Notified but Unaware: Third Party Tracking Online

Larsson, Stefan; Jensen-Urstad, Anders; Heintz, Fredrik

Published in:
Critical Analysis of Law

2021

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Larsson, S., Jensen-Urstad, A., & Heintz, F. (2021). Notified but Unaware: Third Party Tracking Online. *Critical Analysis of Law*, *8*(1), 101-120. https://cal.library.utoronto.ca/index.php/cal/article/view/36282

Total number of authors:
3

Notified But Unaware: Third-Party Tracking Online

Stefan Larsson, Anders Jensen-Urstad & Fredrik Heintz*

Abstract

Drawing from conceptual studies on transparency, particularly with regard to market complexity, user literacy and resignation, this article studies and analyzes the practices of third-party data collection online. Empirically, we map third-party trackers on a sample of Swedish websites in five sectors (media, retail, banking/insurance, public sector, and health), and the trackers are compared to lists of known trackers to determine their main purpose. These results are then used in digital focus groups divided into high-trust and low-trust individuals, to better understand how everyday consumers perceive and cope with the tracking infrastructure. The main results indicate that third-party tracking is omnipresent online, particularly for media and retail sites, showing that data-driven markets depend on the collection, sharing and trade of consumers' personal data. Furthermore, despite regulatory provisions promoting clearer notifications and designs within the EU, many users are still highly unaware of the data collection practices and their underlying purposes. The results indicate a weak consent base for data collection and a market structure and data collecting practices that are highly non-transparent. We recommend more active supervisory authorities, and more stringent requirements primarily in relation to the obscure ad tech infrastructures to improve transparency and promote consumer awareness.

I. Introduction and Purpose of Study

Most web users are, to some extent, tracked by a large number of companies which they are likely to have never heard of. The modern web consists of an intricate network of actors who collect, purchase, convey, and convert data in various ways for a large number of purposes.[1] A common method used in the actual data collection is that the website visited uses cookies, i.e., small text files placed on the user's device, containing, for example, a unique identifier. This makes it possible for third parties—actors who are neither the user nor the visited website—to collect information about users for a variety of purposes. These purposes range from providing enhanced functionality and personalization to tracking browsing habits and providing digital advertising markets with raw material that can be used

[*] Stefan Larsson, Associate Professor in Technology and Social Change at LTH, Lund University, Sweden; Anders Jensen-Urstad, Dataskydd.net; Fredrik Heintz, Associate Professor in Computer Science at Linköping University, Sweden. Acknowledgments: This work was partially supported by (i) the Wallenberg AI, Autonomous Systems and Software Program—Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation, (ii) the Swedish Competition Authority, as well as (iii) The Swedish Retail and Wholesale Council.

[1] Wolfie Christl, Corporate Surveillance in Everyday Life 6 (2017).

ISSN 2291-9732

for targeted advertising. Marketing purposes are a significant driving force for the collection infrastructure, but the collection is also utilized by, for example, data brokers for markets focused on risk assessments in credit and insurance.[2] Many studies on third-party tracking are concerned mainly with privacy, data protection and security issues, while showing that transparency and consumer awareness tend to be low.[3] This article focuses on the web's data collection infrastructure in relation to consumer awareness, under the hypothesis that these two sides are highly disentangled, making the consumers remain ignorant about the extent of tracking, regardless of formal consent. It has been noted that the first party, i.e., the person who provides the website, often also does not know which third parties are embedded[4] for example, when these third parties request additional content, as is commonly done in real-time ad auctions. This obviously further reduces the possibility for end users to understand and control what data is collected about them. It is reasonable to assume that it also affects the working conditions of supervisory authorities.

The ethical aspects of the lack of real consent and low consumer awareness of the presence of tracking has been discussed for many years.[5] Experiment-based studies have recently shown that behavioral design, so-called dark patterns, is used to get consumers to click-consent to third-party tracking,[6] and that "consent fatigue" is widespread.[7] Given how fundamentally important access to some websites is to consumers, the actual choice offered may be negligible.[8] EU regulation in the area of third-party tracking through cookies,

[2] Id.; Kevin Mellet & Thomas Beauvisage, Cookie Monsters: Anatomy of a Digital Market Infrastructure, 23 Consumption Mkts. & Culture 110 (2020); see also Bruce Schneier, Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World (2015); Jockum Hildén, The Politics of Datafication: The Influence of Lobbyists on the EU's Data Protection Reform and Its Consequences for the Legitimacy of the General Data Protection Regulation, Publications of the Faculty of Social Sciences, University of Helsinki (2019).

[3] Timothy Libert & Reuben Binns, Good News for People Who Love Bad News: Centralization, Privacy, and Transparency on US News Sites, in Proceedings of the 10th ACM Conference on Web Science 155 (2019); Elena Maris et al., Tracking Sex: The Implications of Widespread Sexual Data Leakage and Tracking on Porn Websites (2019).

[4] Tobias Urban et al., Beyond the Front Page: Measuring Third Party Dynamics in the Field, in Proceedings of The Web Conference 2020, at 1275 (Yennun Huang et al. eds., 2020).

[5] Daniel E. Palmer, Pop-ups, Cookies, and Spam: Toward a Deeper Analysis of the Ethical Significance of Internet Marketing Practices, 58 J. Bus. Ethics 271 (2005); Nalini Elisa Ramlakhan, Ethical Implications of Third-Party Cookies, 9 Int'l J. Human. 59 (2011).

[6] Arunesh Mathur et al., Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites, 3 Proceedings of the ACM on Human-Computer Interaction 1 (Lorenzo Cavallaro & Johannes Kinder eds., 2019).

[7] Paul Graßl et al., Dark and Bright Patterns in Cookie Consent Requests (2020); Christine Utz et al., (Un)informed Consent: Studying GDPR Consent Notices in the Field, in Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security 973 (2019).

[8] Richards & Hartzog suggest the wider area of online consent flaws to be understood in terms of "pathologies." In line with this, they offer terminology on how to understand different kinds of consent problems. For example, "unwitting" consent, where the user does not understand the legal or the technical settings, or their long-term consequences; and "coerced" consent, when access to a particular website in reality is not voluntary for a user at all. See Neil Richards & Woodrow Hartzog, The Pathologies of Digital Consent, 96 Wash. U. L. Rev. 1461 (2018).

primarily the GDPR, therefore does not appear to function as intended.[9] In other words, there is indeed a challenge of making tracking visible so that end-users understand what they are consenting to. And there is a market-based incentive for using manipulative interfaces to obtain the formal consent of site visitors—i.e., where one option is given visual or interactive precedence over others, such as when the option to opt out of online tracking is not presented together with the opt-in option but can only be reached by clicking through several submenus. The design of the consent largely determines whether consumers make an informed decision rather than be manipulated in the abundance of consent requests for the website owners' purposes of securing legal grounds for data collection.[10] This has led to, among other things, the establishment of the Princeton Web Transparency Project,[11] which continually monitored the top one million websites between 2015 and 2019 to uncover what user data companies collect, and why and how.[12] The presence of Google and Facebook is dominant in all of the most popular websites,[13] but there are local variations, for example among the twenty-seven EU countries (and the United Kingdom).[14]

Many studies have been made on how people perceive data collection. There is extensive evidence showing that individuals are worried about data collection practices, according to a recently published overview of existing empirical research in the area of public perceptions of, attitudes to, and feelings about data collection practices.[15] However, while users are worried, they also find ways of negotiating, managing or resisting the data practices in their everyday lives. Kennedy et al. conclude that people often have conflicting opinions about data practices, simultaneously admitting their advantages, while also worrying about damage they may potentially cause.

Lastly, the particular techniques of third-party tracking can also be framed within a platform-based competitive landscape which may be in transition. Google has signaled future changes in the field, for example planning to phase out third-party cookies in their web

[9] Graßl et al., supra note 7.

[10] Utz et al., supra note 7; see the European Data Protection Board's Guidelines on consent under GDPR. EDPB, Guidelines 05/2020 on consent under Regulation 2016/679, adopted on 4 May 2020. Consent should be "freely given" and "informed" with an "unambiguous indication of the data subject's wishes," under Article 4(11) of the GDPR.

[11] Princeton Web Transparency Project (https://webtap.princeton.edu/).

[12] Arvind Narayanan & Dillon Reisman, The Princeton Web Transparency and Accountability Project, in Transparent Data Mining for Big and Small Data 45 (Tania Cerquitelli et al. eds., 2017).

[13] Steven Englehardt & Arvind Narayanan, Online tracking: A 1-million-site measurement and analysis, in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security 1388 (Edgar Weippl ed., 2016).

[14] Rasmus Helles et al., Infrastructures of Tracking: Mapping the Ecology of Third-Party Services Across Top Sites in the EU, 22 New Media & Soc'y 1957 (2020).

[15] Helen Kennedy et al., Public Understanding and Perceptions of Data Practices: A Review of Existing Research—A Summary (Living with Data), University of Sheffield (2020). For the specific Swedish context, there are survey-based studies pointing to an increased unease over the last few years amongst web users, cf. Datainspektionen [The Swedish Data Protection Authority], Nationell integritetsrapport 2019 (2019); Internetstiftelsen. Svenskarna och internet 2019 (2019).

browser Chrome (with a worldwide market share in excess of sixty-four percent).[16] Apple, relying on different business models than for example Facebook and Google, has recently made changes to the design of iOS as well as Safari to visualize third-party behavior.[17] The transitional aspects of the cookie landscape can be placed into discourses of infrastructures and platforms,[18] serving as a historic basis for the ongoing process of "assetization" or commodification of data,[19] an ongoing datafication[20] where personal data are commodified and traded on advertising or other markets. This is a development relevant beyond third-party cookies, feeding into a more monopolistic turn of platforms.[21]

A. Purpose

By drawing from theoretical notions of transparency for datafied and automated processes,[22] outlined further in the following subsection, the purpose of this study is to better understand the prevalence and user awareness of third-party tracking online in five Swedish sectors. The study investigates and aims to answer the following questions:

- What third-party trackers are present on websites aimed at Swedish consumers?
- What categories of third-party trackers are used, i.e., what are their purposes?
- How does the use of third-party tracking vary between various industries, and between websites within each respective industry?
- How do Swedish consumers and users view and approach third-party tracking?
- What is the level of their awareness and how do they justify their approach?

This study is limited to websites, and does not include, for example, collection in apps.[23] The study does not touch upon the methods used to collect data or exactly what type of data is collected, but primarily deals with who the actors are. It focuses on Swedish sectors and therefore the awareness and data literacy of Swedish users.

[16] Damien Geradin & Dimitrios Katsifis, Taking a Dive into Google's Chrome Cookie Ban (2020) (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3541170).

[17] Todd Haselton, iPhone Update Will Expose How Companies Try to Track You, CNBC, Sept. 2, 2020 (https://www.cnbc.com/2020/09/02/apple-ios-14-for-iphone-will-expose-how-companies-track-you.html).

[18] Jean-Christophe Plantin et al., Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook, 20 New Media & Soc'y 293 (2018).

[19] Mellet & Beauvisage, supra note 2; cf. José van Dijck et al., The Platform Society: Public Values in a Connective World (2018); Shoshana Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (2019).

[20] Ulises A. Mejias & Nick Couldry, Datafication, 8(4) Internet Pol'y Rev. (2019).

[21] Nick Srnicek, Platform Capitalism (2017).

[22] See especially Stefan Larsson & Fredrik Heintz, Transparency in Artificial Intelligence, 9 Internet Pol'y Rev. 2 (2020); Stefan Larsson, The Socio-Legal Relevance of Artificial Intelligence, 103 Droit & Société 573 (2019).

[23] For more information on how apps collect and convey data, see Forbrukerrådet, Out of Control: How Consumers Are Exploited by the Online Advertising Industry (2020); Sophus Lai et al., A Proxy for Privacy Uncovering the Surveillance Ecology of Mobile Apps, 7 Big Data & Soc'y 1 (2020).

B. Conceptual Framework: Transparency, Literacy and Resignation

The perspective on transparency for this study is influenced by how transparency has been conceptualized in relation to datafied and automated processes.[24] Particularly with regards to 1) the *complexity* of commercial data ecosystems, as described by both Christl[25] and Pasquale,[26] where the latter uses the black box metaphor to describe data-driven markets and, ultimately, a "black box society"; 2) the intentional opacity[27] in legal aspects of *proprietorship,* as code and data enter competitive markets.[28] This is sometimes found to be the core problem when markets are opaque, for example for data broker markets,[29] and often comes with direct implications for public transparency or the scrutiny from authorities.[30] And, lastly, we include 3) data/algorithm user *literacy* as a key theoretical basis for this study,[31] indicating that ordinary users' basic understanding has a direct effect on transparency in the automated collection and trading of user data.[32] This transparency framework is related to what Draper and Turow have described as a "digital resignation," based on studies of consumers' perception of data collection.[33] Recent studies have also pointed to "consent fatigue," that is, a sort of overflow of questions asking for consent in a digital context.[34] This overflow of information and focus on individual decision-making in an imbalanced setting lead to what Anja Bechmann has referred to as "non-informed consent cultures."[35]

[24] Larsson & Heintz, supra note 22; Larsson, supra note 22.

[25] Christl, supra note 1.

[26] Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information (2015).

[27] Jenna Burrell, How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms, 3 Big Data & Soc'y 1 (2016).

[28] Pasquale, supra note 26; Sandra Wachter et al., Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR, 31 Harv. J.L. & Tech. 841 (2017).

[29] Matthew Crain, The Limits of Transparency: Data Brokers and Commodification, 20 New Media & Soc'y 88 (2018).

[30] Stefan Larsson, Algorithmic Governance and the Need for Consumer Empowerment in Data-Driven Markets, 7 Internet Pol'y Rev. 1 (2018).

[31] Ina Sander, What Is Critical Big Data Literacy and How Can It Be Implemented? 9 Internet Pol'y Rev. 1 (2020).

[32] Burrell, supra note 27.

[33] Nora A. Draper & Joseph Turow, The Corporate Cultivation of Digital Resignation, 21 New Media & Soc'y 1824 (2019); Joseph Turow & Nora Draper, The Tradeoff Fallacy: How Marketers Are Misrepresenting American Consumers and Opening Them Up to Exploitation (2015) (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2820060).

[34] Utz et al., supra note 7.

[35] Anja Bechmann, Non-Informed Consent Cultures: Privacy Policies and App Contracts on Facebook, 11 J. Media Bus. Stud. 21 (2014).

II. Methodologies

In terms of methodologies, we measured the prevalence of third-party trackers on a Swedish sample of websites divided into five sectors. These were compared to lists of known trackers, each tracker belonging to one or more categories, in order to determine the main purpose of the trackers on the sites. Additionally, the results were tested and discussed in digital focus groups with Swedish users that were sampled based on their own indicated levels of trust towards sharing their personal data for commercial purposes with third-party entities. Here follows a more detailed description of how we picked the sample, collected data, categorized trackers, and conducted focus-group interviews.

A. Tracking the Trackers

There are no entirely reliable statistics identifying the most popular websites in Sweden. The so-called KIA-index, acquired by market research company Kantar Sifo from the Association of Swedish Advertisers at the beginning of 2020, refers to itself as "the official measurement currency for Swedish websites." However, it lacks many major websites as for example Schibsted and Bonnier News, two of the largest media companies on the Swedish market, dropped out a few years ago.[36] Amazon-owned Alexa Internet provides the internationally most well-known website ranking. Their top sites list has been used in numerous studies, for example the study by Englehardt & Narayanan.[37] Alexa's figures are based partly on data sent by end users who installed a browser add-on. It remains unclear how reliable these figures are in relation to Swedish internet users.

The list of websites analyzed for this study was prepared by using various selection methods in different market categories. The following five sectors were analyzed:

1. *Media*: The six largest newspaper groups in Sweden constitute one subcategory.[38] From each of these, the five websites with the greatest reach according to ORVESTO's measurement were selected.[39]
2. *Retail*: Each website included in the so-called E-barometer's table of Swedish consumers' "Favorite actors per industry" (e.g., "health and beauty," "groceries," "sports and leisure").[40]
3. *Banking/insurance*: Includes the ten largest commercial banks in Sweden, the five largest non-life insurance companies and the five largest life insurance companies.[41]

[36] Alex Hartelius, Så ska Kia-index hantera avhoppen, Dagens Media, Nov. 17, 2017 (https://www.dagensmedia.se/medier/digitalt/sa-ska-kia-index-hantera-avhoppen-6884222).

[37] Englehardt & Narayanan, supra note 13.

[38] Ulrika Facht & Jonas Ohlsson, MedieSverige 81 (2019) (https://gupea.ub.gu.se/handle/2077/59560).

[39] Kantar Sifo, ORVESTO Internet December 2019, räckvidd digitalt—Total (2020) (https://www.kantarsifo.se/rapporter-undersokningar/rackviddsmatningar/orvesto-internet).

[40] PostNord, Svensk Digital Handel, & HUI Research, E-barometern 2019, at 19 (2020) (https://www.postnord.se/vara-losningar/e-handel/e-handelsrapporter/e-barometern).

[41] Svenska Bankföreningen, Bank- och finansstatistik 2018 5 (2019) (https://www.swedishbankers.se/media/4339/1910_bank-och-finansstatistik_2018.pdf); Svensk Försäkring, Försäkringar i Sverige 2019, at 19-20

4. *Public sector*: Includes the three largest government authorities for each so-called COFOG category,[42] based on adjusted annual workforce 2019, and also the five largest municipalities (by population).[43]
5. *Healthcare/wellbeing*: Five online medical companies,[44] 1177 Vårdguiden (national platform for healthcare information and services), and the Swedish Public Health Agency.

1. Data Collection

The data collection for third-party tracking was carried out on 18 April 2020. To simulate visits by an ordinary computer user, the visits were made with an ordinary consumer browser, controlled by custom-made software using Google's software library Puppeteer.[45] The software is based partly on the software used by the Swedish NGO Dataskydd.net's public web service, Webbkoll,[46] the experiences from the development of this service, and a tool built to examine municipal websites.[47] Puppeteer controls automated instances of the browser Chromium,[48] which is instructed to visit each of the websites. For each visit, the software logs HTTP requests; in other words, the addresses (URLs) contacted by the browser when the website is loaded. There is no user interaction—for example a mouse click or scrolling—with the website itself. The browser setting, "Do Not Track" (DNT), is normally not activated as a standard in browsers, and neither is it here.

Websites and third-party services can treat visitors differently in a number of ways. One such way is geolocalization.[49] Another is cookies and other data[50] that may remain in the browser since the last visit. To counteract this, and, to the extent possible, to emulate

(2019) (https://www.svenskforsakring.se/globalassets/rapporter/forsakringar-i-sverige/forsakringar-i-sverige-2010-2019.pdf).

[42] Classification of the Functions of Government, an international classification to divide the public sector into various parts according to function or purpose.

[43] Statskontoret, Statskontorets öppna Data: Number of Annual Workforce in Public Authorities 2019 (2019) (http://www.statskontoret.se/globalassets/psidata/arsarbetskrafter-staten-2019.xlsx).

[44] The five mentioned in Grant Thornton, Hur mår den privata vård- och omsorgsmarknaden i Sverige 2019?, Nov. 25, 2019 (https://grantthornton.mynewsdesk.com/documents/hur-maar-den-privata-vaard-och-omsorgsmarknaden-i-sverige-2019-91920).

[45] GitHub, Puppeteer (https://github.com/puppeteer/puppeteer).

[46] Webbkoll, dataskydd.net (https://webbkoll.dataskydd.net/en).

[47] Amelia Andersdotter & Anders Jensen-Urstad, Evaluating Websites and Their Adherence to Data Protection Principles: Tools and Experiences, in Privacy and Identity Management. Facing up to Next Steps: 11th IFIP International Summer School, Revised Selected Papers 39 (2016).

[48] Chromium is the browser on which Google Chrome, and currently Microsoft Edge, are based.

[49] Geolocalization can be carried out in several ways, but the visitor's IP address is sufficient to determine, with significant probability, at least the visitor's country and whether a consumer connection or a company (such as a cloud service) is in question, and to adapt (including blocking) content accordingly.

[50] A newer, lesser known, alternative to cookies is *Web Storage*, which can also be used for tracking purposes, see Web Storage (Second Edition), 3WC Recommendation (19 April 2016) (https://www.w3.org/TR/webstorage/).

an ordinary visitor, the visits for this study were made via an ordinary consumer broadband connection in Stockholm. Each website was visited five times because the third parties with whom a consumer is confronted may vary from visit to visit. Even if this method can map the presence of cookies, it is by no means self-evident what the purposes of these cookies are. That is, additional ways to determine these purposes, and to categorize them in relation to what role they play, are needed.

2. Categorization

Simply measuring the type of trackers present on websites does not indicate explicitly what the data will be used for, in which markets they are used, or the identity of the party behind the tracker. However, to understand how individuals' data is used and what type of data-driven market this relates to, these questions are important. Therefore, to identify third parties who are contacted by users' browsers, we compared the collected third-party addresses against a list of known tracking services. The list is based on the public tracker list compiled by Disconnect, an American company providing software for privacy protection.[51] Disconnect defines tracking as "the collection of data regarding a particular user's or device's activity across multiple websites or applications that aren't owned by the data collector, and the retention, use, or sharing of that data."[52]

The Disconnect list maps individual domain names to companies/organizations and types of service. The list is used by Mozilla Firefox and Microsoft Edge, among others, for their built-in tracking protection. For the purposes of this study, we used Mozilla's slightly adapted[53] version of Disconnect's list. In the original list by Disconnect, domains belonging to Facebook, Google and Twitter are in their own category, apart from everything else. In Mozilla's version, these domains have been moved into "Social," "Advertising," or "Analytics" as appropriate. A third-party domain in the list falls within one or several of the following categories:

- *Advertising*, a tracker which also displays ads or marketing offers;
- *Analytics*, a tracker which collects visitors' information and may build a profile based on a visitor's online activity that can be connected to a real name or other unique personal identifiers;
- *Content*, a mixed category that includes trackers; however, end-users often do not want to block, since it would impair the user experience (for example, embedded YouTube videos and Google Maps);
- *Cryptomining*, that is, letting computer systems perform the mathematical operations required to become the owner of bitcoin or another cryptocurrency. Here, a domain

[51] GitHub, Disconnectme/Disconnect-Tracking-Protection (https://github.com/disconnectme/disconnect-tracking-protection).

[52] Disconnect, Tracking Protection Lists (https://disconnect.me/trackerprotection).

[53] GitHub, Mozilla-Services/Shavar-Prod-Lists (https://github.com/mozilla-services/shavar-prod-lists) (the file disconnect-blacklist.json).

may be classified as cryptomining if it can cause the user's browser to mine cryptocurrencies without explicit user opt-in;[54]

- *Fingerprinting*, a more controversial way to identify particular users or devices based on the properties of the browser, device, network, or any other properties of the computing environment, *without* using client-side storage of cookies or other data; and
- *Social*, software that allows social networks to track users' activities outside the social network, such as by "like" buttons and embedded feeds.

Subsequently, the third-party domains contacted by more than one website, and that were *not* mapped against a third party in the Disconnect list, were examined manually. If it appeared—based on the exact address contacted, data sent, and what could be ascertained about the ownership of the domain—that this is a tracking service that probably belongs to a certain company and a certain tracking category, this was added to the list. Following a review of all such domains, all registered HTTP requests were checked against the new list.

It should be noted that the categories indicate different markets. This is relevant for purposes of understanding the role of individuals' data in a digital web context. It may also be relevant from a competition perspective in the sense that a dominant actor in one category can influence the market conditions of another category. The most intrusive categories that are hardest for consumers to avoid—such as fingerprinting—can also be particularly interesting to understand from both a privacy and competition perspective.

3. Limitations of the Third-Party Tracking Study

While having clear advantages in terms of the ability to get clear indications on the number and types of third parties somehow tied to a specific web page, the tracking methodology utilized here also has clear limitations. Firstly, it does not clearly indicate ownership structures or provide more in-depth knowledge about the organizational complexity of data brokerage. Another limitation relates to the fact that modern websites tend to be dynamic by their nature. A page can load a script that loads another script that loads yet another script, and so on. Numerous factors can influence what data is sent to and from a user's browser. These include user interaction (such as scrolling or cookie consent) and contextual and identifying information that the browser sends to first and third parties. Different subpages can cause different requests. For most websites, one visit is never exactly like another. As previously mentioned regarding trackers, a website visit can also involve more third parties than those that can be observed by the user.

In this study, each website is loaded in a browser, but no further interaction takes place (e.g., no clicking or scrolling). Only the front page of each website is visited. A recent trend is third parties disguising themselves as first parties to bypass blockers through a technique called *CNAME cloaking*.[55] This study does not analyze whether such technology

[54] However, this is not included in the results below, although it has been included in Mozilla's modified Disconnect-list.

[55] Ha Dao et al., Characterizing CNAME Cloaking-Based Tracking on the Web, in IEEE/IFIP TMA2020, 1 (2020).

is used. Lastly, the online tracking landscape is constantly evolving and the list of trackers from Disconnect cannot give a complete picture.

B. Interviewing Consumers

To provide more qualitative depth regarding consumers' awareness and perceptions of the role and value of individuals' information and behavioral patterns, four digital focus group interviews were conducted, in collaboration with the research company Novus. During the interviews, results from the web tracking study described above were presented to the respondents, to stimulate discussion and better understand how web tracking is perceived. The focus group participants were recruited from Novus's "Sweden panel." This is randomly recruited and web-based.[56] A number of questions about data collection and trust had been prepared in advance in order to screen and place participants in different groups depending on their stated level of trust. The target group of the survey is the general Swedish public, aged eighteen to sixty-five:

- High trust, encompassing two groups of individuals with a high level of trust; and
- Low trust, encompassing two groups of individuals with a non-high level of trust (indicated low level or neither high nor low level)

In this context, trust refers to the relationship with companies that collect data mediated through the internet, such as cookies, search habits, history, etc. One advantage of digital focus groups is the possibility of geographic spread, which is harder to achieve in a physical context. All participants participate under the premise that they must answer all questions. The focus groups were conducted online, in a chat in the software provided by Novus. Each group interview lasted around two hours, and forty-six individuals participated, divided into four groups. Working online in the form of a chat is somewhat similar to traditional focus groups with six to twelve persons in a physical room, but lacks the type of dynamics of interviews in physical spaces, where eye contact or other modes of delicate communication occur. It is a relatively quick process once it is active, and there is a significant group dynamic. In this case, ten to twelve individuals per group participated and we conducted two group sessions per evening. The moderator, instructed by a discussion guide prepared in advance, was tasked with following up on potentially interesting diversions, involving all group participants, stimulating thought, and encouraging further discussion.

III. Results

The following section first provides detailed results from the web tracking study, divided into the five sectors. It then compares these results with the results from the interviews of how users and consumers experience and understand this mode of web tracking.

[56] Novus, Novus Sweden Panel (https://novus.se/en/metoder/sverigepanel/).

A. Third-Party Trackers on the Swedish Web

On April 18, 2020, 115 websites were tested in total: thirty in the media category, twenty-five in retail, sixteen in banking/insurance, thirty-seven in the public sector and seven in healthcare/well-being. The five runs conducted generated a total of 54,505 successful HTTP requests[57] to 437 unique domains. In each sector, the overwhelming majority of all websites use tracking tools for analytical purposes. They do this to track website activity such as session duration, pages per session, where visitors came from (referrers), how long they stayed on the website and their geographical position, etc. Overall, Google is the only actor who dominates in all sectors, and 105 out of 115 websites (ninety-one percent) use one of Google's services. Google Analytics is used by most of all websites, regardless of category.

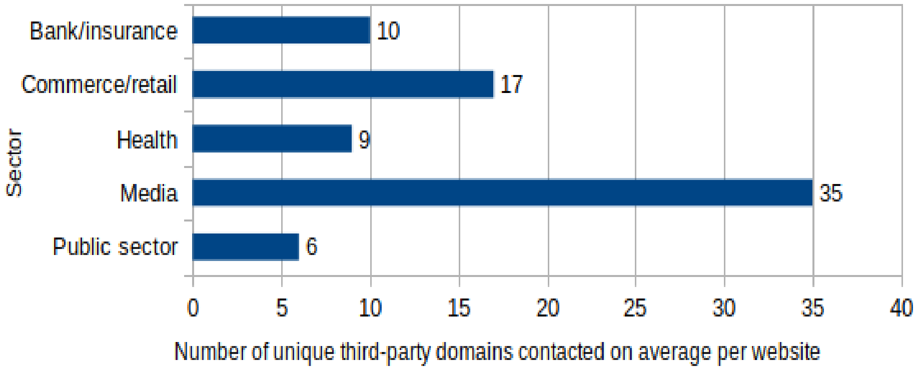

Figure 1: Unique third parties contacted on average on websites in each sector.

A comparison among the five sectors regarding how many unique third-party domains the websites in the respective sectors contact on average shows that the media websites stand out with more than twice as many third-party contacts compared with number two, retail (see Figure 1). Retail, in turn, contacts on average seventy percent more unique third parties than banking/insurance. The websites in the public sector make the smallest number of third-party contacts.

1. Media

By way of example, Figure 2 shows what the distribution can look like with respect to trackers and categories of trackers for the front-page of a regional Swedish newspaper. However, the picture is a simplification that does not show the connections among the various third parties; as mentioned previously, various intermediaries may be involved, in particular in relation to advertising services. A tracker can forward to another, who can forward to a third. During one single visit, the browser can contact a tracker many times. Visits on the front page of the newspaper in Figure 2 caused, on average, 200 third-party

[57] Hypertext Transfer Protocol (HTTP) is the foundational protocol of the web. The client (e.g., a user's web browser) sends *requests* to a server. The server sends *responses* back to the client.

requests: somewhat above the sector average of 158, and substantially below some others in the same sector.

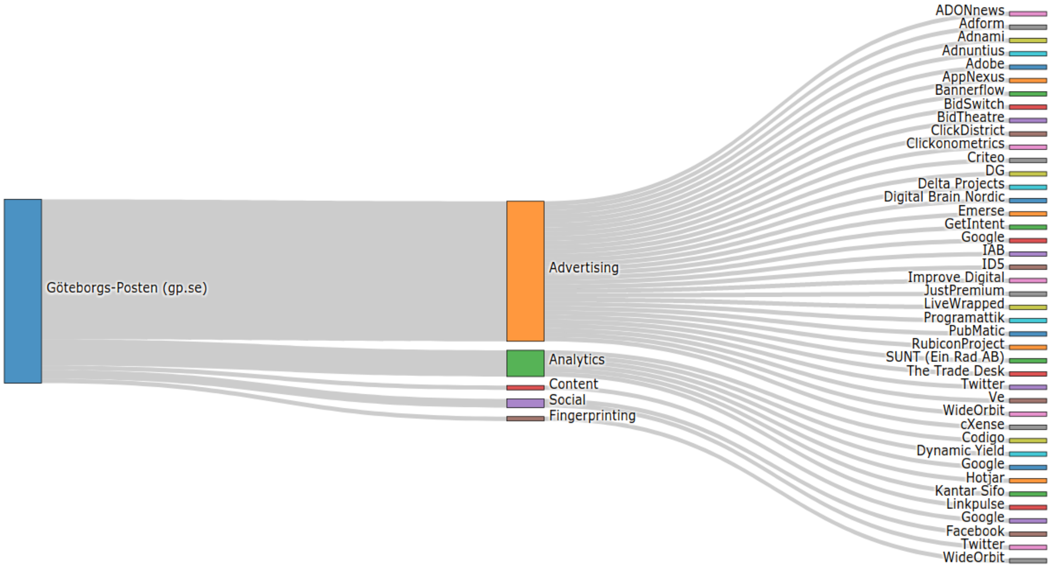

Figure 2: Third-party trackers contacted during a visit to the regional Swedish newspaper Göteborgs-Posten's (GP) front page.

Google enjoys the strongest position in relation to advertising third parties; see Table 1 and Figure 2. Their advertising tool is used by twenty-eight of thirty tested websites and by all newspaper groups. However, there are plenty of other companies with a strong presence among several newspaper groups; see Figure 3. In total, thirty-eight advertising tracker companies were found, twelve of which occur in at least one website in at least half of the groups.

Table 1: The ten most common pairs {company, tracker type} in media

Company	Tracker type	Number of websites	Proportion %
Codigo	Analytics	30	100
Kantar Sifo	Analytics	30	100
Google	Analytics	29	97
Google	Content	29	97
Google	Advertising	28	93
Adform	Advertising	26	87
AppNexus	Advertising	25	83
PubMatic	Advertising	25	83
RubiconProject	Advertising	25	83
LiveWrapped	Advertising	20	67

All of the media group Bonnier's tested websites use the same set of advertising third parties—Adform, Google, PubMatic and RubiconProject. These four companies are also found in all of MittMedia's, another media group, websites. We note that Bonnier News acquired eighty percent of MittMedia in 2019, which likely explains this pattern.[58] No advertising companies other than those listed in Table 1 were found in the websites of Bonnier, Gota Media, MittMedia and Schibsted. NTM and Stampen distinguish themselves with a larger number of contacted advertising parties. A total of twenty-three advertising companies were found on the NTM websites, and of these, seventeen were present on all NTM websites. In the Stampen group, thirty-seven advertising companies were found, of which twenty-two were present on all Stampen websites.

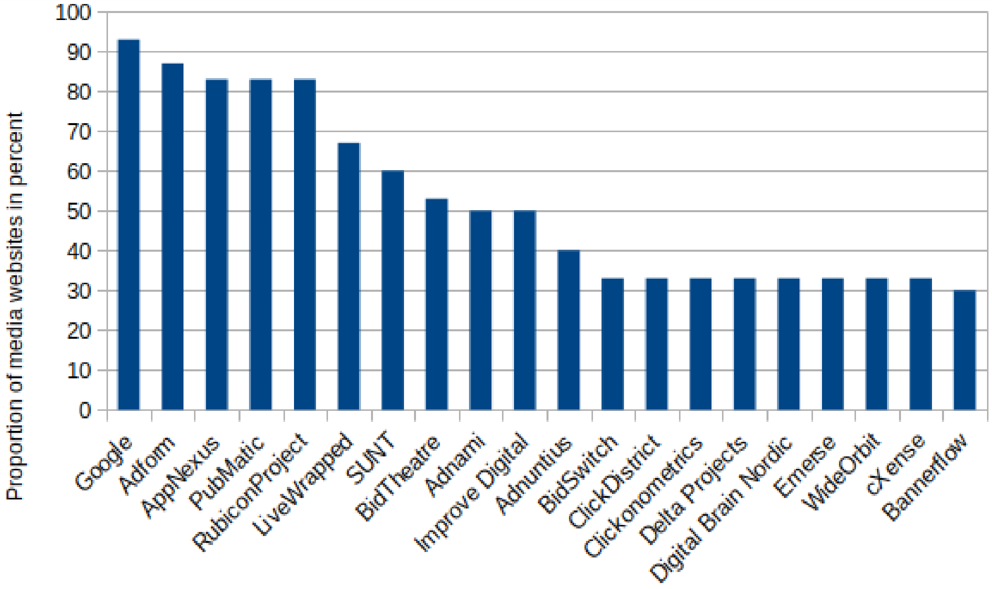

Figure 3: Advertising third-parties' presence on media websites in percent for the 20 most frequently occurring.

Media is the only sector where the tracker category *fingerprinting* is found. Two out of six newspaper groups use this type of tracker: NTM and Stampen. All of their tested websites use WideOrbit. For analytics, all of the tested websites use Kantar Sifo and Codigo; these two collaborate (script is collected from Codigo who then contacts Kantar Sifo); see Table 1. Twenty-nine out of thirty use Google Analytics. It can be noted that the number of advertising trackers is greater among local newspaper groups than among Bonnier and Schibsted, owners of the largest, national newspapers. Stampen stands out with thirty-seven different advertising trackers, of which twenty are found on all of their websites.

2. Retail

Most websites in retail use trackers in the categories of advertising, analytics, content and social (Table 2). Retail is the only sector where most websites—sixty-eight percent—use *social trackers*. This is nearly always Facebook, as found in seventeen out of

[58] Bonnier Nyheter, Bonnier News Acquire Mittmedia, Feb. 7, 2019 (https://www.bonnier.com/sv/news/bonnier-news-forvarvar-mittmedia).

twenty-five websites. The only other social tracker used is Twitter, which appears on only one website.

Table 2: The ten most common pairs {company, tracker type} in retail

Company	Tracker type	Number of websites	Proportion %
Google	Analytics	24	96
Google	Content	23	92
Google	Advertising	22	88
Facebook	Social	17	68
Adtraction	Advertising	11	44
Microsoft	Content	10	40
Hotjar	Analytics	8	32
New Relic	Analytics	7	28
Criteo	Advertising	6	24
Adform	Advertising	3	12

All of the seventeen websites that use Facebook use the analytics tool Facebook Pixel.[59] Briefly, this means that by allowing Facebook to track visitors, the website owner can access the tracking, have the opportunity to conduct targeted advertising campaigns and analyze their campaign's efficiency based on Facebook's analytics. For example, users can be divided into different "customized target groups" depending on their activity on the website—what pages they visit, what they search for, whether they put anything in the basket (and, if so, what), whether they complete the purchase, and so on. The website owner can then create an advertisement on Facebook which is shown to individuals whom Facebook identifies as *similar* to the individuals in a certain customized target group, based on, for example, demographic information and interests.[60] Google Analytics offers a similar functionality if the website owner chooses to activate the advertising functions. Target groups collected with Google Analytics can then be used in Google Ads.[61]

3. Banking and Insurance

In banking and insurance, thirteen out of sixteen websites use a third party categorized as advertising. Of these thirteen, all use either Google or Adobe, both of which are used by eight websites (Table 3). Google also dominates in the analytics category; Google Analytics is used by ten out of sixteen websites. Overall, however, the websites in banking and insurance use analytics services to a lesser extent than other sectors. Six companies use Facebook Pixel.

[59] Facebook Business Help Center, About Facebook Pixel (https://www.facebook.com/business/help/742478679120153?id=1205376682832142).

[60] Facebook Business Help Center, On Lookalike-Target Groups (https://sv-se.facebook.com/business/help/164749007013531?id=401668390442328).

[61] Google Analytics Help, On Advertising Functions (https://support.google.com/analytics/answer/3450482?hl=sv).

Table 3: The ten most common pairs {company, tracker type} in banking/insurance

Company	Tracker type	Number of websites	Proportion %
Google	Analytics	10	63
Google	Content	9	56
Adobe	Advertising	8	50
Google	Advertising	8	50
Facebook	Social	7	44
Adform	Advertising	4	25
Hotjar	Analytics	3	19
Adtraction	Advertising	2	13
Microsoft	Content	2	13
Adtriba	Analytics	1	6

4. Public Sector

Among the five categories of websites tested, the ones in the public sector have the lowest occurrence of trackers. Like other categories, most—eighty-one percent—use some form of analytics services. Google Analytics dominates here as well, with Siteimprove coming in a distant second (Table 4).

Table 4: The nine most common pairs {company, tracker type} in the public sector (only nine were found)

Company	Tracker type	Number of websites	Proportion %
Google	Analytics	27	73
Google	Content	16	43
Google	Advertising	8	22
Siteimprove	Analytics	8	22
Vizzit	Analytics	3	8
Adform	Advertising	2	5
Facebook	Social	2	5
Hotjar	Analytics	2	5
Ontame.io	Analytics	1	3

The fact that one-quarter of the websites—including the Swedish Armed Forces, the Swedish Prison and Probation Service and the Swedish Environmental Protection Agency—use *advertising* trackers may seem surprising. This might be because the site owner has enabled "remarketing and ad reporting features" in Google Analytics for extended tracking capabilities.[62] Facebook Pixel was found on one website (the University of Gothenburg).

5. Health

In the health sector, there is a difference between the five online medical companies and the two public authority websites (the Swedish Public Health Agency ("FHM") and 1177 Vårdguiden ("1177"), the national healthcare platform). FHM and 1177 use Google Analytics, but have no other trackers (Table 5). Three of the five online medical companies

[62] Google Analytics Help, On Advertising Functions (https://support.google.com/analytics/answer/2444872?hl=sv).

have advertising trackers; four have analytics trackers; three have social trackers; all have content trackers.

Table 5: The ten most common pairs {company, tracker type} in health

Company	Tracker type	Number of websites	Proportion %
Google	Analytics	5	71
Google	Content	5	71
Facebook	Social	3	43
Google	Advertising	3	43
Mixpanel	Analytics	2	29
Adform	Advertising	1	14
Amplitude	Analytics	1	14
Branch	Analytics	1	14
Hotjar	Analytics	1	14
LinkedIn	Social	1	14

Three of the online medical companies—Doktor24, Min Doktor and Doktor.se—use above-mentioned Facebook Pixel. The same trio also appears to use Google's advertising integration. In addition to all of them conducting analytics, Doktor.se states that they use third-party cookies to "collect information for advertising to ensure marketing is as relevant as possible," and that an advertising company may map the browsing habits of the visitor on various websites.[63] Min Doktor writes that "directed cookies" are used to map visits for purposes of making the website and advertising "more relevant" and that this information may be shared with third parties.[64] Capio writes that they share information on visits with their "social media, advertising and analytics partners."[65] Doktor24 states that they process personal data, among others, for purposes of "marketing, customer satisfaction surveys or other market communications."[66]

B. Digital Focus Groups

In the opening discussion during the focus-group interviews, there were substantial differences between those with a high and those with a low level of trust in various actors who collect data. Those with a low level of trust did not approve of the rather ubiquitous collection of their data online, since they have no control over what happens with their data. Those in the group with a high level of trust expressed that they have become used to this practice, which ultimately means they must accept it. This group also trusts that the actors will not abuse their information.

[63] Doktor.se, On Cookies (https://web.archive.org/web/20200504101703/https://doktor.se/cookies/).

[64] Mindokter, On Cookies (https://web.archive.org/web/20200504101925/http://www.mindoktor.se/cookies/).

[65] Capio, On Cookies (https://capio.se/cookies/).

[66] Doktor24, On Cookies (https://web.archive.org/web/20200504102128/https://doktor24.se/faq/villkor-for-tjansten/personuppgiftspolicy/).

It is clear that the focus group participants do not generally read cookie information on websites, or make informed decisions regarding cookies. Some expressed that they mainly "click away" the information. Some individuals in these groups state that they are aware of what cookies entail. Some claim to accept cookies even though they do not want to, indicating a kind of digital resignation, as mentioned above in relation to Draper and Turow's studies on American consumers.[67] If they do not approve the agreements, they cannot access all of the website or its information, and they therefore feel they have no choice. Participants expressed that they have a relatively clear idea of why various actors wish to collect information about them—they want to target offers to them based on their purchasing habits, search history, etc. Thus, it appears that they understand that the primary reason for collecting information is commercial. However, they differ as to whether or not this is acceptable. Those with a high level of trust in various actors indicated that they are happy to receive targeted offers, even though this group can sometimes find advertising irritating. On the other hand, the group with a low level of trust is considerably more skeptical and suspects additional, hidden reasons. They also find it irritating, uncertain of where the information ends up and whether it is sold. Ultimately, they feel themselves tracked or monitored.

The individuals in the group with a high level of trust thus are more inclined to perceive that they can benefit from sharing their data, while the individuals in the other group feel that they are made to share more value than what they might get in return. It is simply not worth it. In some cases, attempts are made to side-step or resist the perceived profiling, by not visiting a specific website, letting an acquaintance blip their club card, not buying something online, or even buying something that one normally does not buy to "trick" the collecting side.

After the initial discussion on different types of data collection, the focus group participants were instructed to read through a short text on web tracking, including specific results from the third-party tracking study outlined in this article. The initial reactions were mainly surprise at the extent of third-party monitoring. Participants expressed that they had not been aware that the overall information collection was so extensive when visiting, for example, a media site. Several participants were not surprised that some type of data collection by a third party takes place, but most were surprised at the extent to which it occurs. Some felt uncertain, frustrated and angry, mainly in the group with a low level of trust, while others were not bothered and felt that nothing could be done about it; they were not surprised and perhaps even indifferent. They were also not concerned about it to the same extent.

In brief, the digital focus groups show that:

- With regard to third-party tracking, the respondents feel that it is impossible to read through cookie consent agreements or other information concerning data collection

[67] Draper & Turow, supra note 33; Turow & Draper, supra note 33.

and data management on websites. There are too many agreements, the language is complicated, and the layout tends to be bad or inaccessible;

- The informants are not particularly surprised by the web tracking, but they are surprised by its extent, especially regarding the involvement of third parties;
- Remarkably, neither the group with low trust nor the group with high trust express that they inform themselves about the implications of their choices, that is, read user agreements or cookie notices to the extent that they feel informed. Neither group can therefore be said to be particularly aware of how their data is collected and used in a web context, nor do they think that they have a reasonable possibility to become truly informed.
- The main difference between the two groups seems to be that one group is anxious about not being in control and desires to be more informed, expressing something like resignation, and the other feels more confident that their information is not being misused and even may be of possible benefit, and consequently does not feel anxious about it.

IV. Discussion

The data-collection practices in third-party tracking indicate highly complex, proprietary uses of data which most users and consumers struggle to be aware of and literate about.[68] This study shows that it is *practically impossible* for the typical web user, who has not actively taken action, to avoid being subject to tracking when visiting popular Swedish websites. Google's presence in the Swedish web is nearly universal, mainly due to Google Analytics, but there is also a "long tail" of actors; among the 115 websites in this study, trackers from 90 different companies were found, and 322 third-party domains were contacted. The main results show that third-party tracking is fundamentally present online, particularly for media and retail sites. This indicates that these data-driven markets largely depend on the continuous collection, sharing and trade of consumers' personal data.[69] The public-sector sites had the fewest third-party trackers.

Furthermore, even though regulatory requirements implemented in the EU in recent years, particularly through the GDPR, require more explicit notifications and designs to create more active choices by individual website visitors, a majority of users are still highly unaware of these data collection practices and their underlying purposes or possible consequences. The consent base for data collection seems flawed in this context—widespread consent fatigue[70] and possibly the increased use of dark patterns to nudge users to consent has in recent studies been shown to undermine principles of EU privacy law[71]—indicating

[68] Cf. Christl, supra note 1; Pasquale, supra note 26.

[69] Mellet and Beauvisage, supra note 2; van Dijck et al., supra note 19; Zuboff, supra note 19.

[70] Utz et al., supra note 7.

[71] Graßl et al., supra note 7.

that the market structures and data collecting practices are a black box in itself.[72] At worst, this leads to a risk that users and consumers become exploited—in the sense that they are unaware of what parties are collecting and trading their data,[73] and the purposes for which it is used—based on being manipulated into consent that is more mechanically click-based than informed.[74] Among some groups, it is meaningful to speak of a digital resignation, as suggested by studies on American consumer attitudes towards commercial data collection.[75]

Previous studies have documented that users feel automated individualization is both useful and worrying.[76] This can be compared to the group with a low level of trust in data sharing in the focus group interviews. This group clearly perceives the surveillance dimension of targeted services. Some even feel irritated or upset if they receive targeted offers or recommendations. Some groups feel resigned to the data collection that takes place according to decisions made above their heads and it sometimes causes them to decide not to visit certain websites or search for information, in order to avoid having their personal data collected. They also feel that it is impossible to familiarize themselves with the agreements that are intended to inform them about how their data is used, which creates negative feelings and frustration over a non-functioning approach to consent, even if the direction of blame is not very clear. In economic studies on privacy issues in a digital context, it has also been noted that consumers' ability to make informed decisions on their privacy is seriously hampered, since consumers often have incomplete or asymmetric information about when their data was collected, for what purposes, and with what consequences.[77]

V. Conclusion

This study concludes that the data-collection practices in third-party tracking indicate highly complex, proprietary uses of data which most users and consumers struggle to be aware of and literate about. By mapping third-party trackers on a sample of Swedish websites, we can conclude that third-party tracking is fundamentally present online, particularly for media and retail sites. Google is dominant in all five studied sectors, found on ninety-one

[72] Christl, supra note 1; Pasquale, supra note 26; Larsson, supra note 22; Larsson & Heintz, supra note 22.

[73] Christl, supra note 1; Pasquale, supra note 26. On markets for personal data and the aggregation by data brokers, beyond ad markets, including credit assessments and insurance, see Hildén, supra note 2; Utz et al., supra note 7.

[74] Compare with the EDPB's guidelines on consent under the GDPR, supra note 10. For a critical study on "non-informed" and "blind" consent practices, see Bechmann, supra note 35. For risks of manipulative nudging beyond user interfaces and based in the analytical possibilities of (big) data as such, that is of a "continuously updated, dynamic and pervasive nature," see Karen Yeung, "Hypernudge": Big Data as a Mode of Regulation by Design, 20 Info., Comm. & Soc'y 118 (2017).

[75] Draper & Turow, supra note 33.

[76] Blase Ur et al., Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising, in Proceedings of the Eighth Symposium on Usable Privacy and Security 1 (Lorrie Cranor ed., 2012).

[77] Alessandro Acquisti et al., The Economics of Privacy, 54 J. Econ. Literature 442 (2016).

percent of the sampled websites, and Facebook is particularly prevalent in the retail sector, and also found to be present on commercial online doctors' websites. All in all, this clearly indicates a highly commercialized contemporary data-collecting web, unmistakably interconnected with digital advertising markets. These results were presented in digital focus groups, where many expressed surprise by the extent of third-party tracking, but were differently concerned about its consequences. Neither group expressed that they inform themselves about the implications of their choices, in the sense of reading user agreements or cookie notices, nor that they experience having a realistic possibility to do so.

The study thus shows that it is practically impossible for the typical web user, who has not actively taken action, to avoid being subject to tracking when visiting popular Swedish websites. Therefore, we conclude that there is a need not only to rely on actions for stimulating data literacy amongst users and consumers but also for responsible supervisory authorities to be more active in mapping contemporary markets for data, as well as clarifying what consent practices mean for uninformed and possibly exploited users.[78] For example, the results of this study identify challenges of relevance for data protection and privacy issues, as well as a broader framework of consumer protection. In addition, consent fatigue and lack of awareness of how data is collected and used for commercial purposes are also of direct relevance for competition regulation, bearing in mind the risk of undermined competition. Consumers may be manipulated, unable to choose fairer services or exercise data portability rights. Hence, the inherent imbalance between the complexity of third-party tracking markets and the lack of informed and aware consumers, which ultimately means a highly non-transparent commodification of personal data, are of relevance for data protection, consumers, and competition authorities to investigate further. This is particularly true for media and retail markets. Therefore, we recommend more active supervisory authorities, and more stringent requirements, primarily concerning obscure ad tech infrastructures to improve transparency and promote consumer awareness.

[78] Larsson, supra note 30.