using manual content analysis coupled with clustering techniques. Finally, we run a lab experiment to test how these topics affect individuals' probability to grant access to their data.

Our results outline the dominant topics charactering GDPR permission requests. In particular, we find that these requests center on eight main topics: control, protection, oblivion, transparency, regulation, incentives, gain frame and loss frame. Interestingly some of these themes indicate that firms are treating these communications as legal material, whereas others highlight the firms' wiliness to persuade by reassuring customers or by providing them with some tangible incentives. We also show that persuasive topics, in particular, gain and losses framing, and incentives may increase the consumers' propensity to express permission for the use of their personal details.

We believe that our research, is providing an extremely useful documentation on how EU companies went on asking their customers' permission to use data offers. Our taxonomy of the main themes used is a first step for both academics and practitioners to further investigate how best firms can be put forth their data requests and which themes are more likely to generate wider consent. In this line, our second study by showing how framing differently influences people's likelihood to grant data-use permission starts scratching the untouched surface of our understanding of how consumers will react to regulation restriction on privacy concerns.

2. Literature Review and Conceptual Framework

Although the topic of privacy is a well trotted area of research in sociology (e.g. Wedel and Kannan, 2016; Acquisti, Brandimarte and Loewenstein, 2015; John, Acquisti and Loewenstein, 2010;), only recently marketing researchers have started looking more closely at consumer's preferences for data privacy (e.g. Martin, 2018, Krafft, Arden and Verhoef, 2017; Tucker, 2014; Phelps, Nowak and Ferrel, 2000). Tucker (2014) shows that if people know that they can control their privacy online, they are more willing to provide personal information and to react positively to personalized ads (Tucker, 2014). Athley, Catalini and Tucker (2017) point out that the provision of a clear stated privacy policy lead to increased trust and reduced privacy concerns which positively influence the probability to grant access to personal data.

Furthermore, while taking a cost-benefit perspective researchers have examined whether and how firms can outweigh the perceived costs associated to data disclosure by increasing its perceived benefits (Krafft, Arden and Verhoef, 2017; White, Zahay, Thorbjørnsen and Shavitt, 2008). More specifically, prior research has tried to identify whether or not incentives play a role in increasing consumers' propensity to consent the use of their data. In this line Chellappa and Sin (2005) and Athey, Catalini and Tucker (2017) have shown that people are more inclined to grant firms with information when they receive both monetary and non-monetary incentives. By contrast, Krafft, Arden and Verhoef (2017) report that both monetary incentives and lotteries have no effect in increasing consumers' likelihood to release data. This body of work clearly leaves us with several doubts about the role of incentives to boost privacy permission which we will try to address in this paper.

Prior work has identified a number of factors affecting data disclosure mainly related to framing. For instance, John, Acquisti and Loewenstein (2010) show that the disclosure of private information is responsive to environmental cues such as the way in which people are asked –direct vs indirect questions, the way in which the form to be filled is designed—professional vs unprofessional and, the initial prompt of the request—evocating or not privacy concerns. By the same token Grossklags and Acquisti (2007) document that people prefer to be paid to disclose their information vs, paying to protect their data. Prior work (White, Novak and Hoffman, 2014; Acquisti, John and Loewenstein, 2013) also poses that the likelihood of receiving privacy consent can be affected by the framing of its request. More specifically, consumers seem more likely to grand data-use permission when the denial of data usage is presented as a threat to lose service quality (loss) vs. when the consent to use the data are presented as an opportunity to get better service (gain). Interestingly, however, Ku, Yang and Chang (2018) show that when firms frame their request as a threat to reduce service quality consumers can easily feel mistreated and experience reactance (Ku, Yang and Chang, 2018), in that suggesting that this approach is always viable.

3. Study 1: Mapping the Topics Used to Ask for Data Permission

3.1 Description of the data

In the lead up to the EU's new GDPR compliance deadline on 25th May 2018, many businesses were re-permissioning their databases of customers or registered users through email campaigns. The purpose of a re-permission email is to get explicit opt-in consent to be recorded in the firm CRM and to receive marketing communications. A lack of positive opt-in for the company means that the email address and information about the customer or potential customer should be deleted from the company database. A serious risk for firms is that individuals could decide to opt-out or simply mark these emails as unread messages or spam. Consequently, companies used different strategies to attract consumers 'attention and to encourage them to say yes. We took advantage of this discontinuity to analyze the email campaigns used by an heterogonous group of firms. More specifically, we created a database of 367 companies and we retrieved the re-permission email they sent in the occasion of the EU's GDPR in May, 25 2018. Table 1 summarizes the main industries represented in our database. For each industry, Table 1 also provides an example of firms included in our list.

Additionally, our sample includes companies based in different countries such as USA, UK, Germany, Spain, Italy, France, etc. For each company, we included in the database the text of the email. The mean length of the text is 171.7 words (SD = 110.6).

Table 1: Re-Permission Emails by industry

Industry	Example of Companies Included	Percent
Entertainment, and Recreation	Disneyland, Spotify, Hulu	12%
Clothing, Shoes and Jewelery	Zara, Nike, Clarks, Pandora, Swarovski	10%
Education	City University London, The Case Center, AMA	9%
Browsers & Information Services	Google, Aruba, Doodle	9%
Retailing	Ebay, Selfridges, Aliexpress	8%
Travel & Tourism	Uber, KLM, British Airways, Tripadvisor	8%
Consulting & Marketing Services	Accenture, Pwc	6%
Newspapers, Books, Publishing	The Guardian, The Economist, McGraw-Hill	5%
Food, Restaurants and Wine	McDonalds, JustEat, Barilla	6%
Technology & Electronics	Samsung, Asus	3%
Beauty Sector & Blog	Lancôme, Sephora, EsteeLauder	3%
Finance and Insurance	American Express, Lloyds Bank	3%
Kids and Infancy	Pampers, Toys Center	2%
Social Media & Blog	Facebook, Instagram, Twitter, Linkedin	2%
Automotive	Audi, BMW, Mercedes-Benz	2%
Other	FitBit, Pampers, Waitrose & Partners	12%

3.2 Modelling approach

This database was analyzed through NLP techniques. More specifically we used the Latent Dirichlet Allocation (Blei, Ng, and Jordan, 2003) to analyze in an unsupervised fashion the data permission requests included in our database to detect a number of latent topics.

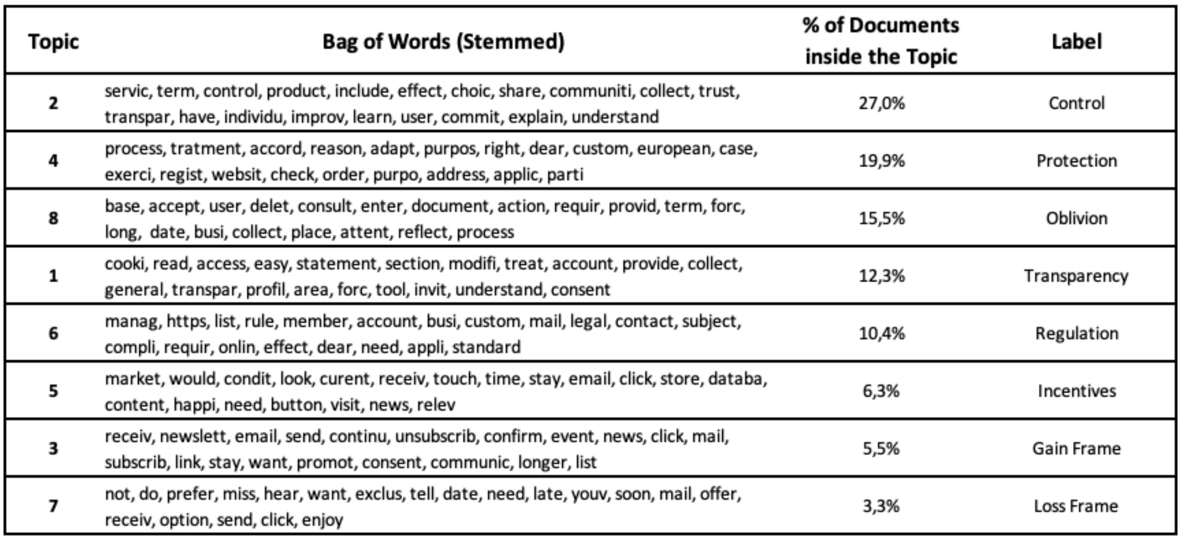
The topic modeling analysis has been conducted using the Gensim package available in Python. Firstly, the usual operation of data cleaning has been brought about: the texts has been firstly split into words (tokenization), then stop words have been removed, bigrams and trigrams have been created and words have been lemmatized. Additionally, numbers and words with a frequency higher than 50% or lower than 5% have been removed since too common or too rare to be useful for the analysis (Griffiths and Steyvers 2004; Lu, Cardie and Tsou, 2011) getting to a final dictionary of 191 unique words. In order to run a LDA model, it is necessary to specify, in an *a priori* fashion, the number of hidden topics to seek in the texts (k); consequently, in order to determine the optimal value for k we have estimated different LDA models with different number of topics and we have compared them using the coherence score measure. According to this measure, the best model is the one with three topics since it returns the highest coherence value score (0.55); however, given our task of capturing the differences which are present in companies' the e-mail, keeping only three topics does not allow us to capture also minor topics. Consequently, we decided to set the number of topics to eight and to, eventually, restrict the set to some coherent topics only, as suggested by computer science literature (Puranam, Narayan and Kadiyali, 2017; Mimno, Wallach, Talley, Leenders and McCallum, 2011; AlSumait, Gentle and Domeniconi, 2009). To evaluate the overlap among different topics we plotted the Jensen-Shannon divergence, which shows that the eight topics were well differentiated.

Table 2 shows each topic with the list of its 20 most probable words; the relevance of the words has been calculated accordingly to Sievert & Shirley (2014):

$$r(w, k|\lambda) = \lambda \log(\phi_{kw}) + (1-\lambda)\log\left(\frac{\phi_{kw}}{p_w}\right) \qquad (1)$$

where w indicates the word, k indicates the topic, ϕ_{kw} denote the probability of term w for topic k, p_w indicates the marginal probability of term w in the corpus and λ is a balance factor that we have set to 0.5 to give equal probability to the lift and the probability of term w for topic k (Sievert and Shirley, 2014). This allows to decrease the rankings of frequent words in the corpus and increase the relevance of more rare words.

Table 2: Topics Most Probable Words



Topic	Bag of Words (Stemmed)	% of Documents inside the Topic	Label
2	servic, term, control, product, include, effect, choic, share, communiti, collect, trust, transpar, have, individu, improv, learn, user, commit, explain, understand	27,0%	Control
4	process, tratment, accord, reason, adapt, purpos, right, dear, custom, european, case, exerci, regist, websit, check, order, purpo, address, applic, parti	19,9%	Protection
8	base, accept, user, delet, consult, enter, document, action, requir, provid, term, forc, long, date, busi, collect, place, attent, reflect, process	15,5%	Oblivion
1	cooki, read, access, easy, statement, section, modifi, treat, account, provide, collect, general, transpar, profil, area, forc, tool, invit, understand, consent	12,3%	Transparency
6	manag, https, list, rule, member, account, busi, custom, mail, legal, contact, subject, compli, requir, onlin, effect, dear, need, appli, standard	10,4%	Regulation
5	market, would, condit, look, curent, receiv, touch, time, stay, email, click, store, databa, content, happi, need, button, visit, news, relev	6,3%	Incentives
3	receiv, newslett, email, send, continu, unsubscrib, confirm, event, news, click, mail, subscrib, link, stay, want, promot, consent, communic, longer, list	5,5%	Gain Frame
7	not, do, prefer, miss, hear, want, exclus, tell, date, need, late, youv, soon, mail, offer, receiv, option, send, click, enjoy	3,3%	Loss Frame

As a robustness check, we also conducted a manual content analysis to validate results obtained through LDA. [1]This analysis confirmed the presence of the eight main topics as well as the labels we associated to each topic. Results highlight that *Control* is the topic most

[1] Two independent judges undertook this manual analysis. The Cohen kappa measure for inter-judgment reliability is 82%.

frequently used (27%) in re-permission emails. *Protection* (20%), *Oblivion* (15%), and *Regulation* (10.4%) topics indicates the choice of a legal type of communication mainly focused on legal aspects of the new GDPR. The topic of *Transparency* is also used (12.3%). *Gain/Loss frames* are present as well as *Incentives* but not frequently used in the re-permission email campaigns.

4. Study 2: The Experiment

The purpose of study 2 is to test the impact of the different topics on the probability of granting access to data beyond the GDPR. The previous analysis indicates that companies used eight main topics. Three of them are mainly legal aspects. *Control* and *Transparency* are not only the most frequent non-legal topics used but are also the most frequently mentioned in each single re-permission email.[2] Additionally, previous literature does not agree about the role of incentives, and its role have never been investigated in combination with other communication topics. Therefore, we decide to focus our experiment on the combined role of incentives and gain/loss frames.

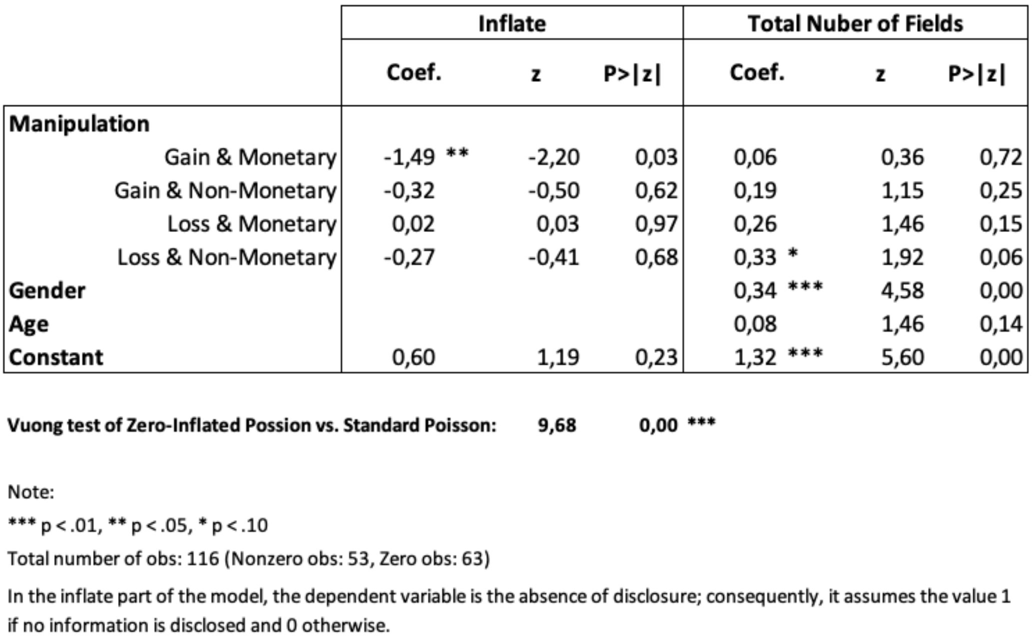
We designed a 5x1 between-subject experimental design where we manipulated two main features of the data request: framing (gain vs. loss) and incentives (monetary vs. non monetary incentives). We also included a control group message that does not have a framing in terms of gain or loss and does not have incentives. We considered a well-known clothing brand to create our scenario-based experimental conditions. Participants who completed the survey are 116 students of a main European university. Respondents were first randomly exposed to one of the five conditions, then they were asked to accept privacy terms. Finally, they were asked to fill a mandatory field (email address) and thirteen optional fields (e.g. gender, birth date, address, phone number, size for clothing, etc.). The acceptance of privacy terms, and the number of optional fields filled represent the focal variable of interest.

We analyze these data by estimating a zero-inflated Poisson model[3]. Results are provided in Table 3. Interestingly, the gain framed combined with the monetary incentive significantly increases the probability to accept the privacy terms and fill at least one field, however it does not increase the probability of providing additional information. By contrast, the loss frame combined with the non-monetary incentive is more effective in inducing individuals to provide the company more personal details.

Table 3: Zero-inflated Poisson Regression

[2] Due to space constrain we did not reported this result, but it is available upon request.

[3] We test the zero-inflated Poisson versus rival specifications



	Inflate			Total Nuber of Fields		
	Coef.	z	P>|z|	Coef.	z	P>|z|
Manipulation						
Gain & Monetary	-1,49 **	-2,20	0,03	0,06	0,36	0,72
Gain & Non-Monetary	-0,32	-0,50	0,62	0,19	1,15	0,25
Loss & Monetary	0,02	0,03	0,97	0,26	1,46	0,15
Loss & Non-Monetary	-0,27	-0,41	0,68	0,33 *	1,92	0,06
Gender				0,34 ***	4,58	0,00
Age				0,08	1,46	0,14
Constant	0,60	1,19	0,23	1,32 ***	5,60	0,00

Vuong test of Zero-Inflated Possion vs. Standard Poisson: 9,68 0,00 ***

Note:
*** p < .01, ** p < .05, * p < .10
Total number of obs: 116 (Nonzero obs: 53, Zero obs: 63)
In the inflate part of the model, the dependent variable is the absence of disclosure; consequently, it assumes the value 1 if no information is disclosed and 0 otherwise.

5. Conclusions

The aim of this paper was to gain a better understand on how firms can effectively ask for data permission. By collecting and analyzing a sample of 367 companies' re-permission emails sent before the introduction of the new GDPR, we were able to identify eight main topics used in of these communications. Legal topics (protection, oblivion and regulation) were frequently used highlighting that most businesses focused their message mainly on legal dimensions, in that missing the opportunity to encourage them to comply. The most frequently used non-legal topics are control and transparency that were also most frequently mentioned in each single re-permission email. Incentives, and framing (gain and/or loss) are the less frequently used topics. We directed our attention to these two topics (framing and incentives) and we conducted a lab experiment to examine their effectiveness in getting access to detailed personal information. We find that people react positively to messages that highlight the gains stemming from data disclosure and contain a monetary incentive (e.g. a discount) to grant a firm's access to non-mandatory personal details. Interestingly, we also show that by presenting data denial as a threat to access a firm's services and providing a non-monetary incentives people are more likely to grand permission to a larger number of personal information. Our paper provides a first empirical analysis of how firms managed the GDPR re-permission email campaigns. This work has implications for the literature on privacy by clarifying the impact of framing combined with different types of incentives. We show that gain and losses affect differently the likelihood of providing access to data and different levels of disclosure. Our findings have some relevant implications for marketers as well. Marketers can use the main findings of our study to more effectively design their on-going data requests beyond the call for re-permission associated with the GDPR.

6. References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. Science, 347(6221), 509-514.

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What is privacy worth?. The Journal of Legal Studies, 42(2), 249-274.

AlSumait, L., Barbará, D., Gentle, J., & Domeniconi, C. (2009). Topic significance ranking of LDA generative models. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 67-82). Springer, Berlin, Heidelberg.

Athey, S., Catalini, C., & Tucker, C. (2017). The digital privacy paradox: small money, small costs, small talk. National Bureau of Economic Research.

Blattberg, R. C. Byung-Do Kim, and Scott A. Neslin (2008), Database Marketing: Analyzing and Managing Customers. International Series in Quantitative Marketing. Springer Science & Business Media.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Chellappa, R. K., & Sin, R. G. (2005). Personalization versus privacy: An empirical examination of the online consumer’s dilemma. Information technology and management, 6(2-3), 181-202.

Goldfarb, A., & Tucker, C. (2012). Shifts in privacy concerns. American Economic Review, 102(3), 349-53.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National academy of Sciences, 101(suppl 1), 5228-5235.

Grossklags, J., & Acquisti, A. (2007). When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information. In WEIS.

John, L. K., Acquisti, A., & Loewenstein, G. (2010). Strangers on a plane: Context-dependent willingness to divulge sensitive information. Journal of consumer research, 37(5), 858-873.

Krafft, M., Arden, C. M., & Verhoef, P. C. (2017). Permission Marketing and Privacy Concerns—Why Do Customers (Not) Grant Permissions?. Journal of interactive marketing, 39, 39-54.

Ku, H. H., Yang, P. H., & Chang, C. L. (2018). Reminding customers to be loyal: does message framing matter?. European Journal of Marketing, 52(3/4), 783-810.

Lu, B., Ott, M., Cardie, C., & Tsou, B. K. (2011). Multi-aspect sentiment analysis with topic models. In 2011 11th IEEE International Conference on Data Mining Workshops (pp. 81-88). IEEE.

Martin, K. (2018). The penalty for privacy violations: How privacy violations impact trust online. Journal of Business Research, 82, 103-116.

Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data privacy: Effects on customer and firm performance. Journal of Marketing, 81(1), 36-58.

Milne, G. R., & Culnan, M. J. (2004). Strategies for reducing online privacy risks: Why consumers read (or don’t read) online privacy notices. Journal of interactive marketing, 18(3), 15-29.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In Proceedings of the conference on empirical methods in natural language processing (pp. 262-272). Association for Computational Linguistics.

Phelps, J., Nowak, G., & Ferrell, E. (2000). Privacy concerns and consumer willingness to provide personal information. Journal of Public Policy & Marketing, 19(1), 27-41.

Puranam, D., Narayan, V., & Kadiyali, V. (2017). The Effect of Calorie Posting Regulation on Consumer Opinion: A Flexible Latent Dirichlet Allocation Model with Informative Priors. Marketing Science, 36(5), 726-746.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces (pp. 63-70).

Tucker, C. E. (2014). Social networks, personalized advertising, and privacy controls. Journal of Marketing Research, 51(5), 546-562.

Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. Journal of Marketing, 80(6), 97-121.

White, T. B., Novak, T. P., & Hoffman, D. L. (2014). No strings attached: When giving it away versus making them pay reduces consumer information disclosure. Journal of Interactive Marketing, 28(3), 184-195.

White, T. B., Zahay, D. L., Thorbjørnsen, H., & Shavitt, S. (2008). Getting too personal: Reactance to highly personalized email solicitations. Marketing Letters, 19(1), 39-50.