International Journal of Communication 11(2017), 2178–2197 1932–8036/20170005

The Reasons Behind Tracing Audience Behavior: A Matter of Paternalism and Transparency

ESTER APPELGREN
Södertörn University, Sweden

This article analyzes privacy agreement texts and cookie consent information collected from 60 news sites in three countries (U.S., UK, and Sweden) within the context of paternalism. The goal of this study is to explore how paternalism is present in news media companies' stated reasons for collecting behavioral data. Twenty-five categories of reasons were identified and divided into six categories: personalization and enhanced user experience, delivery and maintenance of services, internal and corporate use of data, legal reasons, communication with the user, and third-party use of data. The analysis shows that the reasons can be formulated in both paternalistic and nonpaternalistic ways, and that the market-driven logic of Web analytics seems to collide with ethics in a journalistic context.

Keywords: behavioral data, privacy, General Data Protection Regulation (GDPR), paternalism, news media and transparency

For data controllers to obtain informed consent in a digital setting, the law requires an indication of wishes of the data subject (Borgesius, 2015). Nevertheless, many companies currently ask for permission to collect audience data by having users grant consent in a more passive form (e.g., consent is granted when the user continues to click around a website). Indeed, current European legislation permits companies to use passive consent; however, the reform of the data protection rules in the EU (General Data Protection Regulation [GDPR]) will present changes in this area (European Parliament, 2016) and will affect companies and organizations in all member states of the European Union as well as non-EU companies that have websites that target a European audience. Specifically, the new regulation clarifies that consent is not freely given if the data subject did not have genuinely free choice or is unable to withdraw or refuse consent without detriment (Allen & Overy, 2016). The transparency principle discussed in the GDPR (European Commission, 2016) states that when the regulation applies in May 2018, companies and organizations must transparently explain why they are collecting data about users.

Recent research on privacy and the news media focuses on journalistic and user-generated content, however, pointing in different directions. Following Edward Snowden's revelations about the NSA in 2013, Mols and Jansen (2016) found six groupings of privacy attitudes in the public Dutch debate. In the most negative grouping, the attitude was expressed as if the only way to protect individual privacy is

Ester Appelgren: ester.appelgren@sh.se
Date submitted: 2017–01–02

"to leave the digital realm altogether" (p. 14). In a British study, the media was instead found to justify mass surveillance, but the public response following the revelations by Snowden in the UK was remarkably muted (Hintz & Dencik, 2016). Bechmann (2014) argues that "privacy is downplayed" (p. 22) since it is common for companies to ask for consent implicitly in documents on their websites that contain user terms and privacy policies. Perhaps because of this widespread practice of passive consent, the majority of Europeans now feel there is no alternative other than to provide personal information if they want to use digital services (Special Eurobarometer, 2015), and survey studies have recently found that most people do not read privacy agreement texts at all. In Sweden, only 15% of the population state that they read privacy agreement texts (Appelgren & Leckner, 2016). In the UK, 12% ("GB Consumer Privacy Index 2016," 2016) and in the U.S., 16% ("U.S. Consumer Privacy Index 2016," 2016).

Today, 60% of the Swedish population state that they view negatively the media companies' collection of their behavioral data while they are consuming news content online (Appelgren & Leckner, 2016). Because audience trust is a cornerstone for journalism, media companies may face problems if users start to become concerned about their privacy when they are consuming news.

This article takes a closer look at how news companies describe their own actions when collecting behavioral data from their audience. News media companies, like any other organization or company with digital services, are subject to legislation that has shaped the consent-request process and made it more transparent. This regulatory overview includes the privacy- and cookie-policy texts, where the consent request must be presented alongside text describing both the purposes for collected user data and how the audience can opt out. This study is therefore based on a content analysis of 60 privacy texts collected from the most popular news websites in three countries with different legislation present on a European market: the U.S., the UK, and Sweden.

Since people do not generally read the reasons that can be found in user terms and privacy texts and have grown accustomed to the "privacy paradox" (i.e., accepting user terms to get access to services and content to avoid digital isolation; Bechmann, 2014), media companies, using collected data, may take actions that users have not actively agreed to. When someone else makes choices for an individual in this manner, we may speak of paternalism. Paternalistic intervention is presumably in the interest of the individual (see, e.g., Clarke, 2002; Dworkin, 1972; Le Grand & New, 2015). Therefore, if informed consent is obtained from an individual, it is logical to assume that he or she agrees with the possibly paternalistic interventions that may occur as a result of behavioral data collection. However, if the consent is passive and thus uninformed, can the paternalistic reasons for collecting data be considered in the interest of the audience? Journalism strives to promote certain values, but what happens if journalistic values collide with the more commercially oriented culture of Web analytics and design for website technology? This article delves deeper into this question using a normative perspective on paternalism and aims to explore the extent to which paternalism is present in the reasons news media companies give for collecting behavioral data.

This analysis of the study concerns the explicit reasons stated in news media terms and policies for why data are collected. It is important to point out that privacy texts are not considered journalism, but, in the case of news media, they are embedded in a journalistic context.

This article is outlined as follows. I begin with a brief theoretical discussion, followed by the results section, where the content analysis of the privacy texts is presented and illustrated by select quotes from the analyzed privacy texts. The article concludes with a discussion about how the reasons for monitoring users in a news media context are related to both positive and negative aspects of paternalism.

The Many Faces of Paternalism

An intervention is considered paternalistic if it implies a limitation to a person's autonomy by either preventing someone from doing what he or she has decided to do or interfering in the manner in which someone arrives at their decision (Dworkin, 1983). Clarke (2002) extends this definition and suggests that the behavior is paternalistic if "the aim is to close an option or choose for a person when such an intervention is to be carried out for the target's own good" (p. 82). Most definitions of paternalism in previous research are focused on the outcome of a paternalistic intervention. However, according to Dworkin (2015), "the analysis of whether an act is paternalistic or not must also consider the reasons behind the act" (p. 18).

Beginning in 1972, Dworkin evaluated the standpoints of Mill in his famous work on how paternalism interferes with a man's liberty. Dworkin suggested that a person may be more likely to consent to paternalism if the paternalistic act would "enhance the ability of the individual to rationally consider and carry out decisions" (Dworkin, 1972, p. 83). This kind of attempt to convince people that paternalism would be in their interest is also described by Kerr, Barrigar, Burkell, and Black (2006), which notes that business and governments, in their role as organizations that work with personal information, recently have discovered that kindly asking for personal information is just as effective as engaging in traditional surveillance. This implies that the paternalistic action is softened, introducing an option for people to choose if they agree to be monitored by an organization. An example of typical soft surveillance according to Kerr et al. (2006) is when "consumers are asked to 'volunteer' their name, address, telephone number, email, and zip code to be granted access to online services" (p. 3). Kerr et al. describe how such an act of obtaining consent has been designed as an illusion of free choice. In the view of Kerr et al., soft paternalism implies that public or private institutions are changing default rules to steer peoples' choices in directions that will improve their welfare. This is "engineered by the virtue of consent, or rather by the lack of dissent" (Kerr et al., 2006, p. 2).

Dworkin (2002) has suggested a number of different types of paternalism as opposites: pure or impure, hard or soft, broad or narrow, weak or strong, and moral or welfare. Even though many versions of paternalism are debated in the literature, Le Grand and New (2015) argue that the major distinction is between soft and hard paternalism, where the concern is focused on whether the individual is acting autonomously or voluntarily. Pope (2003) defines soft paternalism as a "limitation of a subject's liberty, where the subject does not act substantially autonomously" (p. 662), whereas hard paternalism is "the limitation of a subject's liberty where the subject acts substantially autonomously" (p. 667), meaning that, according to Pope, a person is not making choices voluntarily when they lack the requisite decision-making capacity, for example, if they make decisions when they are not factually informed, coerced, or have not adequately understood. Soft paternalism involves interferences that help people achieve what

they value (Haybron & Alexandrova, 2013). This is also described as libertarian paternalism by Thaler and Sunstein (2008), when people are "nudged" by different types of design to make choices that will make them "better off." Hard paternalism is instead about limiting a person's liberty for that person's own good (Pope, 2003). Note, however, that if the agent restricts a subject's liberty for selfish reasons, then the agent's restriction of the subject's liberty would not be paternalism (Pope, 2003).

Because technology today may encompass elements of artificial intelligence and machine learning, paternalism may also involve ethics of technology (Hofmann, 2003). Paternalism in a technological setting has predominantly negative connotations, and engineers, scientists, and experts are often "accused" of paternalism when technological solutions compromise the autonomy of individuals (Hofmann, 2003). Furthermore, systems that are paternalistic are described as able to "punish" humans even though the punishment may be in their own interest. User experience and personalization are frequent research topics related to ubiquitous computing. Here, researchers strive to make computers invisible, having them "stay out of the way," even though they may be everywhere (Weiser & Brown, 1997). Consequently, computers are being entrusted to make decisions for people and improve the everyday lives of people without the technology creating a disturbance. This process can be described as an act of technological paternalism (Hofmann, 2003; Spiekermann & Pallas, 2006; i.e., machines make decisions for individuals using behavioral data and preprogrammed rules that go into action without the conscious and active consent of the users). When people use websites with interactive features, they both actively and passively submit or leave information about themselves for others to see and use. Marx (2015) questions if autonomy is lessened in such instances. The user may not be aware that the collection is taking place and thus is less in control, but he or she nevertheless at some point consented to the collection of the data.

Paternalism is furthermore present in journalistic content, although according to Thomas (2016), this paternalism does not have to imply fear or scorn. Instead, paternalistic choices made by journalists can, according to Thomas, be considered positive. Tandoc and Thomas (2015) argue that "journalists must preserve their editorial autonomy if they are to meet the functions that come with the communitarian role of journalism" (p. 252). They argue that if there is a drift toward neoliberal atavism, where "the expertise and judgment of the producer are disregarded in favor of the transient needs of the consumer, this must be resisted" (p. 251). Nevertheless, personalization of behavioral data related to news consumption exists. In this context, Borgesius (2015) discusses the difficult nature of news media companies and privacy legislation using the example of a news company providing a smartphone app with personalized news. Behavioral data are needed for analyzing the users' reading habits to recommend news articles; without such data, the service cannot work. However, it is not technologically necessary for the app to use the same personal data for targeted advertising. If consent has not been obtained for both of these reasons, the law does not support both actions. If users also fully grasped that certain data are not always necessary for the reasons a company may provide, this could result in trust problems. Nissenbaum (2015) suggests that technologies that disturb our sense of privacy use inappropriate information flows that violate context-specific informational norms. In a journalistic context that is paired with the market-oriented nature of Web analytics, there are at least two sets of norms present, and if actions result in paternalistic choices made with commercial or technologically deterministic interests in

mind, such as in the case of technology paternalism, these are most often seen as negative (Hoffmann, 2003; Spiekermann & Pallas, 2006; Tandoc & Thomas, 2015).

Journalistic Ideals and Surveillance

Tuchman (1978) describes how news workers identify themselves as a kind of mediator of social reality, placed between the government, the media companies, and the public. She further describes how journalists reveal social reality to the news consumer in the spirit of protecting the public from excess of the government, and the government from excess of the people. Today, the forces from companies with a digital presence are yet another aspect of social reality in which news workers shoulder the role of protecting the public. Campbell and Carlson (2002) argue that "surveillance is a key mechanism of social control, ensuring the 'rationality,' and therefore predictability, of consumers in the marketplace" (p. 587), yet people participate in digital surveillance because they understand to some extent that findings in data are commodities that can be exchanged for perceived benefits. Nevertheless, Campbell and Carlson (2002) argue that our decisions to participate in surveillance are not made in an equal environment. In this context, the concept of paternalism relates to the broader debate on privacy and integrity in the aspect of users' inability to make informed choices in the digital environment. If not made public, Couldry and Turow (2014) state that the findings from big data used within media companies for personalization of content and advertising might "erode democracy" (p. 1711), since few journalists would have the "time and resources to resist the challenges audience data mining may impose on journalistic values" (p. 1718). An example of this is provided by Boczkowski and Peer (2011), who found that journalists favor public affair stories, but audiences tend to choose non-public-affair stories. Journalists then try to close this choice gap by providing fewer public affairs stories and thus contributing to a decrease in journalistic autonomy, a core part of journalistic ideals.

Method

This article is based on a content analysis of privacy texts collected from 60 news sites in the U.S. (n = 20), UK (n = 20), and Sweden (n = 20). The sites were selected using top lists from the market intelligence service, SimilarWeb. According to SimilarWeb, the data behind the top lists are big data sets obtained from a panel of globally monitored devices, local ISPs, Web crawlers, and connected websites and apps ("Top Websites," 2016). A list of the top websites in August 2016 in the three countries in the study was compiled using a combination of the SimilarWeb categories "News and Media" and "Newspapers." The SimilarWeb service was selected as a basis for the sample to uniformly calculate the traffic share for the three countries.

The texts were collected between October 4 and 6, 2016, by entering each of the websites in the sample after deleting the cookies and using the incognito mode in Google Chrome. The texts that were collected consisted of the first text the user encounters either when the company is asking for consent to monitor the user or if the user actively searches for information about privacy or cookies. Therefore, when entering each website for the first time, if a pop-up or banner with privacy information appeared, the first clickable link to text was included in the sample. If there was no banner or pop-up stating that the site collected behavioral data, the privacy policy link was accessed and included in the sample. If no privacy

policy was found, the cookie policy was instead included. If none of these were found, the news company was included in the sample, but no information could be analyzed. Three companies in the sample did not provide any type of privacy policy, text, or cookie policy on their website (see Table 1).

Because privacy texts are created in accordance with current legislation, they tend to include similar formulations and arguments. The reasons given for the collection of data were counted as part of the study, but it is important to stress that more or less the same topics for collection were given in all of the texts as well as the obligatory details about how to opt out of the collection and how data may be shared with third-party companies. One thing that does differ between companies is that even though they may include the same reasons for the collection, they choose to highlight different reasons and provide illustrative examples to explain why they are collecting data.

Table 1. The Media Websites and Texts Included in the Sample Ranked According to SimilarWeb (October 4, 2016).

Rank	U.S.	Type of text	UK	Type of text	Sweden	Type of text
1	yahoo.com	PP	Bbc.co.uk	CP	aftonbladet.se	CP
2	msn.com	PP	yahoo.com	PP	expressen.se	CP
3	cnn.com	CP	dailymail.co.uk	PP	yahoo.com	PP
4	news.google.com	PP	theguardian.com	CP	dn.se	CP
5	espn.com	N/A	msn.com	PP	svd.se	CP
6	nytimes.com	PP	telegraph.co.uk	PP	msn.com	PP
7	huffingtonpost.com	PP	mirror.co.uk	CP	idg.se	CP
8	foxnews.com	PP	independent.co.uk	PP	di.se	CP
9	washingtonpost.com	PP	skysports.com	CP	gp.se	CP
10	buzzfeed.com	PP	newsnow.co.uk	PP	sydvsenskan.se	CP
11	usatoday.com	PP	express.co.uk	CP	feber.se	CP
12	latimes.com	PP	thesun.co.uk	PP	omni.se	PP
13	nbc.com	PP	metro.co.uk	PP	avpixlat.info	N/A
14	cnet.com	PP	ibtimes.co.uk	PP	nyheter24.se	CP
15	businessinsider.com	PP	huffingtonpost.co.uk	CP	metro.se	CP
16	forbes.com	CP	news.google.co.uk	PP	vk.se	N/A
17	nypost.com	PP	Dailystar.co.uk	CP	svt.se	CP
18	sfgate.com	PP	standard.co.uk	PP	hd.se	CP
19	chicagitribune.com	PP	manchestereveningnews.co.uk	CP	corren.se	CP
20	thehill.com	PP	liverpoolecho.co.uk	CP	dt.se	CP

Note. PP = privacy policy; CP = cookie policy.

Results

As with any other website, the audience grants the media companies its consent to be monitored while consuming the news. In this study, 25 reasons for why media companies are collecting data were identified in privacy documents found on the websites of 60 media companies in the U.S., UK, and Sweden. First, the general style of the texts is discussed to illustrate the difficulty of finding explicit reasons in the texts. Second, the 25 identified reasons for collecting data are presented as six aggregated reasons with examples from the analyzed texts.

Style of the Documents

U.S. and British privacy policies are formal and usually at an early point contain definitions of concepts that are used throughout the policy. The Swedish texts are less formal in that the majority of the Swedish news websites included in the sample did not have a privacy policy at all, but rather a cookie policy that used less formal language. Even though the cookie policies are less formally written, they often contain more detailed information about the tracking technologies used and how to opt out from them.

Most of the analyzed texts contain detailed descriptions of the data that are being collected, but very little text describing the reasons why. The texts were often written in a positive, informative tone. For example, *Aftonbladet* ("Personuppgiftspolicy," 2016), the largest newspaper in Sweden, starts its privacy policy by describing how important it is for users to trust the newspaper and that it is concerned about safeguarding its users' integrity, right before the newspaper switches over to more legally bound language. In general, all of the analyzed texts use positive words when describing what was being collected—for example, "sharing data" with the company or phrasing the collection of data as an offer: "When you register or otherwise interact with the Services, you may be invited to provide personal information to enhance your experience on our site" ("Privacy Policy," 2016g, para. 3).

The words *invited* and *enhance* emphasize the advantages associated with the user's activities being monitored, and the implication is that the collection of data is being carried out in the user's interest rather than for the benefit of the news company. In the same positive manner, the motivation for data collection is sometimes described as an offer to the user, as for example in this quote from the Cookie Policy of SVT, the Swedish public television company: "To give you, as a visitor, the best possible experience, SVT uses cookies" ("Cookies och personuppgifter," 2016, para. 1).

Several companies broke the collected data down early in the texts into two categories—personal and impersonal—when distinguishing between, for example, registering for a service (personal) and automatic collection when the audience is using the service (impersonal). Some even used the term *anonymous* for describing data that are collected from users without the users being aware that such data are being recorded when interacting with a service. In the following example, *The New York Times* ("Privacy Policy," 2016d) described the two types of data as follows:

> The information gathered when you interact with the NYT Services falls into two categories: 1) Personal information, which includes personal information you supply

> when you subscribe, order, complete a survey, register for one of our sites, enter a contest or provide your email address and 2) Non-personal information collected through technology, which includes tracking information collected by us as well as third parties. (para. 9)

However, further down in the text, *The New York Times* states that it combines the nonpersonal data with personal data and further clarifies that nonpersonal data encompasses (e.g., IP addresses, geolocation information, and unique device identifiers). When such information is combined with personal data, personal information may potentially be revealed. This ambiguity (i.e., the introduction of safety early on in the text that is then taken back with more complicated technological explanations) was observed in several of the analyzed texts. Nevertheless, the overall data collection processes of the companies are described positively and in a well-thought-out manner. Here, the distinction between personal and impersonal data is most likely an attempt by the companies to make the user feel secure by reducing the fear of technology paternalism.

Paternalism is present in all of the analyzed privacy texts since they all in some way express that the user has agreed to the stated terms involving the various types of decisions that will be made for the user by merely accessing the site. Sentences focused on passive consent are standardized, and a typical example is shown in this quotation: "By using the Services, you agree that your use of our Services, and any dispute over our online privacy practices, is governed by this Privacy Notice" ("Privacy Policy," 2016f, para. 2).

All of the analyzed texts obtained consent from users in this way (i.e., by stating that if usage occurs, then consent is implied). The Swedish media company IDG points out that because the information about data collection is available to everyone who visits the IDG Web pages, the assumption may be made that all users have read the information. Similarly, Schibsted ("Privacy Policy," 2016e), a large Norwegian media company operating on the Nordic market, states that if the users are logged in to one service, and if they agree to auto-login, they have accepted the privacy and cookie policies of all of Schibsted's websites and services. Thus, even if users visit a new service that they have not previously visited, consent for data collection is assumed. Presumably, as survey studies suggest, the majority of the users have not read the terms of use stated in the privacy policy, making these acts therefore paternalistic.

In the majority of the texts, the user has a central role in the description of the data collection. Furthermore, in 30% of the texts, the provided reasons focus on the comfort of the user. In the following quote, the motivation for why data are collected puts the user's needs first and presents the data collection as something the users themselves have asked for:

> By using your information, we can provide the product or service you've asked for. The use of your information helps us understand what your needs and interests are, provide personalized content and match the most relevant adverts and services for you. ("News UK Privacy," 2016, para. 3)

Google News UK ("Privacy Policy," 2016b), in a text available from the link "Learn More" in a pop-up window, provides a reversed account for usage of collected data, presenting what the user may see as a result of the data that has been collected, thus implicitly telling the user that such activities have been recorded: "You may see suggestions based on your Google Web and App Activity and YouTube activity. That includes topics you've searched for, pages you clicked from the search results page, and topics you've asked to get updates on from Google Now" (para. 6).

By saying that the data are being collected for a useful service, the user may accept that the data collection is in their interest, should he or she read the privacy statement. Possibly aiming to gain a similar positive association, *The New York Times* ("Privacy Policy," 2016d) states in its privacy policy that it has received "the TRUSTe's Privacy Seal," followed by a logo of this particular seal, signaling to the users that its processes and privacy text have been reviewed and approved by some sort of authority.

Reasons were given early in the texts, right at the start of the text, or clearly stated in bullet point lists, although the reasons were also sometimes immersed in complicated explanations of when collection takes place or what data are collected. In the latter two cases, it was harder to identify the reasons for collecting data since the focus of the sentences was on how the data are collected and the way the data would be used was mentioned in subordinate clauses.

Reasons for Collecting User Data

The number of reasons found in the texts varied somewhat across the countries. In the U.S. privacy texts, 21 categories were found; in the Swedish texts, 17; and in the UK texts, 24. In total, 25 overall reasons were identified. Table 2 presents the 25 reasons for collecting data, aggregated into six groups.

Table 2. Aggregated Reasons for Collecting User Data, Found in Privacy and Cookie Texts as a Percentage of the Total Number of Reasons Found (N = 60).

Aggregated reasons	*N*	Included categories of reasons
Personalization and enhanced user experience	140	Enhance user experience, make recommendations for the user, personalize or customize ads or content, provide location-based services
Delivery and maintenance of services	97	Deliver the product or service, remember user preferences for log-in, polls, etc.; provide what the user requests; maintain the site; deliver on multiple platforms; save time; enhance efficiency for the user
Internal corporate use of data	50	User metrics and research, enhance shopping experience, develop new products and services, provide services for free, generate revenue
Legal reasons	39	Fulfill laws and regulations, allow comments or content-sharing, prevent fraud, protect copyrights, keep the user safe, send information to the user about the user relationship
Communicate with the user	36	Contact the user, answer questions
Third-party use of data	20	Share data with third parties

Reasons found in four of the six categories were frequently described in a paternalistic manner, whereas the two categories—*legal reasons* and *communicate with the user*—in general were not explicitly paternalistic. Nevertheless, paternalistic exceptions could also be found in these categories. The six categories are illustrated below with quotes from the analyzed texts.

Personalization and Enhanced User Experience

This category includes the most commonly stated reasons. Within the first category, we find the three most common single reasons found in the analyzed documents: personalization of advertisements, found in 80% of the privacy texts; enhanced user experience, found in 65% of the analyzed texts; and personalized content, found in 62% of the analyzed texts. Personalization was in the texts described as in the users' interest. However, acts of personalization can also be interpreted as punishment. The BBC ("Your Information," 2016), for example, describes how the data it collects are used for personalized services, such as recommendations, based on viewing history on their video players or news consumption habits, such as checking the weather forecast for a specific region on BBC Online. The BBC may then choose to present the selected information for the user on their homepage the next time the user visits the site or detect the user's location to automatically send international viewers to the international version of the BBC Web pages. These paternalistic decisions are described in a justifying manner. However, taking the case of the international BBC user who would like to access the bbc.co.uk version of the BBC, the strict paternalistic act of automatically choosing the location will exclude such users from this particular content. Similarly, *Google News* states that it is collecting data about users to "figure out which language you speak" ("Privacy Policy," 2016b, para. 3). This particular example indicates that paternalistic decisions are being made by the content provider. Depending on what the content provider finds, it will make decisions, even though the user did not ask for a geographical judgement and subsequent geo-tailored content when accessing the service. Also, *Yahoo News* uses geo location to present content to the user, although it does reduce the paternalistic effect by offering the user the opportunity to select the country.

Another type of personalization was described by British *Yahoo News* ("Yahoo Privacy Center," 2016):

> Yahoo News Activity allows you to automatically share the Yahoo News articles you view with your friends and easily discover articles that are interesting to them. When you opt-in to Yahoo News Activity, Yahoo will automatically share links to the Yahoo News articles you view to both your friends on Facebook and to your Facebook friends on Yahoo News. (para. 51)

This soft paternalistic personalization feature consists of a process where the user has control through an opt-in solution for personalization rather than the more common preselected opt-out solutions that are described in the majority of the privacy texts, regardless of what service companies offer.

Delivery and Maintenance of Services

This category contains several technology-related aspects of collecting data. Here, a common way of describing the collection was to refer to the technology itself and emphasize how the website needs the information, as if it were a person:

> For almost any modern website to work properly, it needs to collect certain basic information on its users. To do this, a site will create files known as cookies—which are small text files—on its users' computers. These cookies are designed to allow the website to recognise its users on subsequent visits. ("Privacy Policy," 2016c, para. 68)

This description takes the form of technology paternalism, describing the collection from the perspective of the technology and potentially inducing a feeling of lost control for the users. In the example above, neither the users nor the company collecting the information controls the act of collection; rather, it is the site itself that takes on a life of its own while managing the collection of data. Found in 35% of the texts, the cookie technology is referred to as a memory that is used by the company or the website to "remember" the user and preferences that have been either actively stated or passively recorded through previous actions. Dalarnas Tidningar (2016) describes this in a positive manner: "The cookie will make it possible for your device to be recognized. In this manner, you will save time when you surf" (para. 3).

The argument that cookies will save time may be true, but the time saved is possibly a matter of milliseconds. Another way to justify the act of tracking the audience is by simply stating that most sites use cookies, making the user feel that the use of cookies is both normal and justified. Describing the process from the perspective of the site is of course technologically correct, but unnecessary, as the process is at some point authorized by humans and ultimately controlled by humans.

Detailed descriptions of technology were found in over half of the texts in the sample. In 10 of the texts, however, no details at all were given about technology. Five of these were Swedish news companies, four were British, and one American. Notes were found in some of the detailed descriptions of other tracking technologies about not responding to "do not track" signals. This implies that active privacy measures taken by the user, for example, in browser settings, are overruled. By accepting the privacy terms, this hard, paternalistic decision taken by the content provider—to ignore privacy measures that the user has decided on beforehand—could, if read, influence the user to refrain from using the site or service.

Internal and Corporate Use of Data

Internal use of data was expressed, for example, in terms of understanding site usage and demographics. The uses for this category were often described as circular references, for example, to collect data to produce statistics. No particular analysis or deeper meaning of why the company is in need of statistics were usually described in close connection with the reasons in this category, and therefore such uses did not explicitly imply paternalism. Only two news companies mention explicitly a transaction

of money as a reason for data collection, *The Guardian* and Sky Sports. In the following quote, *The Guardian* ("Privacy," 2016a) mentions revenue explicitly:

> One of the ways we generate revenue to safeguard our journalism for the future is by using your data (for example when you sign-in to the Guardian's websites, and/or by using cookies when you browse our website) to make advertising more relevant to you.
>
> This enables us to charge advertisers money, which helps us keep the website open to all, while at the same time making the Guardian and Observer's own products and services more relevant to you. (para. 1)

This transparent explanation of the major reason behind why news media companies collect behavioral data from their audiences blames neither the technology, third parties, nor the users themselves for wanting the data collection to take place. The simple act of describing the transaction of money may reduce the fear of paternalism, since users will understand that news is not free. Somewhat less explicit, but still describing a transaction caused by the collection of behavioral data with the use of cookies, is the privacy notice in Sky Sports ("Sky Account Security," 2016): "We sell space on some of our websites to advertisers. The resulting adverts often contain cookies" (para. 64).

In these two sentences the cookies are thus linked to the implicit revenue that the company can make from selling ad space. In particular, the analyzed American news companies mention that advertisements finance their business but do not explain this relationship as a reason for collecting behavioral data.

Legal Reasons

This category seldom included illustrative examples, and reasons were written in a legally bound, strict language. The reasons, for example, were to prevent fraud or copyright infringement, to comply with the law, to respond to requests from the authorities, or to protect the safety of the users, but the texts did not state how the data are used to achieve such reasons. Because the reasons were only mentioned, they were not described in actions that could be labeled as paternalism. For example, Schibsted ("Privacy Policy," 2016e) mentions that it collects data to prevent misuse: "We use personal information about user activities and technological data to prevent, limit or investigate different types of misuse of our Services. We define misuse as establishing fake profiles, spamming, harassment, contrary to the User terms" (para. 148).

Communicate With the User

This category was found mostly in the U.S. documents; more than half of the U.S. texts stated that data are collected in order for the company to contact the user. Examples of situations where the users can be contacted were customer service related or to notify a user that he or she had won a contest, about surveys, or simply to receive newsletters. Depending on how the company chooses to contact the

user, the contact itself could potentially be intrusive, but perhaps not paternalistic. For example, the *Los Angeles Times* ("Privacy Policy and Your Privacy Rights," 2016) states:

> We use the information that we collect for the following purposes: . . . To send you alerts or other communications via SMS messages or other media or networks. . . . To contact you with information that we believe will be of interest to you. (para. 2)

The vague description of when users can be contacted in combination with the use of SMS might imply an intrusion, but not necessarily in a paternalistic manner, as the user most likely consciously has submitted contact details to the company. The Swedish news organization IDG ("Finstilt," 2016) states, however, in a slightly more paternalistic manner, that merely opening a newsletter or clicking on a link may result in a user being contacted by the IDG organization.

Third-Party Use of Data

The act of sharing collected data with third parties was generally included in the analyzed privacy texts; however, only one third of the texts mentioned third-party use of behavioral data as a reason for collecting data. Similar to how technology is described, privacy policies mention third parties as middlemen that may analyze collected behavioral data like a black box (i.e., out of the company's control). For example, the policy of Swedish website Nyheter24 ("Integritetspolicy," 2016), begins with a detailed description of what data third parties are in need of and avoids the reason behind usage of such data. It then goes into the site's own uses for the information:

> The measurements are used to validate the information given by the website to the advertisers regarding number of visitors, size of traffic, target groups reached and outcome of campaigns. . . . We use cookies to enhance usage of our websites and to personalize parts of the content. (para. 10, translated)

These reasons thus enhance the uncertainty of what is being done with the collected data, since the data are collected and analyzed by someone other than the news company.

One type of technology, Web beacons, was mentioned in several of the documents in relation to third-party sharing of data. Web beacons were mainly described as something positive, for example, to get relevant advertisements, but they were also found to make it possible for ad networks to view, edit, or set their own cookies as if the user had requested a Web page from their site ("Privacy Policy," 2016a). In this case, the paternalistic act is aided through technology, but blamed on the ad network companies.

Reasons in this category often took the form of a punishment. For example, in the policy document of SfGate (2016), the user is informed that it is possible to block cookies, although doing so will exclude the user from certain content, thus punishing the user for not accepting the monitoring of their behavior. The following example illustrates how it is possible to opt out from targeted advertising, but the user is simultaneously informed that it is not possible to prevent information from being collected for other purposes, suggesting a form of hard paternalism: "You can opt out of this 'targeted advertising.' The

effect of opting out is controlled by third parties; they may still collect your usage information for analytic, research or internal operations purposes" ("Privacy," 2016b, para. 30).

Facebook often had its own space in privacy policies. Notably, no reasons were found for why the news media companies collect Facebook profile information. Instead, the detailed privacy policies informed users that if they have accessed the news site by clicking on a link in their Facebook flow, the news company will collect data provided by Facebook, such as profile information. Furthermore, in some instances, privacy texts also expressed that the news media companies may give information on their site usage back to Facebook. Detailed descriptions are provided about the data that may be collected from social networks and that this data may be combined with data, often impersonal, that the news company may already possess. Note that the information about third-party sharing, and social networking sites in particular, is usually stated at the end of these documents, with the BBC as the lone exception, stating already in the pop-up window when entering the site for the first time that cookies from social networking sites are used for user experience purposes.

Conclusions

Journalists often act in a paternalistic manner when producing news because journalistic choices are in the interest of individuals and are there "to give people the tools they need to flourish in a democracy and address the issues that prevent people from doing so" (Thomas, 2016, p. 96). While this type of paternalism is not considered in this study, from the perspective of the user, different manifestations of paternalism, in the news content and in accompanying corporate texts, such as privacy policies, are blended in the overall experience of using digital news services. For this reason, it is therefore also important that reasons for why media companies collect data on its users are manifested in a transparent manner and justified with reasons that the audience can relate to and are in tune with journalistic ideals. Otherwise, trust may be compromised. As this study assessed reasons for data collection, the normative nature of the study is a limitation of the selected research design and future research could investigate how reasons in policy documents are also understood by the general public.

The findings of this study indicate that privacy policies of media companies seem to be formulated using a similar approach as any company with a digital presence. Similar to Tuchman's (1978) description of the in-between position news workers strive to have between governmental power and the public, in this case the position between the media corporation's power and the journalistic context of the website is disrupted since the news media—which is often claimed to be open and transparent—may treat its audiences just as any corporation does in user agreements. This is troubling, since Karlsson (2011) argues that the news industry "feeds on audience trust" (p. 292). A solution would be, as Karlsson (2011) suggests, to bring different transparency activities forward to the audience to revitalize journalism both as a profession and as an authoritative source of information.

Because survey studies suggest that users in general have not read privacy policies, they are not in general aware that they have agreed to data collection, and they are not familiar with the detailed reasons stated by the companies. The nature of how consent was obtained is thus the determining factor if reasons are paternalistic or not. Consent is related to concepts of control and self-determination

(Bechmann, 2014), and limited individual control is a component in paternalistic decisions that affect an individual. Privacy notices that appear on websites constitute one example of a "nudge," as described by Thaler and Sunstein (2008), where a message, often in the form of a pop-up, is presented to the user to prompt a choice about accepting the user terms. If this nudge is not provided, as was often the case on the analyzed sites, we cannot speak of an available choice. This particular design solution goes against current legislation, as well as the idea of libertarian paternalism as suggested by Thaler and Sunstein (2008), where it is important to protect the autonomy of the individual. As long as passive consent is in use, all of the reasons stated in the analyzed privacy policies may thus be regarded as hard paternalism, since the reasons clearly interfere with the user's rights to make voluntary decisions, and they are furthermore formulated as if the collection of data is actually in the interest of the user. The implicit message transmitted through the reasons at present also implies soft paternalism, since the method for collecting the data is coercive and users must accept privacy policies if they wish to have access to the news content. Nonetheless, even if the ability to choose would increase, such as if users could actively allow themselves to be tracked when consuming news, there could still be elements of technological paternalism present. Spiekermann and Pallas (2006) suggest that bringing all choices to the foreground for people to assess individually could actually imply a form of coercive selection, since every action would require approval. Thus, there is a fine line between how much choice a company can torment the audience with and the number of choices that are necessary for people to still feel that they are in control of their privacy when consuming news.

The analysis shows that reasons for collecting data can be formulated in both paternalistic and nonpaternalistic ways, but previous research has found that the illusion of free choice (i.e., still a form of soft paternalism) makes people more willing to share data (Kerr et al., 2006). The reasons for collecting data may thus indeed be paternalistic in a positive sense (i.e., beneficial to users)—for example, when used to enhance the user's experience or personalizing content tailored to individual audience members—and, at the same time, paternalistic in a negative sense, as choices may be imposed on users although users have not actively agreed, thus potentially resulting in an undesired outcome.

Paternalistic actions, as they are described in the analyzed texts, involve personalization of news flows, advertisements, website design or functionality, as well as the sharing of data with third-party business partners. Since 60% of the Swedish population are opposed to news companies collecting data to enhance the user experience (Appelgren & Leckner, 2016), it is therefore troubling that enhancing user experience was one of the most common reasons found in the privacy texts of the news websites active on the Swedish market. Furthermore, the act of cradling the audience into believing that it is safe to submit their digital traces, or that data are impersonal, while carefully nested sentences later contradict such statements—for example, when Facebook profile data can be processed together with impersonal news site behavioral data—contributes to a somewhat questionable practice that may not be in tune with journalistic ideals, as "journalism serves a purpose above and beyond its immediate commercial audience" (Tandoc & Thomas, 2015, p. 252) This is an example of how the market-driven logic of Web analytics seems to collide with ethics in a journalistic context.

I argue that the current practice of presenting reasons in privacy policies is at odds with the journalistic context and may result in what Nissenbaum (2015) describes as an inappropriate information

flow, since people expect journalism to be in their interest rather than in the corporate interest of the news media company. Hence, the norms of a more market-driven culture that signifies the origins of Web analytics (Tandoc & Thomas, 2015) dominates what the stated reasons in the texts may imply in terms of outcome. A more explicit description of revenue, as already provided by *The Guardian* ("Privacy," 2016a) as the main reason for collecting audience data, would increase the transparency into why advertisements and third parties are given such prominence in privacy policies and cookie texts and perhaps would also be more in line with how users may interpret the context-specific values and norms of journalism. However, only two of the analyzed documents mentioned transactions of money explicitly as a reason for collecting data. Other identified reasons, such as sharing data with third parties or personalizing advertisements, are implicitly referring to making revenue, albeit using different wording. Similarly, justifying the functionality of the website in terms of user experience may implicitly be derived from generating revenue, since the satisfaction of users is crucial for encouraging them to stay on the website.

Given the pending EU regulation, we can reasonably expect that the act of how companies ask for consent will soon change, and users might in the future be alerted in a clearer way about data collection. This regulation may thus reduce the knowledge gap between companies and users in that people will realize they are being monitored, and the reasons for why collection takes place may then possibly begin to matter more to users.

References

Allen & Overy LLP. (2016). *The EU general data protection regulation*. Retrieved from http://www.allenovery.com/SiteCollectionDocuments/Radical%20changes%20to%20European%20data%20protection%20legislation.pdf

Appelgren, E., & Leckner, S. (2016). Att dela eller inte dela—Internetanvändarnas inställning till insamling av personlig data [To share or not to share—Attitudes toward the collection of personal data among Internet users]. *Telecommunications Policy*, *38*, 1–17.

Bechmann, A. (2014). Non-informed consent cultures: Privacy policies and app contracts on Facebook. *Journal of Media Business Studies*, *11*(1), 21–38. doi:10.1080/16522354.2014.11073574

Boczkowski, P. J., & Peer, L. (2011). The choice gap: The divergent online news preferences of journalists and consumers. *Journal of Communication*, *61*(5), 857–876. doi:10.1111/j.1460-2466.2011.01582.x

Borgesius, F. J. Z. (2015). Personal data processing for behavioral targeting: Which legal basis? *International Data Privacy Law, 5*(3), 163–176. doi:10.1093/idpl/ipv011

Campbell, J. E., & Carlson, M. (2002). Panopticon.com: Online surveillance and the commodification of privacy. *Journal of Broadcasting & Electronic Media*, *46*(4), 586–606. doi:10.1207/s15506878jobem4604_6

Clarke, S. (2002). A definition of paternalism. *Critical Review of International Social and Political Philosophy*, *5*(1), 81–91.

Cookies och personuppgifter [Cookies and personal data]. (2016). *SVT*. Retrieved from https://kundo.se/org/svt-tittarservice/d/cookies-och-personuppgifter/

Couldry, N., & Turow, J. (2014). Advertising, big data and the clearance of the public realm: Marketers' new approaches to the content subsidy. *International Journal of Communication*, *8*, 1710–1726.

Dalarnas Tidningar. (2016). *Cookies*. Retrieved from http://www.dt.se/information/cookies

Dworkin, G. (1972). Paternalism. *The Monist*, *56*(1), 64–84.

Dworkin, G. (1983). Paternalism: Some second thoughts. In R. Sartorius (Ed.), *Paternalism* (pp. 105–112). Minneapolis, MN: University of Minnesota Press.

Dworkin, G. (2002). Paternalism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from https://stanford.library.sydney.edu.au/entries/paternalism/

Dworkin, G. (2015). Defining paternalism. In T. Schramme (Ed.), *New perspectives on paternalism and health care* (pp. 17–29). Hamburg, Germany: Springer International Publishing.

European Commission. (2016). *Regulation (EU) of the European Parliament and of the Council of 27 April 2016*. Retrieved from http://eur-lex.europa.eu/eli/reg/2016/679/oj

European Parliament. (2016). *The EU general data protection regulation*. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC

Finstilt [Fine print]. (2016). *IDG*. Retrieved from http://www.idg.se/2.1085/1.635883

GB consumer privacy index 2016. (2016). *TRUSTe*. Retrieved from https://www.truste.com/resources/privacy-research/ncsa-consumer-privacy-index-gb/

Haybron, D., & Alexandrova, A. (2013). Paternalism in economics. In C. Coons & M. Weber (Ed.), *Paternalism: Theory and Practice* (pp. 157–177). Cambridge, UK: Cambridge University Press.

Hintz, A., & Dencik, L. (2016). The politics of surveillance policy: UK regulatory dynamics after Snowden. *Internet Policy Review*, *5*(3), 1–16. doi:10.14763/2016.3.424

Hofmann, B. (2003). Technological paternalism: On how medicine has reformed ethics and how technology can refine moral theory. *Science and Engineering Ethics*, *9*(3), 343–352. doi:10.1007/s11948-003-0031-z

Integritetspolicy [Integrity policy]. (2016). *Nyheter24*. Retrieved from http://nyheter24gruppen.se/integritetspolicy/

Karlsson, M. (2011). The immediacy of online news, the visibility of journalistic processes and a restructuring of journalistic authority. *Journalism*, *12*(3), 279–295. doi:10.1177/1464884910388223

Kerr, I. R., Barrigar, J., Burkell, J., & Black, K. (2006). Soft surveillance, hard consent. *Personally Yours*, *6*, 1–14.

Le Grand, J., & New, B. (2015). *Government paternalism: Nanny state or helpful friend?* Princeton, NJ: Princeton University Press.

Marx, G. T. (2015). *Windows into the soul: Surveillance and society in an age of high technology*. Chicago, IL: University of Chicago Press.

Mols, A., & Janssen, S. (2016). Not interesting enough to be followed by the NSA: An analysis of Dutch privacy attitudes. *Digital Journalism*, 1–22. doi:10.1080/21670811.2016.1234938

News UK privacy and cookie policy. (2016). *The Sun*. Retrieved from http://www.newsprivacy.co.uk/single/

Nissenbaum, H. (2015). Respecting context to protect privacy: Why meaning matters. *Science and Engineering Ethics*, 1–22. doi:10.1007/s11948-015-9674-9

Personuppgiftspolicy [Personal data policy]. (2016). *Aftonbladet*. Retrieved from http://www.aftonbladet.se/hjalpinfo/vanligafragor/article23620200.ab

Pope, T. M. (2003). Counting the dragon's teeth and claws: The definition of hard paternalism. *Georgia State University Law Review*, *20*, 659–722.

Privacy. (2016a). *The Guardian*. Retrieved from https://www.theguardian.com/info/privacy

Privacy. (2016b). *NBC*. Retrieved from http://www.nbcuniversal.com/privacy/

Privacy policy. (2016a). *Buzzfeed*. Retrieved from https://www.buzzfeed.com/about/privacy

Privacy policy. (2016b). *Google*. Retrieved from https://www.google.com/intl/en_uk/policies/privacy/

Privacy policy. (2016c). *The Independent*. Retrieved from http://www.independent.co.uk/service/privacy-policy-a6184181.html

Privacy policy. (2016d). *The New York Times*. Retrieved from http://www.nytimes.com/content/help/rights/privacy/policy/privacy-policy.html

Privacy policy. (2016e). *Schibsted*. Retrieved from https://login.schibsted.com/privacy

Privacy policy. (2016f). *USA Today*. Retrieved from http://static.usatoday.com/privacy

Privacy policy. (2016g). *The Washington Post*. Retrieved from https://www.washingtonpost.com/privacy-policy/2011/11/18/gIQASIiaiN_story.html?utm_term=.0b3b18a49daf

Privacy policy and your privacy rights. (2016). *Los Angeles Times*. Retrieved from http://www.tronc.com/privacy-policy/

SfGate. (2016). *Privacy policy*. Retrieved from http://www.sfgate.com/privacy-policy

Sky account security. (2016). *Sky*. Retrieved from https://www.sky.com/help/articles/sky-privacy-and-cookies-notice

Special Eurobarometer. (2015). *Data protection Eurobarometer: Factsheet*. Retrieved from http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_eurobarometer_240615_en.pdf

Spiekermann, S., & Pallas, F. (2006). Technology paternalism—Wider implications of ubiquitous computing. *Poiesis & Praxis*, *4*(1), 6–18. doi:10.1007/s10202-005-0010-3

Tandoc, E. C., Jr., & Thomas, R. J. (2015). The ethics of Web analytics: Implications of using audience metrics in news construction. *Digital Journalism*, *3*(2), 243–258. doi:10.1080/21670811.2014.909122

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Thomas, R. J. (2016). In defense of journalistic paternalism. *Journal of Media Ethics*, *31*(2), 86–99. doi:10.1080/23736992.2016.1152895

Top websites [Data set]. (2016). *SimilarWeb*. Retrieved from https://www.similarweb.com/top-websites/category/news-and-media

Tuchman, G. (1978). *Making news: A study in the construction of reality*. New York, NY: Free Press.

U.S. consumer privacy index 2016. (2016). *TRUSTe*. Retrieved from https://www.truste.com/resources/privacy-research/ncsa-consumer-privacy-index-us/

Weiser, M., & Brown, J. (1997). The coming age of calm technology. In P. J. Denning & R. M. Metcalfe (Eds.), *Beyond calculation* (pp. 75–85). New York, NY: Springer Science and Business Media.

Yahoo privacy center. (2016). *Yahoo News*. Retrieved from https://policies.yahoo.com/us/en/yahoo/privacy/index.htm

Your information and privacy. (2016). *BBC*. Retrieved from http://www.bbc.com/usingthebbc/privacy/