AARHUS UNIVERSITY

Cover sheet

This is the accepted manuscript (post-print version) of the article.

The content in the accepted manuscript version is identical to the final published version, although typography and layout may differ.

How to cite this publication

Please cite the final published version:

Tudoran, A. A. (2019). Why do Internet consumers block ads? New evidence from consumer opinion mining and sentiment analysis. Internet Research, 29(1), 144-166. <https://doi.org/10.1108/IntR-06-2017-0221>

Publication metadata

Title:	Why do Internet consumers block ads? New evidence from consumer opinion mining and sentiment analysis
Author(s):	Tudoran, Ana Alina
Journal:	Internet Research, 29(1), 144-166
DOI/Link:	10.1108/IntR-06-2017-0221
Document version:	Accepted manuscript (post-print)
Document license:	

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

Citation: Tudoran A.A. (2018). Why do Internet consumers block ads? New evidence from consumer opinion mining and sentiment analysis, *Internet Research*, Forthcoming.

Why do Internet consumers block ads? New evidence from consumer opinion mining and sentiment analysis

Abstract

Purpose: The number of Internet consumers who adopt ad-blocking is increasing rapidly all over the world. The purpose of the paper is to evaluate this phenomenon by: (1) assembling the existing considerations and key theoretical aspects of the determinants of online ad-blocking; and (2) by exploring the consumers' beliefs and sentiments towards online ads and expected outcomes of ad-blocking behavior.

Methodology: Data consist of 4,093 consumers' opinions in response to the news items about ad-blocking, published by a leading news and technology website in the period 2010-2016. The unstructured data is analyzed using probabilistic topic modeling and sentiment analysis.

Findings: Five main topics are identified, unveiling the hidden structure of consumers' beliefs. A sentiment analysis profiling the clustered opinions reveals that the opinions that are focused on the behavioral characteristics of ads express the strongest negative sentiment, while the opinions centered on the possibility to subscribe to an ad-free fee-financed website are characterized on average by a positive sentiment.

Practical implications: The findings provide useful insights for practitioners to create/adopt more acceptable ads that translate into less ad-blocking and improved Internet surfing experience. It brings insights on the question of whether ad-free subscription websites have or do not have the potential to become a viable business opportunity.

Originality: The research (1) improves the current understanding of the determinants of ad-blocking by introducing a conceptual framework and testing it empirically; (2) makes use of consumer-generated data on the Internet; and (3) implements novel techniques from the data mining literature.

Keywords: ads, ad-blocking, consumer, topic modeling, sentiment analysis, text-mining.

1. Introduction

New forms of online ads are emerging to compete for consumers' attention. Ads are a necessary evil of the Web, but the consumers may not find them that evil, especially if they conform to acceptable ads guidelines (Broida, 2015, p. 32). Online ads can help companies to increase long-term brand awareness (Drèze and Hussherr, 2003) and can play a significant role in brand recognition (Keller, 2010; Madhavaram and Appan, 2010) and purchase decisions and sales (Lewis and Reiley, 2014). Online ads can facilitate more targeted advertisement and personalized ads adapted to consumer needs and purchase intent, provided that they do so in a non-intrusive and safe way (Edwards *et al.*, 2002; Goldfarb and Tucker, 2011a; Goldfarb and Tucker, 2011b). It is, however, increasingly recognized that some online ads are invasive and drive many consumers to seek ways to block them (Verlegh *et al.*, 2015). In the digital world, the growing popularity of ad-blocking is a serious problem (Vratonjic *et al.*, 2013). The number of Internet consumers adopting ad-blocking has grown worldwide over the past five years. According to the 2015 Global Report from (PageFair and Adobe, 2015), ad-blocking usage in Europe alone grew by 35% during 2015 and reached 77 million monthly active consumers. An increased interest in ad-blocking is also reflected in the data from Google Trends (2016), where the worldwide search interest volume in the phrase 'adblock' rose by 1,950% between 2004 and 2016 (Figure 1). A recent U.S. survey by Securities and Optimal.com (2016) reveals that about 45% of those who do not use ad-blocking today report that they were not aware of its existence. As the Internet consumers become increasingly conscious of the possibility of ad-blocking, even higher adoption rates must be expected globally.¹

[Figure 1 here]

With a view to discourage the ad-blocking behavior, the ad industry has embraced several strategies. Parts of the ad industry have questioned the legitimacy of the ad-blocking, claiming that

¹ The adoption rate of ad-blockers has changed significantly with the launch of the mobile platform

consumers have an implicit *quid pro quo* contract with the website and should accept ads in exchange for web content (Goldstein *et al.*, 2014). Other initiatives like the one promoted by the article ‘Why Ad-blocking is devastating to the sites you love’ (Fisher, 2010) have focused on stimulating the consumer empathy with the online publishers and have designed emotional messages to change the consumer affect. Magazines like Forbes and GQ have taken more radical steps to force a behavioral change by preventing ad-blocker consumers from viewing the contents of the websites before the ad-blocker has been de-activated. But none of these actions have been sufficiently effective in overcoming consumer resistance and, in some cases, they have actually caused a significant number of individuals to stop reading the websites in question (Dvorkin, 2016). Consumers’ generalized ad avoidance can ultimately carry forward to no revenue for content providers, and content would not be funded (Anderson and Gans, 2011). In a less extreme situation, ad-blocking may cause content providers to impose fees for the content, reduce the quality of the content, and raise the advertising clutter for ad-viewers less sensitive to ads in order to maintain revenues (Anderson and Gans, 2011). Using a consumer science approach, this paper analyzes how consumers experience online advertising and why they use ad-blocking. The objectives of this study are to assemble and explain the key theoretical aspects of the determinants of online ad-blocking and to bring empirical evidence on the consumers’ beliefs and sentiments to validate the theoretical assumptions.

This research contributes to the literature on online consumer behavior in several ways. First, there is limited scientific literature exploring consumers’ cognitions and affect with respect to online ads in the context of ad-blocking behavior. The research in advertising has almost exclusively focused on the ad effectiveness to attract attention, comprehension, search, ad-clicking, or conversion (Burns and Lutz, 2006; Kireyev *et al.*, 2015; Lewis and Reiley, 2014; Li and Kannan, 2014; Xu *et al.*, 2014; Fang *et al.*, 2007; Sundar and Kalyanaraman, 2004) (to cite just a few). One recent branch of theoretical studies in information economics addresses the economic and legal consequences of ad-blocking and the content providers’ design of countermeasures to prevent ad-blockers (Clifford and Verdoodt, 2016; Vratonjic *et al.*, 2013; Vallade, 2009; Anderson and Gans, 2011). Another branch, following Cho and Cheon (2004), develops structural models based on surveys to explain the consumer tendency to avoid ads in general (Seyedghorban *et al.*, 2016) or specific types of ads in

particular, such as unsolicited commercial e-mails (Baek and Morimoto, 2012). Singh and Potdar (2009) bring insights into the reasons behind blocking online ads by reviewing previous research on ad-avoidance, and a few studies analyze the effect of specific factors such as online privacy and ad animation (e.g. Baek and Morimoto, 2012; Youn, 2009; Wirtz *et al.*, 2007; Neben and Schneider, 2015). Because the few studies who refer to online ad-blocking commonly consider the determinants in isolation, this research aims to provide an integrated picture of the main determinants and test it empirically.

The second contribution of the study is the implementation of text-mining techniques on consumer-generated data from the Internet to explore consumers' opinions. We test our propositions on a set of 4,093 opinions published by a leading news and technology website in the period 2010-2016. Extracting information from online consumers' opinions is a relatively new approach in consumer studies that should facilitate a finer representation of the users' cognitions than focus-group-based methods using a limited number of subjects. The dimension of the new types of data require analytical techniques developed in the area of data mining to yield connections between and within thousands of opinions; connections that are impossible to establish by hand (Blei and Lafferty, 2009). We use two specific text-mining techniques: probabilistic topic modeling (Blei *et al.*, 2003) to identify the main topics pervading consumers' beliefs towards ads and ad-blocking behavior; and sentiment analysis (Taboada *et al.*, 2011; Jurka, 2012) to evaluate the sentiment produced by the feelings associated with online ads and ad-blocking.

Third, the present paper adds to previous research that calls for interventions aimed at improving the quality of user experience in relation to future Internet use (Clifford and Verdoodt, 2016; Goldstein *et al.*, 2014) and at defining the legal framework of advertising for protecting consumers' privacy and security on the Internet (Wirtz *et al.*, 2007). The analysis aims to raise the awareness and provide website owners with empirically-based recommendations for choosing better forms of advertisements in order to stimulate the website users' long-term satisfaction and entice them to return (Neben and Schneider, 2015).

2. Research background

In the advertising literature, ad-avoidance is defined as “*all actions by media users that differentially reduce their exposure to ad content*” (Speck and Elliott, 1997, p. 61). Previous research has documented that the ad-avoidance is part of everyday media style. For example, it is well-known that many consumers zap television commercials, leave the room, or switch radio channels as the rule rather than the exception when they are exposed to ads (Speck and Elliott, 1997; Cronin and Menelly, 1992; Abernethy, 1991). All broadcast media are affected by ad-avoidance, but over the last years, ad-avoidance has proliferated tremendously in the online setting. On the Internet, the ad-avoidance concept was advanced by Cho and Cheon (2004) who conceptualized it as a three-dimensional – cognitive, affective and behavioral – response to advertising stimuli. By definition, then, ad-avoidance is an attitude (Fishbein and Ajzen, 1977) – a learned predisposition and/or a state of mind of an individual to evaluate and respond either in a favorable or unfavorable manner towards a specific object, i.e. online ads. The cognitive components of ad-avoidance are evaluative judgments consisting of the beliefs about the attributes of the ads and the expected outcomes of avoiding the ads, and the evaluation of the expected outcomes. An important function of the beliefs is that, if negative, they lead to an intentional or automatic psychological mechanism of eluding dealing with the object, e.g. by deliberately or unconsciously ignoring the object. However, behavior is not only determined by cognitions, but also by feelings and emotions (Eagly and Chaiken, 1998). A consumer’s negative feeling and emotional reaction associated with an ad is defined as the affective component of ad-avoidance (Cho and Cheon, 2004). The affective avoidance involves the development of affective states of mind towards the ads based on (un)satisfying exposure to the ads. The behavioral ad-avoidance is the actual engagement in actions to avoid the ads (Speck and Elliott, 1997). Tang *et al.* (2015) note that the avoidance behavior can be active or passive. Passive avoidance behavior refers to the action that requires little behavioral effort to get away from the online ads, such as looking away from the ads and waiting for the ads to go away by themselves. Active behavioral avoidance of ads refers to the effortful actions undertaken by consumers to escape or get rid of online ads. Examples of active ad-avoidance behavior include: closing the ad, abandoning the website, and ad-blocking. Ad-blocking is defined as a conscious action to avoid online ads by installing a computer program that automatically blocks the ads.

Drawing on the social-psychology theory (Fishbein and Ajzen, 1977; Eagly and Chaiken, 1998), the cognitive-affective system of ad-avoidance can be viewed as a foundation to explain ad-blocking behavior. The purpose of this work is to identify the specific beliefs about the attributes of the ads or the expected outcomes of blocking the ads, and the evaluation of the expected outcomes that have to be perceived by the individual in order to behave. By reviewing existing literature, we define the key elements of this behavior and explain it in more details in the next section.

Online ads characteristics

The online ads characteristics refer to the attributes of the online commercials. In recent years, online advertising has evolved particularly with respect to the behavioral characteristics, e.g. interactivity, display duration, mouse over, movement, frequency, and onset timing, to include more obtrusive features that strive to be highly visible relative to the site content and to stimulate more purchases (Goldfarb and Tucker, 2011b). Analyzing the advertising clutter, Ha and McCann (2008) note that ad intrusiveness has become one of the most important features of online ads because advances in advertising technology have facilitated forced exposure to advertising. Online ads normally aim at capturing consumers' attention; however, most often they distract people from their planned activities (Tang *et al.*, 2015, p. 519). Ad intrusiveness is the degree to which ads interrupt the flow of an editorial media content unit (Ha and McCann, 2008; Neben and Schneider, 2015). Prior research indicates that ad intrusiveness makes it difficult for consumers to process the information, decreases task performance and recall, and it is generally associated with negative attitudes towards using the website (Burke *et al.*, 2005; Edwards *et al.*, 2002; Hong *et al.*, 2004; Hong *et al.*, 2007; Goldfarb and Tucker, 2011a; Neben and Schneider, 2015).

The Information Theory (Shannon and Weaver, 1948) helps to explain why the intrusive characteristics of ads determine most of the ad-avoidance reaction on the Internet (Goldfarb and Tucker, 2011a; Seyedghorban *et al.*, 2016). The tenets of the Information Theory suggest that when the information is perceived to progress in an unexpected direction, motivating the individual to pay particular attention, the interference between the ad and the individual's goal impedes the process of obtaining the information and the ad will be perceived as a noise. Goal impediment leads to

perceptions of search hindrance, disruption, and distraction, and except for the case where the consumer's goal is to search for the products, these perceptions increase consumer resistance to ads independently of perceptions of ad utility or incentives associated with the ad message (Seyedghorban *et al.*, 2016). For example, drawing on the model of advertising avoidance proposed by Cho and Cheon (2004) and a replication study by Seyedghorban *et al.* (2016), evidence reveals that perceived goal impediment has the strongest effect on the cognitive, affective, and behavioral ad-avoidance. With a focal emphasis on the cognitive ad-avoidance, other studies also find that disturbance, distraction and information overload significantly affect memory, attention, information processing, reading, and search time (Edwards *et al.*, 2002; Goldstein *et al.*, 2014; McCoy *et al.*, 2008; Zha and Wu, 2014).

The avoidance reaction is not only determined by the cognitive processes, but also by the emotional stimulation that interacts with cognitions and arises with or without conscious control (Scherer *et al.*, 2001). Consistent with the Psychological Reactance Theory (Brehm, 2009), by forcefully obstructing the consumers' view of the content or by allowing them to close a certain ad only after a certain amount of time, the ads create a negative reaction induced by the perceived lack of control or violation of freedom to do things (Ha and McCann, 2008). A study based on neurophysiological data of the autonomous nervous system (Neben and Schneider, 2015) reveals that the ability to control and the belief that the control over what they see is not in their hands lead to significant frustration, physiological stress, and negative affect. Thus ad-avoidance behavior is, to a large extent, a result of the interaction between the cognitive processes (appraisal of the characteristics of the ads) and the emotional arousal activated by physiological stress and the cognitive interpretation of the arousal. The appraisal of the intrusive characteristics of online ads is expected to significantly govern consumers' beliefs and to be associated with a significant negative affect and ad-blocking.

Hence, it is proposed that:

H_{1a}: The perceived intrusive characteristics of online ads pervade the consumers' beliefs on ads and ad-blocking.

H_{1b}: The perceived intrusive characteristics of online ads are associated with a significant negative sentiment.

The hidden costs of online ads

The hidden costs of online ads are expenses not explicitly included in the purchase price of the Internet access. Technical studies on the performance of computers and mobile devices in the presence of ads (Gui *et al.*, 2015; Gao *et al.*, 2016) report that the online ads incur various hidden costs for the consumers, who indirectly pay to download the ads to their platforms in order to view or interact with the content of the website. For example, the amount of bandwidth consumed for downloading content with ads is calculated to be between 25% to 35.06% of the total available bandwidth (Gui *et al.*, 2015). The decrease in the available bandwidth for downloading useful content leads to increased download time with particularly negative effects for the slow Internet connections (Singh and Potdar, 2009). Also, it is estimated that the apps with ads consume on average 48% more CPU time, 16% more energy, and 79% more network data than mobile apps without ads (Gui *et al.*, 2015).

In the advertising literature, there is limited evidence about the relevance of the hidden costs of online ads on consumers' behavior. One exception is a commercial survey among 1,543 Adblock Plus users (Pallant, 2011) showing that increased loading time and bandwidth consumption are the second most important reasons to avoid ads and install ad-blockers. Consistent with the utility maximization model (Becker, 1996), any factor that impacts on the value of the expenditure, in addition to availability of desired content, is perceived as a burden by the consumers (Seyedghorban *et al.*, 2016). Therefore, it could be expected that the hidden perceived costs of the online ads would determine a lower perceived value of the Internet access and thus trigger ad-blocking in consumers. The following statement is proposed:

H_{2a}: The perceived hidden costs of online ads pervade the consumers' beliefs on ads and ad-blocking.

H_{2b}: The perceived hidden costs of online ads are associated with a significant negative sentiment.

Consumer power and empowerment

"Power" is a key human concern that constantly influences behavior and constitutes a fundamental component of social systems and hierarchies (Labrecque *et al.*, 2013). In the Internet, power is defined as the perceived asymmetric ability to control people or valued resources in online

social relations (Rucker *et al.*, 2012; Labrecque *et al.*, 2013). Power involves an interaction between two or more parties, where the valued resources (e.g. actual monetary resources, ability to reward/legitimate authority, or intellectual capital) must be important to at least one of the parties (Rucker *et al.*, 2012). Empowerment refers to the dynamic process of gaining power through action by changing the status quo in current power balances (Labrecque *et al.*, 2013).

In the Internet, the actors get empowered by multiple sources (see a review by Labrecque *et al.*, 2013), but information-based power is particularly relevant in the context of this research. The Internet has originally empowered consumers through increased information consumption, production, and choice. Marketers empowered themselves with the right of collecting and using the information that consumers produce in relation to their online patterns. Such empowerment occurred by tracing and tracking the consumers' behavioral patterns and personal information in order to deliver more targeted advertisements (Leenes, 2015). The asymmetric distribution of power has eroded the consumer power perceptions and made consumers determined to take active steps to reduce transparency and preserve their power (Labrecque *et al.*, 2013). Consistent with the theory of power-responsibility equilibrium (PRE) (Emerson, 1962), if one party chooses a strategy of greater domination, the other party will take defensive actions to reduce the power. In the online context, the power holders in this exchange are the marketers who know far more about the consumers through behavioral tracking, while the consumers are the less dominant party who know nothing about how the business models use their personal information and the risks and benefits of sharing this information (Clifford and Verdoodt, 2016). When the power holders are dominating or are not viewed as socially responsible, the consumers would be motivated to generate balancing actions. From this perspective, the ad-blocker is not only seen as a tool for improved online experience or reduced costs; rather it is seen as a more general manifestation of consumer attempting to regain power balance or empowerment (Clifford and Verdoodt, 2016; Wirtz *et al.*, 2007). Notably, the perceived legitimacy of ad-blocking is expected to facilitate the consumer empowerment and defensive reaction, i.e. ad-blocking adoption. Although the industry claims to have an implicit *quid pro quo* contract, according to which the consumers should accept ads in exchange for web content (Goldstein *et al.*, 2014), such an agreement does not exist from the consumer perspective (Clifford and Verdoodt, 2016, p. 8). It is therefore proposed that:

H_{3a}: The perceived legitimacy of ad-blocking and consumer power pervades the consumers' beliefs on ads and ad-blocking.

H_{3b}: The perceived legitimacy of ad-blocking and consumer power is associated with a significant negative sentiment.

Online privacy and data security

Tracking services are widely used by the advertising networks to collect and correlate several pieces of information that can identify the consumer interests and predict their preferences for targeted advertising. Tracking has significant privacy implications for consumers. Consumer privacy is defined as the ability of individuals to seclude information about themselves or to control when, how, and to which extent their personal information is to be transmitted to others (Youn, 2009; Wirtz *et al.*, 2007). Privacy concerns are not only determined by the disclosure of the information per se and loss of anonymity but more importantly by the uncertainty associated with how personal information is used by the other parties and for which purposes it is used (Bart *et al.*, 2005). This uncertainty may arise from a myriad of potential actions that could lead to disadvantages in real life (Thode *et al.*, 2015), such as annoying commercial solicitations, ads with embarrassing or suggestive content, a less favorable offer, denial or benefit (Agarwal *et al.*, 2013; Mayer and Mitchell, 2012; Ur *et al.*, 2012; Youn, 2009). The privacy concerns vary widely across cultures (Ur *et al.*, 2012; Youn, 2009) and depend on both external factors (Zhou, 2017) and internal factors such as personality traits, trust (Lwin *et al.*, 2016), and knowledge (Agarwal *et al.*, 2013). For instance, if consumers do not possess the basic knowledge of advertising tracking methods and technologies, privacy is not of concern (Youn, 2009). However, as they become more knowledgeable, privacy beliefs about online ads motivate the consumers to adopt risk-protective behaviors such as seeking advice and refraining from using certain websites (Youn, 2009; Wirtz *et al.*, 2007). Because ad-blockers not only block ads, but also prevent the mechanisms that deliver tracking from being installed on the consumer's equipment, the privacy concerns can be expected to be associated with ad-blocking behavior. At the same time, the extent to which consumers perceive that the online ads expose them to malicious software (e.g. adware, malware, spyware), which can access or harm their devices may also influence the decision to adopt

ad-blockers. Drawing on the protection motivation theory (Rogers, 1975), the individual's belief that he/she can protect himself/herself from a risky behavior and the perception that the protective behavior is effective in reducing risk, determine the consumer risk protective behavior. Based on these arguments, it is expected that:

H_{4a}: The perceived online privacy and security pervades the consumers' beliefs on ads and ad-blocking.

H_{4b}: The perceived online privacy and security is associated with a significant negative sentiment.

4. Methodology

4.1. Sample

A total of 4,093 consumer reviews linked to the news items about “ad-blocking” published during 2010-2016 and publicly available from arstechnica.com² were retrieved through an automated retrieval procedure. The search procedure using the key terms ‘ad-blocking’ and ‘adblocking’ in the ‘Business’ area of the website, returned fourteen relevant news items published within the time frame 2010-2016 (Table 1).

[Table 1 here]

4.2. Data pre-processing

To prepare the text data for analysis, various cleaning tasks were performed using a personalized software that runs on Node.js (Tilkov and Vinoski, 2010) and the ‘SnowballC’ package in R (Bouchet-Valat, 2014). Specifically, we: i) removed opinions with less than fifty characters; ii) converted all text to lowercase; iii) removed punctuation and problematic symbols; iv) removed the numbers contained in the text; v) removed uninformative words such as prepositions and conjunctions; vi) removed extra whitespaces; vii) removed web links occurring in the original text; viii) removed the

² Ars Technica is a high-ranking technology news website in USA, which in 2010 temporarily adopted an anti-ad-blocking strategy that prevented readers who used ad-blocking software from viewing their site (McGann, 2010).

terms ‘adblocking’, ‘adblock’, ‘adblocker’, ‘ad-blocking’, ‘ad-block’, and ‘ad-blocker’ to increase the discriminative information due to presence in almost all the opinions; ix) stemmed the document to retain only the stem of a selection of similar words. Following the data pre-processing, the remaining corpus included 10,627 words of which 2,325 (about 22%) occurred more than once. The 10 terms with the highest document frequency were: site, block, content, people, advertise, pay, time, consumer, flash, and read.

A document-term matrix was created using the inverse document frequencies transformation function (tf_{idf}) (Robertson, 2004) to represent the raw data set numerically and build a proper data structure. The tf_{idf} value increases proportionally with the number of times a word appears in the document, but this is offset by the frequency of the word in the corpus, which makes it possible to adjust for uninformative words appearing more frequently in general.

$$tf_{idf} = tf \times \log \frac{N}{m_n} \quad (1)$$

where tf = term frequency; N = total number of documents; m = total number of documents containing the n^{th} term. Terms obtaining a tf_{idf} below 0.10 were removed. This process led to a reduced document-term matrix of 4,092 documents and 9,382 terms to be used for estimating the model.

4.3. Topic modeling analysis

To test the hypotheses H_{1a} - H_{4a} , the Latent Dirichlet Allocation model (LDA) (Blei, 2012; Blei *et al.*, 2003) was applied to catch the most important topics within the consumers’ opinions and their distribution across all the documents. LDA is the most frequently used topic-modeling algorithm.³ A

³ We also implemented a correlated topic model (CTM) (Blei and Lafferty, 2007), which constraints the different topics to be weak correlated. We expected that if topics were indeed correlated, CTM would reduce the overlapping between the topic distributions. However, the results using CTM for our data made the learned topics less interpretable, i.e. the topic distributions could not be clearly associated to a semantic meaning and numerous semantic relevant top-words repeated in several topics.

topic is formally defined as a distribution over a fixed vocabulary (Blei, 2012). The set of topics are defined on that vocabulary, albeit the words of each topic have different probabilities associated. For instance, the Internet security topic has words associated with ‘spyware’ and ‘malware’ with higher probability than other words in the vocabulary. LDA thus assumes that all the documents in a collection share the same set of topics and that each document exhibits these topics to a varying degree (Blei, 2012).

The LDA model implies a joint probability distribution over the observed variables (i.e. the words of the document) and the hidden random variables (the topic structure) *cf.* equation (2):

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{k=1}^K p(\beta_k) \prod_{d=1}^D p(\theta_d) \left(\prod_{n=1}^N p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right) \quad (2)$$

where:

K = the number of topics, where each β_k is the distribution over the vocabulary for topic k ;

θ_d = the topic proportions for the d^{th} document; $\theta_{d,k}$ = the topic proportion for topic k in document d ;

z_d = the topic assignments for the d^{th} document; $z_{d,n}$ = the topic assignment for the n^{th} word in document d ;

w_d = the observed words for the d^{th} document; $w_{d,n}$ = the n^{th} word in document d .

The LDA model estimation consists of the following three steps:

- Determining β_k , the term distribution over the vocabulary for topic k

$$\beta_k \propto \text{Dirichlet}(\delta) \quad (3)$$

- Determining θ_d , the proportions of the topic distribution in document d

$$\theta_d \propto \text{Dirichlet}(\alpha) \quad (4)$$

- For each of the N words, w_n
 - Choose a topic $z_n \propto \text{Multinomial}(\theta)$.

- Choose a word w_n , conditioned on the topic z_n : $p(w_n | z_n, \beta)$.

The model to be estimated is expressed as the conditional distribution of the hidden variables given the observed variables, *cf.* equation (5):

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D} | w_{1:D}) = \frac{p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})}{p(w_{1:D})} \quad (5)$$

The denominator in equation (5) is an integral whose value cannot be computed analytically. Efficient sampling-based algorithms such as Gibbs sampling allow to sample from the posterior distribution that asymptotically follows $p(\beta_{1:K}, \theta_{1:D}, z_{1:D} | w_{1:D})$ and to obtain a conceptually straightforward approximation of the model parameters (Andrieu *et al.*, 2003). This approximation is based on the Markov Chain Monte Carlo technique, which implies a probabilistic walk through a state space whose dimensions correspond to the variables in the model and where the next state (value) visited for each variable depends only on the current state. The initialization and the number of iterations chosen in the Gibbs sampling produce values for each of the variables in the model. After a certain number of iterations, the Gibbs sample reaches a point where all the values of the estimated parameters are coming from the stationary distribution of the Markov Chain. The approximated value of each parameter (which is reported in the results section) is calculated as the average taken over iterations considered after the ‘the burn-in’ period and only for every L^{th} value to avoid autocorrelation problems (Andrieu *et al.*, 2003).

4.4. Selecting the number of topics

The specification of the number of topics to extract (K) should ensure meaningful and interpretable topics and, most importantly, a number of topics that are robust to perturbations in the original data. Choosing a value of K that is too low generates topics that are excessively broad, while choosing a value that is too high results in over-clustered data (Greene *et al.*, 2014). There are several algorithms used to select K in a systematic way (De Waal and Barnard, 2008; Steyvers and Griffiths, 2007). We used the term-centric stability approach proposed by Greene *et al.* (2014), where different values of K are analyzed and then the K value with the highest stability is chosen and used as the value for the number of topics. Explicitly, given a range of K values (e.g. 2 - 20), the term-centric stability

measure fits 18 topic models, one for each value of K . For each K value candidate, a number of sub-samples (e.g. 10) would be generated from the dataset. Then 10 topic models are generated, one for each sub-sample, which results in 180 topic models in total. As topic models are described via lists of top-ranked terms, one ranked list for each topic, the method uses the similarity between these term rankings over the 10 runs of the same algorithm on data from the same source to estimate the term-centric stability score. The stability score for each K is calculated as an average Jaccard measure among all 10 topic models. The average Jaccard measure provides a way of comparing the agreement of the outputs from two different runs of a modeling algorithm when the outputs are represented as term ranking lists. The topic model with the highest stability score is the most stable and is selected for further interpretation.

5. Estimation and results

5.1. Term-centric stability procedure results

We employed the term-centric stability procedure and set $K = 2 - 20$ based on the rationale that there were fourteen news items that the consumers might specifically reflect upon, and in addition, they may discuss other topics that they consider relevant but unrelated to the news. To ensure a fair degree of detail, but still keep the analysis practically feasible, $K \leq 20$ was considered. Figure 2 illustrates the stability scores. Excluding the model with two topics (it would only group the documents very coarsely) (Hornik and Grün, 2011), $K = 5$ registered the highest stability score. Thus, $K = 5$ was selected for parameterizing the LDA model.

[Figure 2 here]

5.2. LDA estimation and results

The topic modeling analysis was performed in R with the packages ‘tm’ (Meyer *et al.*, 2008) and ‘topic models’ (Hornik and Grün, 2011). The document-term matrix served as the input for the LDA function with the Gibbs sampling. Estimation using Gibbs required a specification of the values of the hyper-parameters of the prior distribution. The default values were set to 0.1 for the parameter

of Dirichlet prior on the per-topic word distribution, β , and $50/K$ for the parameter of Dirichlet prior on the per-document topic distribution, θ . The burn-in parameter was set to 2,000. Following the burn-in period, 2,500 iterations were performed taking every 100th iteration for approximating the final value. Furthermore, by setting different starting points, three independent runs were used to check the stability of the results. Table 2 summarizes the five topics identified.

[Table 2 here]

Topic 1 contains words such as ‘advertising’, ‘annoy’, ‘animation’, ‘video’, ‘sound’, ‘intrusive’, as well as words related to actions such as ‘click’, ‘start’, ‘turn’, ‘show’, ‘block’, confirming the relevance of the intrusive characteristics of online ads on the consumer perceptions about online ads and ad-blocking (H_{1a}). We examined some of the representative comments to which the fitted model had assigned a high proportion of topic 1 (more than 60%) and found that many comments refer to the dynamic features of ads and the consumers’ attitude towards ads if they do not interfere with the consumers’ access to the web content. For example, one consumer complains about “*animated or video ads*” and about feeling that his concentration “*is being assaulted and harassed*”. However, he says that “*if your ads are static [...] then OK. I will not block them and might even read them*”. Another consumer says that “*Internet ads are annoying, they're often distracting and obnoxious*” but “*if content providers actually adopted static, unobtrusive advertising [...] I would probably consider whitelisting certain ad-supported websites*”.

Topic 2 contains keywords such as ‘pay’, ‘Internet’, ‘cost’, ‘data’, ‘bandwidth’, ‘charge’, and ‘connect’. This topic summarizes the consumers’ opinions about how the online ads impact the Internet bandwidth usage, browser speed, and other performance-related costs. Some of these concerns are found among the consumers who are on a limited data plan or without a data plan, living in certain areas with limited bandwidth, and mostly among mobile consumers. Some representative comments to which the fitted model has assigned a high proportion (more than 60%) of topic 2, read as follows: “*I don't see ads now, but since I pay for the data I consume, I'm perfectly valid in determining what data I consume & pay to have delivered*”. Another consumer says: “*Epecially for when people are*

browsing the Internet on cellular data connections with small data caps, having to load megabytes of Flash ads not only slows your loading speeds to a crawl, but is costing people a significant amount of money". The results confirm H_{2a}, that the perceived hidden costs associated with online ads make up one of the main topics that pervade the consumers' beliefs on ads and ad-blocking.

Topic 3 includes high-frequency terms such as 'article', 'read', 'comment', 'subscription', 'good', and 'support'. This topic was not expected according to our theoretical framework. A closer examination of the opinions to which the model assigned the highest proportions indicated that the topic was generated by comments that focused on the possibility to subscribe to an ad-free website as an alternative to using the ad-blockers. Some comments to which the fitted model has assigned the highest proportion (36.1%) of topic 3, read as follows: "*I fear if Ars increases the number of intrusive ads, like the egregious LG ad, to combat ad-blockers they might alienate people who might, like me, take time to appreciate Ars and seriously consider subscribing after a long perusal period. Perhaps Ars should offer different levels of subscription, i.e. the current no ads one and a second one involving no ads, tracking, analytics, etc.; just the content and nothing else*". Another consumer says "*I'm not entirely against subscriptions - when Pandora asked me to start paying if I was going to continue consuming voraciously, I accepted. Everything Pandora One offers I appreciate though - no ads, higher quality sound, a separate application, and longer timeouts*".

Topic 4 is defined by keywords such as 'website view', 'reason', 'argument', and 'legal' as well as keywords related to actions such as 'agree', 'understand', 'change', and 'specify'. While not particularly self-explanatory keywords, by reading the opinions to which the model assigned the highest proportions, we discovered that the core of the topic was related to the consumers' thoughts about their right to view the website when using ad-blockers under certain terms and conditions. A representative comment reads as follows: "*Well so far no one I know has been able to point out what law I'd be breaking when blocking adverts. I don't think it's against the law as such to break T&Cs (i.e. website terms and conditions). I think it just defines the circumstances under which the website owner might revoke 'privileges' like your ability to browse the page. It is difficult to prove that someone has read, understood and agreed to the T&Cs and there is no law, that specifically grants someone the right to have their website viewed in a certain manner, or that limits people to viewing a*

website in its original ‘intended’ form.” Thus, we related this topic with the consumers’ perceived legitimacy of blocking the ads and the manifestation of their power and right to choose, confirming H_{3a} .

Topic 5 is defined by keywords such as ‘flash’, ‘block’, ‘whitelist’, ‘track’, ‘security’, ‘malware’, and ‘trust’. Confirming H_{4a} , we find that the privacy and security perceptions play a significant role in the consumer opinions about online ads and ad-blocking. Some comments describe that: *“If you want to fix advertising, start by delivering ads based upon the content of the page being visited. Do not track individual consumers and do not use any data created by the consumer (e.g. the browsing history). In other words, address privacy”, and “I don’t block ads on any sites [...]. I do block, disable or never even install Flash [...]. I will continue to block Flash but not ads in general. Flash can cause my browser to consume huge numbers of CPU cycles. Flash is not secure; it’s a huge malware and security vulnerability”.*

To evaluate the extent to which the subject of a news item is reflected in the consumers’ opinions, the proportions of each topic in relation to each individual news item were analyzed. This analysis also served as a validity check of the finding to ensure that the topic model was able to accurately recognize the news topic from the consumers’ comments. For example, looking at news item C5 (‘France’s second-largest ISP suspends ad-blocking for now’, which is focused on the Internet service providers beginning blocking ads at the network level because of bandwidth consumption), Figure 3b indicates a relatively high proportion of topic 2 (the performance costs) reflected in the consumers’ opinions underlying C5. However, it is evident that the five latent topics identified are rather evenly distributed across the fourteen news items, indicating that the consumers’ concerns about these topics persist throughout the news items, regardless of the specific news content.

[Figure 3 here]

5.3. Identification of opinion clusters and sentiment analysis

By definition, topic analysis assumes that no document is “pure” but contains different slices of the K topics albeit with varying distribution. Therefore, to be able to infer the consumers’ sentiment with respect to the topics discussed (hypotheses H_{1b} - H_{4b}), we first identified a typology of opinions

based on their topic distribution similarity. Clustering the opinions also provided an understanding of the characteristics and size of the clusters with similar opinions. For this purpose, we used different clustering techniques and observed comparable results. In the classical Ward hierarchical clustering analysis (Hair *et al.*, 2006), the dendrogram indicated the 4, 5, and 6-cluster solutions as three possible candidate models. For the K-means clustering (Wedel and Kamakura, 2002), the decrease in the within-group's sum of square distance errors proved the 6-cluster solution as being the most probable model. For a model-based clustering (Fraley and Raftery, 2002), considering the ICL-BIC criterion, which penalizes weakly separated clusters (McLachlan and Peel, 2000), the best model was the ellipsoidal equal shape and orientation model with 6 clusters. The empirical analysis reported in Table 3 uses the K-means cluster solution with significant evidence on the statistical differences between the clusters for the five clustering variables (Hollander *et al.*, 2013). The clusters were labeled based on the topics where members of the cluster had an average score one-half of a standard deviation above the total average or, if no topic was above average, the most prominent topic.

[Table 3 here]

Thus, we find that cluster 1 represents 12% of the sample and is characterized by a significant number of opinions discussing topic 3 'Subscription'. Cluster 2 represents 9% of the sample and compared to the other clusters, it assigns much more importance to topic 2 'Performance costs'. Cluster 3 represents 13% of the sample and is characterized by opinions scoring predominantly high on topic 5 'Security and Privacy'. Cluster 4 represents 12% of the sample and has significantly high levels on topic 1 'Annoying ads characteristics'. Cluster 5 represents 7% of the sample and is characterized by opinions scoring predominantly high on topic 4 'Perceived legitimacy of ad-blocking and consumer power'. Cluster 6 represents 45% of the sample and is characterized by opinions scattered amongst different topics with similar distribution.

We assessed the sentiment of opinions in each cluster using a lexicon-based method suggested by (Taboada *et al.*, 2011), which makes it possible to calculate a sentiment score considering the number of positive and negative words in each opinion cluster. For that purpose, we used the

dictionary developed by Hu and Liu (2004), which categorizes the polarity of 6,800 words (2,006 positive and 4,783 negative words). The results reported in Table 4 confirm H_{1b} , that the opinions in cluster 4 (Irritated consumers) are associated with a significantly high negative sentiment (p -value < .0001). Confirming H_{3b} and H_{4b} , we also find a negative sentiment related to the opinions in cluster 5 (Law-focused consumers) (p -value < .001) and cluster 3 (Protectors) (p -value < .0001). We find that the opinions in cluster 1 (Subscribers) display a significant positive sentiment (p -value < .0001), and contrary to H_{2b} , we find that the average sentiment of opinions in cluster 2 (Cost-anxious consumers) is non-negative (p -value = 0.3093). To strengthen these results and validate the polarity of the opinions, we applied a Naïve Bayes classifier trained on Wiebe's subjectivity lexicon (Jurka, 2012). The results of this analysis, illustrated in Figure 4, reveal that in cluster 3 (Protectors) and cluster 4 (Irritated consumers) more than 65% of the opinions have a negative sentiment and only about 10% have a positive sentiment score. Comparatively, in cluster 1 (Subscribers) about 60% of the opinions have a non-negative score with 25% being strictly positive.

[Table 4 here]

[Figure 4 here]

6. Discussions

6.1. Limitations and future research

Before discussing the results and drawing conclusions from this study, a number of limitations and opportunities for future research can be mentioned. First and foremost, the study is based on a relatively new type of data consisting in 4,093 opinions publicly available online. The main focus of this study is on consumers who are interested in the IT-technology and who are aware of the ad-blockers. The fact that the population considered represents the most likely adopters and promoters of

ad-blocking among the population⁴ strengthens the confidence that one can have in the results, but it is important that future research tests the role of these factors in relation to the general population too. As a direct response to the threat posed by ad-blockers, the advertising industry has already started to change its strategy, and thus it is also reasonable to replicate the present study at a later point in time, also to test the temporal stability of the factors identified. Second, like most other qualitative-based research, the present work is a small-scale exploratory study. This implies that causality statements cannot be made. Hence, future studies are needed to test these relationships in an experimental setting complemented with quantitative methods. Third, the results of this study suggest that a substantial number of consumers have the intention to subscribe to ad-free web contents, but more research is required to identify how much they are willing to pay and which websites would benefit the most from implementing the subscription model. Whether or not the consumers with ad-blocking enabled are still valuable to the site by, for example, linking the website they block elsewhere on the Internet and helping to drive more readers to the site, including readers who do not use ad-blocking is also a relevant issue for future research. Fourth, as this analysis serves for deepening our understanding of consumers' attitudes towards paid online advertising in general, further research on a special case of online ads such as promoted tweets (Muñoz-Expósito *et al.*, 2017), advergame (Vashisht and Sreejesh, 2017), social media advertising (Duffett, 2015), microblogging ads (Zhang and Peng, 2015), and entertaining ads (Cicchirillo and Mabry, 2016) would be beneficial for broadening the academic online advertising literature.

6.2. Conclusions and implications

This study identified five main topics unveiling the hidden structure of consumers' beliefs about online ads and the expected outcomes of ad-blocking. While some of these factors have been previously investigated in other research areas such as information economics (Tåg, 2009; Anderson

⁴ The average age of the Ars readership is 35, 83% of Ars readers said that "my friends often come to me for advice on products and services", and 95% of Ars readers said that they have influenced the purchasing decisions of their friends or peers (Anthony, 2015).

and Gans, 2011), computer and communication sciences (Vratonjic *et al.*, 2013; Singh and Potdar, 2009), no research has to date empirically validated them in an integrated framework. Also, this research fills a gap in the literature mainly focused on the consequences of ad-blocking from the business perspective (Tåg, 2009; Vratonjic *et al.*, 2013; Anderson and Gans, 2011) and not so explicitly with respect to the consumer's interest. The findings extend our understanding of consumer behavior, in the light of the on-going debate about how the advertisers can prevent ad-blocking, either by designing acceptable ads to address the users' concerns or by finding alternative actions to force consumers not to use ad-blockers and accept the existent ad model.

The results showed that perceptions of intrusiveness and obtrusiveness make up one of the most important reasons explaining online ad-blocking behavior. The relevance of behavioral ad characteristics in consumer opinions, especially animation, video, and sound, corroborates Pallant (2011)'s research who found animation and sound as being the two major reasons for consumers using Adblock Plus.⁵ The fact that consumers very often invoked the word "text" suggests that an emphasis on less intrusive forms of ads would be more effective to trigger consumer ad acceptance and prevent the use of ad-blocking technologies. This is in line with Kyun *et al.* (2017)' finding that consumers are less likely to block text-based ads. The results also show that the opinions predominantly discussing behavioral ad characteristics registered the strongest negative sentiment among all the opinions expressed by consumers, indicating that the consumers' negative emotions produced by the feelings associated with the characteristics of ads account for most of the ad-blocking adoption.

Furthermore, the results confirmed that the hidden performance costs related to the online ads are of concern to many users. The consumers invoked that ads pose a greater threat in mobile devices with small data caps, which is in line with (Gao *et al.*, 2016) who found that the costs related to memory/CPU overhead in mobile applications might influence the consumer behavior. Notably, the average score of the sentiment associated with this topic was not significantly negative, arguably suggesting that this factor is the main reason underlying ad-blocking behavior. However, the results suggest that eliminating or reducing the hidden performance costs associated with ads can reduce the

⁵ Adblock Plus is the most popular ad-blocker worldwide.

number of users adopting ad-blocking especially in the mobile devices.

The analyses also showed that a significant number of opinions referred to the possibility to subscribe to an ad-free website. An interpretation of this result is that offering ad-free content could create a window of opportunity for the publishers' business regardless of the ad-based model. Previous researchers have argued that if the website owners made it clear why ads were necessary but offer users the opportunity to view the content ad-free, they may find there is a substantial portion of their audience who would be willing to register or even to pay to visit an ad-free version of their favorite sites (Vallade, 2009). For example, using a game-theoretic model, Vratonjic *et al.*, (2013) found that a strategically applied fee-financed or ad-financed monetization strategy and individual treatment of consumers might yield higher revenues for publishers, compared to deploying one strategy across all consumers. However, an important research question remains about how the new strategy will affect the overall consumer welfare. Theoretical studies in economics (Tåg, 2009; Anderson and Gans, 2011) predict that introducing the option to pay may initially decrease the overall consumer welfare. Firstly, such a decrease occurs because, at the beginning, consumers using the free version will be exposed to more advertisements to compensate revenue losses; and secondly, the ad-blockers' penetration will diminish the quality of the website content. While ad avoidance technologies might eventually drive increased user fees and lower advertising levels in the long term, more empirically-focused studies are required to understand the impact of strategies in which publishers offer two versions of their content that differ in advertising quantity. Because in this study consumers invoked that the value of the site content lead them to accept the subscription-based model, a better understanding of the consumers' valuation of the content is essential in order for publishers to make a well-informed decision regarding the adoption of different strategies (Vratonjic *et al.*, 2013).

One of the topics discovered conveyed the consumers' beliefs about the legitimacy of ad-blocking and their right to choose the content they want to see, in the absence of any proof to the contrary. Discussing the legality of commercial skipping, Vallade (2009) documented that apparently there is no legal remedy available that can force users to watch ads; furthermore, the advertisers' efforts to stop the spread of ad-blockers based on the theory of copyright infringement are not sustainable (Vallade, 2009). Finally, the results confirmed that consumers are concerned about their

privacy and the security of online ads. Notably, the high frequency of the word “trust” implies that there were relatively many individuals emphasizing their lack of trust towards tracking, and they would probably accept ads if given more transparency. This finding is also in line with notions from Thode *et al.* (2015, p. 455), namely that “building up trust on tracking on the customer’s side seems to be an obvious need and one possible way to go”. The concerns of the individuals about the security of their information were very often attributed to the Flash technology. Ad developers and online marketers might find this information useful when deciding to develop/adopt ad applications based on Flash in order to mitigate the perceived risks associated with ads.

Summing up, the study aimed to investigate the key consumer beliefs towards online ads characteristics and the expected outcomes underlying ad-blocking behavior. The study calls for businesses to put more effort into meeting the expectations of consumers and for regulators to improve the legal framework for online advertisements and the behavior of entities in the advertising industry (Clifford and Verdoodt, 2016; Leenes, 2015). We hope that similar studies can be conducted using other methods aimed at deepening these findings.

Accepted

References

- Abernethy, A.M. (1991), "Differences between advertising and program exposure for car radio listening", *Journal of Advertising Research*, Vol. 31 No. 2, pp. 33-42.
- Agarwal, L., Shrivastava, N., Jaiswal, S. and Panjwani, S. (2013), "Do not embarrass: re-examining user concerns for online tracking and advertising", in *Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUPS'13)*, 24-26 July, ACM, New York, NY, pp. 1-13.
- Anderson, S.P. and Gans, J.S. (2011), "Platform siphoning: ad-avoidance and media content", *American Economic Journal: Microeconomics*, Vol. 3 No. 4, pp. 1-34.
- Andrieu, C., De Freitas, N., Doucet, A. and Jordan, M.I. (2003), "An introduction to MCMC for machine learning", *Machine Learning*, Vol. 50 No. 1-2, pp. 5-43.
- Anthony, S. (2015), "9.5 out of 10 Ars Technica readers prefer Ars Technica", available at: <https://arstechnica.co.uk/staff/2015/12/9-5-out-of-10-ars-technica-readers-trust-what-they-read-on-ars-technica/> (accessed 10 August 2017).
- Baek, T.H. and Morimoto, M. (2012), "Stay away from me", *Journal of Advertising*, Vol. 41 No. 1, pp. 59-76.
- Bart, Y., Shankar, V., Sultan, F. and Urban, G.L. (2005), "Are the drivers and role of online trust the same for all web sites and consumers? A large-scale exploratory empirical study", *Journal of Marketing*, Vol. 69 No. 4, pp. 133-152.
- Becker, G.S. (1996), *Accounting for Tastes*, Harvard University Press, Cambridge, MA.
- Blei, D.M. (2012), "Probabilistic topic models", *Communications of the ACM*, Vol. 55 No. 4, pp. 77-84.
- Blei, D.M. and Lafferty, J.D. (2007), "A correlated topic model of science", *The Annals of Applied Statistics*, Vol. 1 No. 1, pp. 17-35.
- Blei, D.M. and Lafferty, J.D. (2009), "Topic models", *Text mining: Classification, Clustering, and Applications*, Vol. 10 No. 71, pp. 71-89.

- Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, Vol. 3 No. 1, pp. 993-1022.
- Bouchet-Valat, M. (2014), "SnowballC: snowball stemmers based on the C libstemmer UTF-8 library", available at: <https://cran.r-project.org/web/packages/SnowballC/index.html> (accessed 3 March 2016).
- Brehm, J.W. (2009), "A theory of psychological reactance", in Burke, W.W., Lake, D.G. and Paine, J.W. (Ed.), *Organization Change: A Comprehensive Reader*, John Wiley and Sons Inc., San Francisco, pp. 377-392.
- Broida, R. (2015), "Why I changed my mind about ad-blocking software", available at: <https://www.cnet.com/news/why-i-changed-my-mind-about-adblock-software/> (accessed 8 April 2016).
- Burke, M., Hornof, A., Nilsen, E. and Gorman, N. (2005), "High-cost banner blindness: ads increase perceived workload, hinder visual search, and are forgotten", *ACM Transactions on Computer-Human Interaction*, Vol. 12 No. 4, pp. 423-445.
- Burns, K.S. and Lutz, R.J. (2006), "The function of format: consumer responses to six on-line advertising formats", *Journal of Advertising*, Vol. 35 No. 1, pp. 53-63.
- Cho, C.H. and Cheon, J.H. (2004), "Why do people avoid advertising on the Internet?", *Journal of Advertising*, Vol. 33 No. 4, pp. 89-97.
- Cicchirillo, V. and Mabry, A. (2016), "Advergaming and healthy eating involvement: how healthy eating inclinations impact processing of advergame content", *Internet Research*, Vol. 26 No. 3, pp. 587-603.
- Clifford, D. and Verdoodt, V. (2016), "Ad-blocking-the dark side of consumer empowerment: a new hope or will the empire strike back?", paper presented at the BILETA Conference, 11-12 April, University of Hertfordshire, Hertfordshire, UK, available at: https://lirias.kuleuven.be/bitstream/123456789/538851/2/BILETA2016_D.Clifford_V.Verdoodt_Ad-blocking_25_04_2016.pdf (accessed 8 May 2017).

- Cronin, J.J. and Menelly, N.E. (1992), "Discrimination vs. avoidance: "zipping" of television commercials", *Journal of Advertising*, Vol. 21 No. 2, pp. 1-7.
- De Waal, A. and Barnard, E. (2008), "Evaluating topic models with stability", in *Proceedings of the Nineteenth Annual Symposium of the Pattern Recognition Association of South Africa*, IAPR, Cape Town, pp. 79-84.
- Drèze, X. and Hussherr, F.X. (2003), "Internet advertising: is anybody watching?" *Journal of Interactive Marketing*, Vol. 17 No. 4, pp. 8-23.
- Duffett, R.G. (2015), "Facebook advertising's influence on intention-to-purchase and purchase amongst Millennials", *Internet Research*, Vol. 25 No. 4, pp. 498-526.
- Dvorkin, L. (2016), "Inside Forbes: what journalists must know - and can do - about new upheavals in the ad world", available at: <http://www.forbes.com/sites/lewisdvorkin/2016/01/05/inside-forbes-from-original-sin-to-ad-blockers-and-what-the-future-holds/-49230a6e18ae> (accessed 23 April 2016).
- Eagly, A.H., and Chaiken, S. (1998), "Attitude structure and function", in Gilbert, D.T., Fiske, S.T. and Lindzey, G. (Ed.), *The Handbook of Social Psychology*, McGraw-Hill, New York, NY, pp. 269-322.
- Edwards, S.M., Li, H. and Lee, J.H. (2002), "Forced exposure and psychological reactance: antecedents and consequences of the perceived intrusiveness of pop-up ads", *Journal of Advertising*, Vol. 31 No. 3, pp. 83-95.
- Emerson, R.M. (1962), "Power-dependence relations", *American Sociological Review*, Vol. 27 No. 1, pp. 31-41.
- Fang, X., Singh, S. and Ahluwalia, R. (2007), "An examination of different explanations for the mere exposure effect", *Journal of Consumer Research*, Vol. 34 No. 1, pp. 97-103.
- Fishbein, M. and Ajzen, I. (1977), "Belief, attitude, intention, and behavior: an introduction to theory and research", *Philosophy and Rhetoric*, Vol. 10 No. 2, pp. 130-132.

- Fisher, K. (2010), “Why ad blocking is devastating to the sites you love”, available at: <http://arstechnica.com/business/2010/03/why-ad-blocking-is-devastating-to-the-sites-you-love/> (accessed 25 March 2016).
- Fraley, C. and Raftery, A.E. (2002), “Model-based clustering, discriminant analysis, and density estimation”, *Journal of the American Statistical Association*, Vol. 97 No. 458, pp. 611-631.
- Gao, C., Xu, H., Man, Y., Zhou, Y. and Lyu, M.R. (2016), “IntelliAd understanding in-APP ad costs from users’ perspective”, available at: <https://arxiv.org/pdf/1607.03575.pdf> (accessed 4 April 2017).
- Goldfarb, A. and Tucker, C. (2011a), “Online display advertising: targeting and obtrusiveness”, *Marketing Science*, Vol. 30 No. 3, pp. 389-404.
- Goldfarb, A. and Tucker, C. (2011b), “Search engine advertising: channel substitution when pricing ads to context”, *Management Science*, Vol. 57 No. 3, pp. 458-470.
- Goldstein, D.G., Suri, S., McAfee, R.P., Ekstrand-Abueg, M. and Diaz, F. (2014), “The economic and cognitive costs of annoying display advertisements”, *Journal of Marketing Research*, Vol. 51 No. 6, pp. 742-752.
- Google Trends (2016), *Google Trends*, available at: <https://www.google.com/trends/> (accessed 4 January 2016).
- Greene, D., O’Callaghan, D. and Cunningham, P. (2014), “How many topics? Stability analysis for topic models”, in Calders, T., Esposito, F., Hüllermeier, E. and Meo, R. (Ed.), *Conference Proceedings ECML PKDD: Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, Berlin, pp. 498-513.
- Gui, J., Mcilroy, S., Nagappan, M. and Halfond, W.G. (2015), “Truth in advertising: the hidden cost of mobile ads for software developers”, in *Proceedings of the 37th International Conference on Software Engineering (ICSE)*, ACM, Florence/Firenze, Italy, pp. 527-538.
- Ha, L. and McCann, K. (2008), “An integrated model of advertising clutter in offline and online media”, *International Journal of Advertising*, Vol. 27 No. 4, pp. 569-592.

- Hair, J.F., Black, W.C., Babin, B.J. Anderson, R.E. and Tatham, R.L. (2006), *Multivariate Data Analysis (Vol. 6)*, Pearson Prentice Hall Upper Saddle River, New Jersey, NJ.
- Hollander, M., Wolfe, D.A. and Chicken, E. (2013), *Nonparametric Statistical Methods*, John Wiley and Sons Inc., New York, NY.
- Hong, W., Thong, J.Y. and Tam, K.Y. (2004), “Does animation attract online users’ attention? The effects of flash on information search performance and perceptions”, *Information Systems Research*, Vol. 15 No. 1, pp. 60-86.
- Hong, W., Thong, J.Y.L. and Tam, K.Y. (2007), “How do Web users respond to non-banner-ads animation? The effects of task type and user experience”, *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 10, pp. 1467-1482.
- Hornik, K. and Grün, B. (2011), “topicmodels: an R package for fitting topic models”, *Journal of Statistical Software*, Vol. 40 No. 13, pp. 1-30.
- Hu, M. and Liu, B. (2004), “Mining and summarizing customer reviews”, in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD’04*, New York, NY, pp. 168-177.
- Jurka, T. (2012), “Sentiment: tools for sentiment analysis”, available at: <https://github.com/timjurka/sentiment> (accessed 20 March 2016).
- Keller, K.L. (2010), “Brand equity management in a multichannel, multimedia retail environment”, *Journal of Interactive Marketing*, Vol. 24 No. 2, pp. 58-70.
- Kireyev, P., Pauwels, K. and Gupta, S. (2015), “Do display ads influence search? Attribution and dynamics in online advertising”, *International Journal of Research in Marketing*, Vol. 33 No. 3, pp. 475-490.
- Kyun, C.Y., Yuri, S. and Sukki, Y. (2017), “E-WOM messaging on social media: social ties, temporal distance, and message concreteness”, *Internet Research*, Vol. 27 No. 3, pp. 495-505.

- Labrecque, L.I., vor dem Esche, J., Mathwick, C., Novak, T.P. and Hofacker, C.F. (2013), “Consumer power: evolution in the digital age”, *Journal of Interactive Marketing*, Vol. 27 No. 4, pp. 257-269.
- Leenes, R. (2015), “The cookiewars: from regulatory failure to user empowerment?”, in van Lieshout, M. and Hoepman, J.H. (Ed.), *The Privacy & Identity Lab: 4 years later*, The Privacy & Identity Lab, Nijmegen, pp. 31-49.
- Lewis, R.A. and Reiley, D.H. (2014), “Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo!”, *Quantitative Marketing and Economics*, Vol. 12 No. 3, pp. 235-266.
- Li, H. and Kannan, P. (2014), “Attributing conversions in a multichannel online marketing environment: an empirical model and a field experiment”, *Journal of Marketing Research*, Vol. 51 No. 1, pp. 40-56.
- Lwin, M.O., Wirtz, J. and Stanaland, A.J.S. (2016), “The privacy dyad: antecedents of promotion- and prevention-focused online privacy behaviors and the mediating role of trust and privacy concern”, *Internet Research*, Vol. 26 No. 4, pp. 919-941.
- Madhavaram, S. and Appan, R. (2010), “The potential implications of web-based marketing communications for consumers' implicit and explicit brand attitudes: a call for research”, *Psychology and Marketing*, Vol. 27 No. 2, pp. 186-202.
- Mayer, J.R. and Mitchell, J.C. (2012), “Third-party web tracking: policy and technology”, paper presented at the Security and Privacy (SP) 2012 IEEE Symposium, 20-23 May, San Francisco, CA, available at: https://jonathanmayer.org/papers_data/trackingsurvey12.pdf (accessed 17 March 2018).
- McCoy, S., Everard, A., Polak, P. and Galletta, D.F. (2008), “An experimental study of antecedents and consequences of online ad intrusiveness”, *International Journal of Human-Computer Interaction*, Vol. 24 No. 7, pp. 672-699.

- McGann, L. (2010), “How Ars Technica’s “experiment” with ad-blocking readers built on its community’s affection for the site”, available at: <http://www.niemanlab.org/2010/03/how-ars-technica-made-the-ask-of-ad-blocking-readers/> (accessed 17 March 2016).
- McLachlan, G. and Peel, D. (2000), “Mixtures of factor analyzers”, in McLachlan, G. and Peel, D. (Ed.), *Finite Mixture Models*, John Wiley and Sons Inc., Hoboken, NJ, pp. 238-256.
- Meyer, D., Hornik, K. and Feinerer, I. (2008), “Text mining infrastructure in R”, *Journal of Statistical Software*, Vol. 25 No. 5, pp. 1-54.
- Muñoz-Expósito, M., Oviedo-García, M.Á. and Castellanos-Verdugo, M. (2017), “How to measure engagement in Twitter: advancing a metric”, *Internet Research*, Vol. 27 No. 5, pp. 1122-1148.
- Neben, T. and Schneider, C. (2015), “Ad intrusiveness, loss of control, and stress: a psychophysiological study”, in *Proceedings of the Thirty Sixth International Conference on Information Systems (ICIS 2015)*, AIS Electronic Library, Fort Worth, Texas, pp. 1-11.
- PageFair and Adobe. (2015), “The cost of ad blocking”, available at: https://downloads.pagefair.com/wp-content/uploads/2016/05/2015_report-the_cost_of_ad_blocking.pdf (accessed 22 March 2016).
- Pallant, W. (2011), “Adblock Plus user survey”, available at: <https://adblockplus.org/blog/adblock-plus-user-survey-results-part-2> (accessed 23 May 2016).
- Robertson, S. (2004), “Understanding inverse document frequency: on theoretical arguments for IDF”, *Journal of Documentation*, Vol. 60 No. 5, pp. 503-520.
- Rogers, R. W. (1975), “A protection motivation theory of fear appeals and attitude change”, *Journal of Psychology*, Vol. 91 No. 1, pp. 93-114.
- Rucker, D.D., Galinsky, A.D. and Dubois, D. (2012), “Power and consumer behavior: how power shapes who and what consumers value”, *Journal of Consumer Psychology*, Vol. 22 No. 3, pp. 352-368.

- Scherer, K.R., Schorr, A. and Johnstone, T. (2001), *Appraisal Processes in Emotion: Theory, Methods, Research*, Oxford University Press, New York, NY.
- Securities and Optimal.com. (2016), “Ad-blocking survey and forecast”, available at: <https://optimal.com/optimal-com-wells-fargo-ad-blocking-survey/> (accessed 7 April 2016).
- Seyedghorban, Z., Tahernejad, H. and Matanda, M.J. (2016), “Reinquiry into advertising avoidance on the Internet: a conceptual replication and extension”, *Journal of Advertising*, Vol. 45 No. 1, pp. 120-129.
- Shannon, C.E. and Weaver, W. (1948), “A mathematical theory of communication”, *The Bell System Technical Journal*, Vol. 27 No. 3, pp. 379-423.
- Singh, A.K. and Potdar, V. (2009), “Blocking online advertising - A state of the art”, in *Proceedings of the IEEE International Conference on Industrial Technology, 10-13 February*, IEEE, Gippsland, Victoria, Australia, pp. 1-6.
- Speck, P.S. and Elliott, M.T. (1997), “Predictors of advertising avoidance in print and broadcast media”, *Journal of Advertising*, Vol. 26 No. 3, pp. 61-76.
- Steyvers, M. and Griffiths, T. (2007), “Probabilistic topic models”, *Handbook of Latent Semantic Analysis*, Vol. 427 No. 7, pp. 424-440.
- Sundar, S.S. and Kalyanaraman, S. (2004), “Arousal, memory, and impression-formation effects of animation speed in web advertising”, *Journal of Advertising*, Vol. 33 No. 1, pp. 7-17.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M. (2011), “Lexicon-based methods for sentiment analysis”, *Computational Linguistics*, Vol. 37 No. 2, pp. 267-307.
- Tåg, J. (2009), “Paying to remove advertisements”, *Information Economics and Policy*, Vol. 21 No. 4, pp. 245-252.
- Tang, J., Zhang, P. and Wu, P.F. (2015), “Categorizing consumer behavioral responses and artifact design features: the case of online advertising”, *Information Systems Frontiers*, Vol. 17 No. 3, pp. 513-532.

- Thode, W., Griesbaum, J. and Mandl, T. (2015), "'I would have never allowed it': user perception of third-party tracking and implications for display advertising", in Pehar, F., Schlögl, C. and Wolff, C. (Ed.). *Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015) 19-21 May*, Verlag Werner Hülsbusch, Zadar, Croatia, pp. 455-456.
- Tilkov, S. and Vinoski, S. (2010), "Node.js: using javascript to build high-performance network programs", *IEEE Internet Computing*, Vol. 14 No. 6, pp. 80-83.
- Ur, B., Leon, P.G., Cranor, L.F., Shay, R. and Wang, Y. (2012), "Smart, useful, scary, creepy: perceptions of online behavioral advertising", paper presented at the 8th Symposium on Usable Privacy and Security (SOUPS), 11-13 July, Washington, DC, available at: https://www.blaseur.com/papers/soups2012-oba_ur.pdf (accessed 17 March 2018).
- Vallade, J. (2009), "Adblock plus and the legal implications of online commercial-skipping", *Rutgers Law Review*, Vol. 61 No. 3, pp. 823-853.
- Vashisht, D. and Sreejesh, S. (2017), "Effect of nature of the game on ad-persuasion in online gaming context: moderating roles of game-product congruence and consumer's need for cognition", *Internet Research*, Vol. 27 No. 1, pp. 52-73.
- Verlegh, P.W.J., Fransen, M.L. and Kirmani, A. (2015), "Persuasion in advertising: when does it work, and when does it not?", *International Journal of Advertising*, Vol. 34 No. 1, pp. 3-5.
- Vratonjic, N., Manshaei, M.H., Grossklags, J. and Hubaux, J.P. (2013), "Ad-blocking games: monetizing online content under the threat of ad avoidance", in Böhme, R. (Ed.), *The Economics of Information Security and Privacy*, Springer, Berlin, Heidelberg, pp. 49-73.
- Wedel, M. and Kamakura, W.A. (2002), "Introduction to the special issue on market segmentation", *International Journal of Research in Marketing*, Vol. 19 No. 3, pp. 181-183.
- Wirtz, J., Lwin, M.O. and Williams, J.D. (2007), "Causes and consequences of consumer online privacy concern", *International Journal of Service Industry Management*, Vol. 18 No. 4, pp. 326-348.

- Xu, L., Duan, J.A. and Whinston, A. (2014), "Path to purchase: a mutually exciting point process model for online advertising and conversion", *Management Science*, Vol. 60 No. 6, pp. 1392-1412.
- Youn, S. (2009), "Determinants of online privacy concern and its influence on privacy protection behaviors among young adolescents", *Journal of Consumer Affairs*, Vol. 43 No. 3, pp. 389-418.
- Zha, W. and Wu, H.D. (2014), "The impact of online disruptive ads on users' comprehension, evaluation of site credibility and sentiment of intrusiveness", *American Communication Journal*, Vol. 16 No. 2, pp. 15-28.
- Zhang, L. and Peng, T.Q. (2015), "Breadth, depth, and speed: diffusion of advertising messages on microblogging sites", *Internet Research*, Vol. 25 No. 3, pp. 453-470.
- Zhou, T. (2017), "Understanding location-based services users' privacy concern: an elaboration likelihood model perspective", *Internet Research*, Vol. 27 No. 3, pp. 506-519.

Accepted

Table 1.

Data source description

Date	Article name	No. of documents retrieved
2010/03	C1: Why Ad-blocking is devastating to the sites you love	1,937
2010/03	C2: Safely whitelist your favorite sites and opt out of tracking (updated rules)	272
2010/03	C3: Pew: readers prefer ad-supported news to pay walls	80
2013/01	C4: France's second-largest ISP deploys ad-blocking via firmware update	178
2013/01	C5: France's second-largest ISP suspends ad-blocking for now	62
2013/03	C6: Google evicts ad-blocking software from Google play store	109
2015/02	C7: Over 300 businesses now whitelisted on Adblock Plus. 10% pay to play	200
2015/05	C8: EU carriers plan to block ads, demand money from Google	176
2015/10	C9: Mobile carrier to Google. Yahoo. Facebook: Pay up or we'll block your ads	180
2015/10	C10: Online advertisers admit they 'messed up' and promise lighter ads	353
2016/01	C11: Mozilla co-founder unveils Brave, a browser that blocks ads by default	150
2016/02	C12: NY Times recommends ad blockers after CEO mulls ad-block ban	117
2016/03	C13: In the name of free speech, adblock serves-up ads, just for a day	136
2016/05	C14: Adblock Plus will help consumers pay publishers and keep a cut for itself	143

Table 2.

Topics learned and the top-thirty words in the learned topic-word distribution

Topic	Topic label	Top-thirty words
1	Annoying ads characteristics	advertisement block page web annoy click start stop turn show visit problem product mind animation video day find sound put move care text stuff bad good play fine intrusiveness obnoxious
2	Performance costs	pay google internet free money user business company provider model service cost customer isp (internet service provider) paid data consumption bandwidth revenue access give buy app lot sell charge connect market call force
3	Subscription	article read comment subscript reader subscribe good support post guy work month interest feel experience thread option link idea long year worth love great news thought lot bit hope pretty
4	Legitimacy of ad-blocking and consumer power	website view reason work point issue legal person simple argument case fact agree number revenue user base understand software change mean world effect small exist specify kind system result matter
5	Privacy and security	flash block run whitelist browser server load track install computer security problem malware network static server display trust user script brows disable noscript doubleclick accept partial control flashblock firefox blocker

Table 3.

Characteristics of the clusters

Cluster	1	2	3	4	5	6	KW test*
n	505	400	549	500	287	1851	
%	12.34	9.77	13.41	12.21	7.01	45.23	
<i>By news</i>							
C1	0.19	0.05	0.14	0.13	0.09	0.40	< .0001
C2	0.11	0.01	0.35	0.07	0.03	0.43	< .0001
C3	0.15	0.11	0.04	0.10	0.13	0.48	< .0001
C4	0.06	0.15	0.13	0.16	0.08	0.42	< .0001
C5	0.02	0.58	0.02	0.03	0.03	0.32	< .0001
C6	0.01	0.25	0.06	0.04	0.06	0.60	< .0001
C7	0.05	0.03	0.16	0.14	0.04	0.58	< .0001
C8	0.01	0.49	0.02	0.01	0.05	0.42	< .0001
C9	0.03	0.35	0.04	0.03	0.07	0.47	< .0001
C10	0.06	0.04	0.10	0.26	0.07	0.47	< .0001
C11	0.03	0.06	0.09	0.07	0.01	0.74	< .0001
C12	0.11	0.05	0.14	0.18	0.03	0.50	< .0001
C13	0.02	0.07	0.16	0.15	0.08	0.52	< .0001
C14	0.15	0.10	0.08	0.11	0.03	0.52	< .0001
<i>By topics</i>							
1. Annoying ads characteristics	0.18	0.17	0.18	0.29	0.17	0.20	< .0001
2. Performance costs	0.17	0.32	0.16	0.17	0.19	0.20	< .0001
3. Subscription	0.30	0.17	0.17	0.18	0.18	0.20	< .0001
4. Legitimacy of ad-blocking and consumer	0.19	0.18	0.17	0.18	0.32	0.20	< .0001

power

5. Privacy and security	0.16	0.16	0.31	0.19	0.15	0.20	< .0001
-------------------------	------	------	-------------	------	------	------	---------

Notes: *Non-parametric Kruskal-Wallis test of significant differences between cluster means. The reported value is the significance level (p -value) associated with the test.

Accepted

Table 4.

Sentiment analysis: valence and strength

Cluster number (label)	Sentiment average score	<i>p</i> -value*
1 (Subscribers)	0.64	< 0.0001
2 (Cost-anxious consumers)	0.13	0.3093
3 (Protectors)	-0.77	< 0.0001
4 (Irritated consumers)	-1.46	< 0.0001
5 (Law-focused consumers)	-0.87	0.0002
6 (General consumers)	-0.10	0.003

Notes: *Significance level for the test of hypothesis H_0 : mean = 0.

Accepted

Figure 1.

Relative search interest in the phrase 'adblock' over time

Notes: The data were collected from Google Trends (2016). The search interest in the phrase 'adblock' over time was normalized as a proportion of all searches on all topics on Google at a particular time. The data were indexed to 100, where 100 is the maximum search interest for the timeframe selected (Google Trends, 2016).

Figure 2.

Topic stability analysis

Figure 3.

Distribution of the topics across opinions and by news

a.

b.

c.

d.

c.

Figure 4.

Percentage of opinions with negative, neutral and positive sentiment within each cluster