Algorithmic marketing and EU law on unfair commercial practices. Rethinking consumer protection with AI

Algorithmic marketing and EU law on unfair commercial practices

Federico Galli

Springer

1 Data-driven surveillance

1.1 The new marketing research

Over the last twenty years, the Internet, e-commerce platforms, social networks, and mobile devices have turned the average Internet user into a constant data generator.[1] Today, data trails left by users cover an increasing array of information about their private lives: from simple socio-demographic information and purchase choices to networks of social connections, information on life events, and real-time location. Marketers have an incentive to collect this information to better understand how individuals behave, improve their communication with the public, and efficiently position their brand in the market.

To this end, marketing organisations implement organisational and technological processes to translate data into actionable knowledge.[2] This set of processes is known as marketing analytics or customer analytics.[3]

As we saw in the previous chapter, businesses have spent the past forty years refining their IT systems to provide actionable knowledge for their marketing decisions. After a first era of "artisanal analytics," primarily based on descriptive statistical methods applied to internal static databases (not connected via the Web), the rise of the Internet data infrastructure has progressively prompted business organisations to incorporate techniques and algorithms developed in data mining and machine learning.[4]

Consumer analytics based on big data and machine learning differ from traditional tools based on statistics. With new methods, the data analyst does not need to form hypotheses about the relationship between the variables in the model, but instead relies on inductive learning, with the system itself building a model that represents which variables in the

[1] On big data and related transformations in social and economic relationships, see McAfee and Brynjolfsson (2012); Mayer-Schönberger and Cukier (2013); Davenport (2014); Morrison (2015).

[2] McAfee and Brynjolfsson (2012) describe five new management challenges for a company to succeed in big data management. Companies need to have leadership teams that set clear goals and ask the right questions, and they need to employ IT workers with expertise in computer science and data management. Companies need to develop excellent software tools to manage the volume, velocity, and variety of big data, and need to use the knowledge so extracted to maximise cross-functional cooperation between business departments and customer relationships. Also needed, crucially, is a cultural shift in mentality: the right question for managers is no longer "what do we think?" but "what do we know?" This change requires moving away from acting solely on instinct and making use of data to build strategies and decision-making.

[3] In particular, large quantities of data require "big data consumer analytics", defined as "the extraction of hidden insight about consumer behavior from Big Data and the exploitation of that insight through advantageous interpretation" (Erevelles, Fukawa, and Swayne 2016).

[4] According to Davenport (2018), companies today are transitioning from analytics 3.0—which he describes as data-economy analytics, in which companies in traditional industries also embrace big data and analytics by resorting to machine learning models—to analytics 4.0, where we have the highest degree of sophistication in analytics thanks to AI and cognitive technologies (NLP, speech recognition, computer vision). Analytics 4.0 enhance the previous stages and form a backdrop that merges analytics and automation. On the stages of this transition see, more generally, Davenport and Harris (2017).

data should be retained as meaningful and which should be dropped. Current analytics tools can accommodate a virtually infinite number of variables to provide a generalised knowledge model for decision-making in customer relationships. Consumer-data variables may include not simply structured data but the whole variety of textual, visual, and aural data, which do not have a predefined structure.[5] To deal with this new variety of data, business-information systems are progressively incorporating algorithms and techniques developed in subfields of AI, such as natural language processing and image recognition. Once a machine learning model is in place, the AI system can automatically be retrained—or possibly self-adapt—based on the new data that is generated and can respond to modifications of data input in the dataset. This adaptability makes machine learning the perfect way to cope with the speed at which new consumer data is produced today, and at which existing data therefore becomes obsolete.

Big data analytics is said to be driving a new information and research strategy in companies.[6] Such an understanding means that big data analytics should not merely be regarded as a set of new technical tools enabling companies to gain more refined customer information. This is also a new organisational setting through which information becomes available to the enterprise and decisions are made. It is important to better understand this transition. This new approach is reflected in the so-called data-driven orientation, where the word *driven* does not only mean that marketing decisions are made looking at the data but indicates the way strategy is formed and decisions are subsequently made.[7]

Therefore, the first step in investigating developments in the relationship between algorithmic business and consumers consists in paying attention to the practices underlying the new data-driven strategy.

1.1.1 Data gathering and exchanges

Today, marketing organisations can gather data about consumers from a variety of sources.[8]

---

[5] Balducci and Marinova (2018) define unstructured data as "information that either does not have a pre-defined data model or is not organised in a pre-defined manner." This kind of data can be either verbal (spoken and written data) or nonverbal, which in turn can be human (e.g., facial cues) and inanimate (e.g., radio frequency identification, location data).

[6] On the impact of big data on strategy formation, see Constantiou and Kallinikos (2015, p. 52), who look at big data as a new organisational context. While traditional business strategies would respond to the need to collect data intentionally and purposefully to inform specific theoretical models and provide predefined input to decision-making process, big data can be made business-relevant subsequent to its collection, but its relevance may not straightforwardly derive from the original data records. Much of what comes under the heading of big data is not collected intentionally; rather, they say, "it is haphazard, hugely heterogeneous, and, not infrequently, trivial, messy and agnostic, as it happens with much user-generated content and data logs of various kinds".

[7] Beyond business, this new setting has been summarised in the motto "First data then search for any possible uses of what is already available as data." The quote is from Anderson (2008), which elaborates on the obsolescence of traditional scientific methods in research.

[8] In providing a framework for different data-collection methods, I am following Clarke (2019).

Some data is acquired by businesses through direct digital transactions with consumers. In this case, individuals may be conscious of a transaction taking place, or, at least, of the fact that data have been acquired as part of the transaction. Sometimes consumers themselves are requested to disclose personal information (e.g., email, credit card, billing address) when registering for a service, purchasing items, or signing into social media. Otherwise, they may volunteer personal information, as when posting a review on a website or updating their social media profile.[9] Consumers are generally assumed to read the privacy notices on data-collection practices and therefore to be conscious of which data items are acquired and for what purposes, for which they give express or implied consent. As is often noted, however, consumers are, at best, very confused about the information given by data collectors about the acquisition and processing of data.[10] Even when they expressly consent to data collection, doubts remain as to whether they have a genuine understanding of how much data they are providing to the collecting entity.[11] As will be discussed later in this chapter, this issue is closely linked to the mechanisms put in place by policies, laws, and regulations to protect consumer privacy.

A crucial data asset of marketers is acquired from consumers through various activities the consumer cannot appreciate even when disclosed in privacy statements. Such data are typically a by-product of the transaction mediated by computer protocols, acquiring data such as consumers' IP address and the type of software used, and the data may even include their physical location.[12] Consumers seldom appreciate how much information they are giving up simply by using their computers. This ranges from macro-measures, such as the site visited, the number of pages or screen viewed, the time spent on-site or in-app, the average time spent per page, the time elapsed since the last visit; to deep behavioural data, such as time spent in page-specific contexts, mouse movements, scrolling path and speed, mouse hovering on specific content, and time of inactivity.

Similarly, covert modes of data collection may be engaged not only when the consumers are navigating the website of the collecting entity itself but also when they are interacting with other websites. This purpose may be served by surreptitious tracking technologies such as third-party cookies, web bugs, web beacons, tracking pixels and browser fingerprints, and adware. These technologies cause an instant transfer of customer data to a party other than the one who owns the website.[13] Surreptitiously acquired data can be then exchanged via each marketer's networks of partners and/or with data brokers and advertising-service providers, resulting in the data being further exploited by many more

[9] On "volunteered information," see Mitchell (2010). Before Web-based commerce, marketing had only had two sets of information to deal with: market-aggregated information, which is primarily qualitative and statistical, and transaction-related data, which are granular and personally identifiable.

[10] According to Yong Jin Park (2013), privacy literacy online consists of the following elements: familiarity with technical aspects of the Internet, an awareness of common institutional practices, and an understanding of current privacy policies. Many empirical studies exist already on the problem of online consumers' privacy literacy. See, among others, Bartsch and Dienlin (2016) and more recently Prince et al. (2021).

[11] Ridley-Siegert (2015); Walker (2016).

[12] This kind of data acquisition is traditionally referred to as "clickstream data." See Montgomery et al. (2004).

[13] Sipior, Ward, and Mendoza (2011).

organisations.[14] The key players in this space are advertising-service providers such as DoubleClick (owned by Google) and Quantcast. Consumer tools such as Ghostery and AdBlock Plus, and their associated lists, can be used to identify many hundreds of organisations that are privy to individuals' Web traffic through such means,[15] even though the reliability of these tools has been questioned.[16]

A great deal of data is acquired from sources asserted to be in the public domain and free for use for virtually any purpose. The most relevant source is social media, which includes public profile data, shared postings, and other multimedia content. User photos, texts, and video data are examples of unstructured consumer data collected by the average company with a profile and followers or a connection with its customers. Other public environments, such as blogs and content-sharing platforms, can be accessed to gather data about consumers. Social media platforms are also relevant for data-collection services provided to business partners. The level of these services is generally dependent on how much data the platform already knows and collects and on company policy.[17] For example, Pandora, Google, and Facebook offer sophisticated tools for gathering consumers data, while Amazon's tools are limited to keyword search.

Marketing organisations can acquire consumer data from other organisations by adopting collaborative or pay-per-data schemes. This may be by purchase, barter, or other form of sharing, commonly designated in privacy policies using vague terms such as "strategic partnership" or "business-affiliated entities."[18] The general public is rarely aware of such data transfers. Many of these exchanges are conducted privately because of their dubious legality and the risk of media coverage and public outrage.

Data-privacy reports increasingly point to the systematic overcollection and oversharing of consumer data among service providers. Such practices are barely, if at all, compliant with data-protection regulations on data exchanges, and happen outside the public consciousness. Hundreds of thousands of different companies participate in networks of transcontinental collection, sharing, and selling of information. For example, in a 2020 report, the Norwegian Consumer Council reviewed ten privacy policies of common app providers and found that consumer data are transmitted to at least 135 different third parties involved in advertising and marketing services.[19] While some of these data transmissions prove to be necessary for the apps to function, personal data are in many instances sent to companies that use this information for purposes that consumers cannot know in advance.

---

[14] Christl and Spiekermann (2016), p. 40: "The marketing data industry is arguably the main driving force behind ubiquitous consumer surveillance. It consists of a wide range of different types of companies, including marketing and advertising agencies, list brokers, database management providers, online and mobile advertisers, and firms engaged in direct mail, telephone sales services, and data-driven commerce, as well as companies offering loyalty programs. Marketing data companies, which are often called "data brokers", mostly offer a smaller or bigger selection of these services."

[15] Schaub et al. (2016).

[16] Privacy Choice (2010).

[17] For an account of methods of users' data collection and disclosure, see Lomborg and Bechmann (2014).

[18] H. Schneider et al. (2017).

[19] Forbrukerrådet (2020).

Marketers, who might also have a brick-and-mortar retail shops, can also merge online consumer data with data about what consumers do in the store.[20] The increased displacement of closed-circuit televisions (CCTV) or radio-frequency identification (RFID) and tracking devices (e.g., digital points of sale, in-store table surveys, touchscreen kiosks) in stores allow traders (vendors) to collect data about the way shoppers behave and interact with various products as they move through the store. Thanks to image-processing-and-recognition technologies, multimedia data can be converted into actionable insights about behaviours.

The many streams of online consumer data add to the plethora of data-acquisition methods antedating the emergence of e-commerce. These methods include mailing lists, loyalty programs, telephone directories, and electoral rolls. Use may also be made of publicly available data, such as voter records, name and address changes, and marital status.

1.1.2 Profiling and audience creation

Companies collect different types of data in anonymised or pseudo-anonymous form through identifiers (e.g., IP address, cookies, device IDs) that usually do not contain information about natural persons. For example, a consumer's repeated interactions with an e-commerce website can be collected and compiled into a digital consumer record, which may contain data on personal attributes provided in setting up or updating a personal account, as well as data on items purchased and clickstream data. A commonly used term in the industry is "customer profile," referring to the consolidated record of first-, second-, or third-party data about a particular consumer.

Profiling also has another meaning, referring to the different analytical interventions performed on collected data.[21] Except for declared data, which may directly represent certain personal information about the customer (e.g., age, sex, marital status, specific consumption preferences), data stored in customer profiles are not self-representative and must be analysed to acquire actionable knowledge. As noted, in big-data settings, such analytical activities occur through automated means based on statistical models, including machine learning.

From a machine learning perspective, automated profiling is "the activity of building statistical models from large amounts of data from many individuals, after which the profiles themselves can be exploited to derive novel information about particular individuals."[22] The process is twofold. First, using a learning algorithm, a model is built from the dataset containing aggregate knowledge found in the data. Here, businesses can

[20] Brynjolfsson, Hu, and Rahman (2013). Omni-channel is a strategy based on merging consumers' physical and online presence to provide a seamless customer journey across e-commerce websites and brick-and-mortar stores.

[21] The term *profile* derives from the Italian *profilo*, from *profilare*, originally "to draw a line," especially the contour of an object. That is precisely the idea behind profiling through data processing, in which the data that is available about individuals or groups is expanded so as to describe their traits and propensities.

[22] Van Otterlo (2013).

use the totality of customers' data to infer a general distribution of patterns. Thus, the model could include a correlation between two attributes declared by consumers, e.g., the fact that eighteen-year-old girls living in a certain suburb of Milan like punk music. Second, the resulting model will be applied to a particular consumer or group of consumers to uncover new knowledge about them that was not previously present or observable in the data. Suppose, for example, that consumer Anna has explicitly declared in her profile that she is an eighteen-year-old woman living in the suburbs of Milan, but she hasn't mentioned any interests in music. A company can apply the statistical model to infer that Anna may like punk music. In this case, consumer Anna has been profiled as possessing a "taste for punk music."

Using profiling techniques, marketers can carry out different kinds of analysis. One of the most important is *audience segmentation*, premised on the understanding that customers vary substantially in their behaviour, needs, wants, and characteristics. The main goal is thus to divide the existing or potential customers into groups of similar consumers to select the segment to be considered for a similar marketing initiative. As we saw in the previous chapter, this form of profiling can be performed through unsupervised learning methods, such as cluster analysis.

Clustering allows marketers to iteratively divide available consumers into a certain number of non-predefined segments without having to make assumptions about similar categories. Clustering provides a significant advantage, as it broadens the spectrum of audience segments from classical socio-demographic groups (sorted, for example, by age or sex) to people sharing similar behaviours. For example, instead of dividing consumers according to the category "age 18–25," the data clustering would instead consider as relevant audience a group fitting the description "(*i*) mother, (*ii*) born in August, (*iii*) has two children, and (*iv*) listens to Iron Maiden." Such an analysis would enable marketers to observe that in its potential customer base, mothers who have two children and are born in August are more likely than others to buy Iron Maiden music—a correlation that would be unexpected. Moreover, clustering algorithms are designed to learn directly from data distribution and its changes. This adaptability means that as long as new data about consumers are available and the profiling model is retrained on those data, the distribution of clusters and partition of customers may change over time, ensuring that resulting segments always reflect current behaviours.

Cluster analysis can be performed by any business equipped with enough data about consumers and analytical software.[23] Customer-clustering tools are generally included in many AI-powered data analytics software products and are one of the basic analytics processes that can be performed in analytics API served by Google and the like.[24] Alternatively, customer segments can be selected through marketing API services available in social media environments. Providers such as Facebook and Google use their in-house AI analytics tools to segment their users into various groups based on different categories. The possibility of reaching such groups is offered to marketers advertising on their platforms as "customised audiences." These are targeted advertising services that

[23] Diaz Ruiz and Kjellberg (2020).

[24] Google Ad Manager Help, Create first-party audience segments, 2020, *https://support.google.com/admanager/answer/2423498?hl=en*. See, more generally, Procter, Voss, and Lvov (2015)

allow marketers to find new customers among platform users on the basis of the characteristics the same marketers want to find in the audience. For example, *Tables 3.1* and *3.2* provide an overview of the different attributes that Facebook appears to infer from user data.[25]

Style	**Category**	**Groups (examples)**
Demographics	Age	A range of 13–65+ can be customised
	Education	High school, college, doctorate degree, professional degree
	Ethnic affinity	Hispanic, Afro-American, Asian-American
	Financial	A range between 30K and 500K can be selected
	Gender	All, men, women
	Generation	Baby boomers, Generation X, millennial
	Home	Apartment, single-family home, renter, young and hip
	Language	Select one language
	Life events	Anniversary within 30 days, away from hometown, new relationships, recently moved, newly engaged
	Location	Marketers can select one or more countries, regions, cities, and ZIP/postal codes, and then select "everyone in the location," "people recently in the location," and "people travelling in the location"
	Parents	New parents, parents with teenagers, green mums, stay-at-home mums etc.
	Politics	Marketers can select the political orientation of users (e.g., liberal, very liberal, conservative, very conservative)
	Relationship	Interest (men, women, men and women) relation (e.g., engaged, married, widowed, divorced)
	Work	Marketers can type in a job title and select an industry or office type

Table 3.1. Facebook's customised audience (demographics)

Style	Category	Groups (examples)
Interest	Business and Industry	Healthcare, design, real estate, retail, aviation, etc.
	Entertainments	Games, movies, music, reading, TV
	Family and Relationships	Fatherhood, dating, wedding, parenting
	Fitness and Wellness	Dieting, physical fitness, yoga, rumba
	Food and Drinks	Alcoholic beverages, type of cousin, restaurants

[25] Facebook For Business, Audience Targeting Options, 2020 *https://www.facebook.com/business/help/633474486707199*.

	Hobbies and Activities	Politics and social issues, garden, home, arts and music
	Shopping and Fashion	Beauty, clothing, toys
	Sport and Outdoors	Outdoor recreation, sports (types)
	Technology	Computers (types), consumer electronics
Behaviours	Automotive	New vehicle buyers, used vehicles, types of vehicles
	Business-to-business	New vehicle buyers, used vehicles, types of vehicles
	Charitable donations	Cancer causes, political, health, religious, animal welfare
	Digital activities	Console gamers, photo uploaders, operating system used, Internet browsers used
	Expats	Select country
	Financial	Spending methods (lines of credit)
	Job Role	Professional, corporate executive
	Media	Radio, TV
	Mobile device user	Mobile by brand, mobile by OS, network connection, smartphone and tablet owners
	Purchase behaviours	Marketers can select categories of products or purchasing habits (e.g., clothing, kids' products, store types, technology)
	Residential	Like to move, recent home buyers, recently took out a home loan
	Seasonal and events	Football match, rugby, etc.
	Travel	Frequent traveller, personal traveller, commuter, family vacations, returned from a trip
Connections	Page visitors	People who like your page, friend of people who like your page
	App users	People who used your app, friends of people who used your app
	Events	People going to your event, friend of people going to your event

Table 3.2. Facebook's customised audience (interests, behaviours, connections)

Marketers can use profiling not only to segment audiences but also to better understand the specific characteristics of new customers. Suppose, for example, that a shopping site had 50 visits in the last week and only five customers have signed up for a service. These five customers represent their best (i.e., most profitable) consumer segment. Therefore, a company might want to find out if among thousands of potential customers, there are customers similar to their best customers, so as to expand the reach of a successful campaign. Acquiring additional information about customers in marketing is often done through the development of so-called "buyer persona."

The notion of buyer personas was developed in 1993 by marketing professor Angus Jenkinson and responded to the mission of developing fictional characters showing typical and recurrent traits of a brand's target buyers.[26] Today, data-driven companies can

[26] Jenkinson (1994). In the early days of marketing, buyer personas were generally constructed by conducting interviews and surveys with real consumers. Using their expressions and thoughts, and maybe even the photos they took or the activities they did, marketers would "step into the shoes" such real

build their buyer personas by directly looking at the data and performing analytics to find similar customers, that is, ones that share traits.[27]

Developing a data-driven persona is another case of automated profiling, making it possible to infer additional characteristics about a specific consumer based on past data about similar individuals. Unlike what is the case in segmentation, here the marketer does not attempt to group consumers according to similarities ("look-alike modelling")[28] but rather tries to determine "who the consumer is" on the basis of a presumptive description that can be inferred by considering and confronting a consumer's past behaviour with the behaviours of other customers. For example, imagine an online travel agency that offers its clients tourist packages for sports activities or romantic getaways. A romantic getaway is more suited to a young couple looking for intimate moments, while a sports weekend is for people who enjoy outdoor activities and love adrenaline. In this case, the marketer, who has no clue about the identity of their new website visitor, could use data-driven persona research to guess her marital status. This can be done based on data about previous consumer data and their marital status, i.e., single, engaged, married (in online travel agencies such as Booking.com or Expedia, this kind of information is typically part of the information.

1.1.3 Scores and other predictions

Instead of assessing the individual's characteristics, "customer scoring" practices assess consumers' behaviours in terms of their probability of bringing a particular benefit or loss to the firm.[29] Not all customers are alike in their response to a certain marketing campaign, so the marketer may opt to predict which customer profiles are most likely to engage positively with the offers.

Customer scoring practices can employ algorithmic systems powered with machine learning and predictive analytics, using past data to make predictions. For example, a travel company that wants to understand whether its customers might respond positively to a certain package offer for mountain holidays can use machine learning on a dataset containing attributes and behaviours describing all its consumers. The system can map a function that correlates the consumers' characteristics with a score determining the probability of being responsive to the offer in question.[30] After enough training, and each

individuals and rely on the imagination and creativity to build profiles that described their hypothetical target consumer.

[27] Burgess and Burgess (2020) explain that a buyer persona "is a semifictional representation of your ideal customer based on market research and real data about your existing customers." The authors describe this process of creating a buyer persona through data in seven steps: gathering buyers' data, assembling the team, getting to know the buyer, evaluating pain points, crafting the persona, continuously revisiting and revising the persona, and validating the persona. Based on the validated persona, the marketing team will be able to find a lookalike audience in new incoming data.

[28] Popov and Iakovleva (2018).

[29] On predictive analytics see, generally, Siegel (2013).

[30] This is an example of regression. However, consumer scoring can also be modelled as a classification task. In such a case, consumers would be classed as either "potential buyer" or "non-buyer."

time new customer's data flows into the company database, the tool discovers and suggests customers who are more likely to respond. The company can target its commercial offers at them rather than at unprofitable consumers and can thus maximise returns on its marketing campaigns.

Customer scoring is typically used in direct marketing to generate "leads" (so-called lead scoring), as with the example above. The model inductively derived from the data contains a list of leads, i.e., consumers who will be potentially responsive to the marketing trigger on the basis of specific attributes or past purchases.[31] Other times, customer-scoring practices can be used to predict the long-term profitability of each consumer. This indicator is known as customer lifetime value (CLV) and measures (predicts) the profits that can be extracted from a company's relationship with a particular customer considering the customer's attributes and purchasing behaviour.[32] Finally, another common type of score is "customer churns." In marketing language, churns are consumers who show signs of switching between sellers. They need to be identified in advance and lured with specific offers before they switch to another seller.

Consumers-scoring and predictive-analytics practices are becoming popular among all market players across industries.[33] Given the instant access to analytics software and the democratisation of data-science and machine-learning knowledge and tools, making predictions to optimise marketing actions is likely to prove to be the cornerstone characteristic of data-driven marketing.

Undoubtedly, the primary practice where consumer scores are involved is targeted advertising. In the last fifteen years, advertising companies and intermediaries have been at the forefront in developing services to provide their clients with the ability to target consumers that are more likely to be influenced by marketing campaigns. The use of AI systems for targeted advertising uses the same tools as consumer scoring. Predictive models are used to estimate the probability of a positive engagement of specific consumers with a particular advertising campaign and select them for bidding. The practice is more complex than this. I will say more about targeted advertising in the next section. But first, it is essential to reflect on this complex of algorithmic practices as they already carry several significant implications for consumers and their privacy.

1.2 Data research through the prism of surveillance

The monitoring of consumer behaviours by businesses has been an object of study in the last thirty years under the study of "commercial surveillance". According to the most neutral account, such commercial surveillance comprises all activities that public or

[31] Barnhard (2020).

[32] S. Chen (2018).

[33] Camilleri (2020).

private organisations, or even individuals, carry out in closely watching someone or something.[34]

Since the emergence of database marketing and customer relationship management, corporate information practices have caught the attention of surveillance scholars and critical marketing studies because they play themselves out in an imposed activity of constantly monitoring of consumers' behaviours.[35] Big data and analytics applications are just the latest evolution in a line of consumer-surveillance technologies. By using these technologies, marketers seek to extract value from an extensive array of new data-generating sources used by consumers, where the aim is to gain new insights into consumer behaviours and preferences.[36]

Nevertheless, when viewed from a corporate surveillance perspective, big data analytics proves to be particularly problematic for several reasons.

1.2.1 The characteristics of big data surveillance

While traditional surveillance practices have always relied on the purposeful collection of specific data to create models, corporate surveillance in big data environments is best characterised as "populational."[37] According to the original view of marketing, information was used to decide how to position products among consumers and create the best communication strategies. Consequently, early surveillance strategies were aimed at capturing specific information assets that were instrumental in creating and implementing the marketing strategy, be it a promotional, discount, or advertising campaign. In big data, information extraction and decision-making rely on inductive

[34] In the Foucauldian tradition, Lyon (2007), defines surveillance (from French *surveiller*, "to watch over") as "the focused, systematic and routine attention to personal details for purposes of influence, management, protection or direction" (p. 14). The author cautions that, contrary to what is commonly assumed, the term *surveillance* does not carry any evaluative overtones, but it is merely descriptive. It includes everything from face-to-face encounters to mediated arrangements dependent on a comprehensive and ever-growing range of information technologies.

[35] Together with governmental surveillance, the corporate use of information technologies has been among the recurrent research interests of surveillance scholars. For example, in one of the earliest studies on corporate surveillance, Gandy Jr. (1993, p. 1) relied on Bentham's idea of the panopticon to describe technology-mediated corporate practices as "a kind of high-tech, cybernetic triage through which individuals and groups of people are being sorted according to their presumed economic value." Similarly, Clarke (1988) was concerned about the increasing use of information technologies in commercial organisations and coined the term *dataveillance* to stress how longstanding forms of visual and communication surveillance had been progressively replaced by more economically viable and technically efficient computerised means. Others who have spoken of corporate surveillance and technology, include Lyon (1994, 2010); Haggerty and Ericson (2000); Zwick and Dholakia (2004). For an historical overview of this research strand, see Pridmore and Zwick (2011).

[36] Ball, Di Domenico, and Nunan (2016), p. 61: "thus we are forced to ask what the 'big' in 'big data' refers to. While the etymology of the term encourages a focus on the volume of data, it refers in fact instead more to the ubiquity of data, the completeness of coverage over contemporary lives. It is this ubiquity, the knowledge of a near-complete record of individual lives, which removes the need for a priori decisions on commencing surveillance.").

[37] Andrejevic and Gates (2014).

logic, whereby surveillance of the entire customer base is required. The more data are available, the more correlations and patterns can be subsequently detected to create a more effective context for decision-making. This new "omnivorous approach"[38] to data facilitates, and may even induce, a surveillance culture that extends beyond constrained informational assets (e.g., age, sex, gender, purchases) to encompass any possible data trail left by consumer activities on the Internet. Data in itself is less important than the information that can subsequently be extracted from it. The single data point becomes valuable only as long as it can be stored and aggregated with another myriad data points and analysed (or mined). Because machine-learning systems are agnostic,[39] any information can be relevant in extracting future knowledge. This is why commercial surveillance ranges across the full spectrum of available information about consumer behaviours. Every visited website, every link clicked, even hovering across the page with a mouse, may result in detecting useful patterns and correlations in consumer behaviours. Actually, the frequently used "behaviour" is a misnomer, as it may refer to anything that can be digitally monitored or that leaves some digital footprint.[40] The asset in the algorithmic business lies not so much in quality as in scale.

In addition, big data surveillance practices are marked by an expanding movement not only in the scale of data collection, but also in scope. The emergence of Web 2.0 and then mobile commerce has expanded the terrain for corporate surveillance practices towards a wide-ranging scope of social life activities.[41] Social networks and content-generation platforms have enabled marketing organisations to acquire information directly from the "prosumer."[42] In contrast to traditional mass media, which could only track the address of commercial content, social media and content-generation platforms now make it possible to extend the commercial gaze to consumers who upload or watch videos on YouTube, upload or browse personal images on Flickr, or accumulate friends and connections on Facebook and LinkedIn. Prosumers are more receptive to commercial social media content than consumers are to TV advertising or radio content. This due to the Internet's decentralised, one-to-many communication. Big data surveillance is set up to use the work of prosumers, with the additional effect of making no difference between passive consumers and active producers of social and cultural work.[43] A buyer profile can be created relying on past purchases and website visits, which may fit with a marketing

[38] Mayer-Schönberger and Cukier (2013).

[39] Hildebrandt (2019), for example, describes machine learning as agnostic in the sense that the algorithms are oblivious to human bias or independent of the design choices that determine its accuracy.

[40] Fisher and Mehozay (2019). See also Krasmann (2020), who argues for a specific epistemological approach to algorithmic systems. Since these systems have no access to the "world of human sense-making," they reduce reality and paint a "behaviourist picture" of human modes of existence.

[41] Fuchs (2011).

[42] The concept of a prosumer was introduced by Toffler (1980) to welcome the uptake of a new political, economic, and democratic order characterised by "progressive blurring of the line that separates producer from consumers." However, surveillance scholars have argued that Toffler overlooked that "prosumption" would also affect power relations in decentralised working environments, reconfiguring production and the autonomy of labour. In particular, Fuchs (2011, p. 295) criticises the practice of digital platforms outsourcing work to users and consumers, who now work for free. That is the case, he argues, with social networks, which exploit user-generated content to make a profit through advertising.

[43] Fuchs (2011, p. 299), who on a Marxist approach argues that big data companies exploit prosumers as much as industrial capitalists exploit factory workers.

orientation. However, they also include postings and likes on social media, video uploads on YouTube, and comments in a private blog. Therefore, big data analysis implies a de-contextualisation of any variable of "behaviour," be it economic or commercial nature (e.g., the purchase of a product) or social (e.g., a post on Facebook). Any variable can be fit into a spurious correlation for the objective pursued by companies.

Another aspect of contemporary surveillance is the constant and incessant capture of data, blurring the line between online and "offline" activities. In fact, while online social life has been decontextualised, the offline social life of consumers is constantly being recontextualized through geo-localisation devices and live tracking.[44] If commercial surveillance started in the world of physical places, such as supermarkets and stores, the use of information technologies such as the Internet subsequently virtualised that world. Browsing data has become significant within the virtual journeys of the Internet user. The advent of mobile phones, apps, and lately wearable technologies has given birth to a significant exchange of information through which consumers are located, identified, and authenticated, thereby linking consumers' activities to places. So-called "m-commerce" (or mobile commerce) has increasingly made it possible to disseminate commercial message to consumers as they move, thus making it possible to serve consumers with just-in-time personal messages specific to their location.[45]

Big data analytics practices are also increasingly relying on predictions rather than explanations. This transition marks a notable achievement in consumer surveillance, which is focused not only on generating knowledge about consumers' current and past behaviours, but also on attempting to predict future behaviours.[46] The focus is no longer on data patterns as they emerge from behaviours but on "(meta)data patterns to individual's potential behavior which yields powerful information about what we will do."[47]

Finally, contemporary consumer surveillance practices have a networked nature and are highly organized and strategic.[48] Marketers have always been concerned with controlling consumers across time and space but were primarily able to do so within the remit of bilateral transactions. Clear examples are direct and email marketing. Today, marketers can follow consumers everywhere, turning the mobility of everyday life into input for the "more diffuse and expanded systems of production." The information strategy of data-

[44] See Lyon (2010), claiming that there exists a new type of surveillance he calls "mobi-veillance."

[45] The digitally mediated contextualisation of consumers has been analysed by surveillance scholars under the label "brand-scape," defined by Wood and Ball (2013) as "a new mode of ordering that seeks to simultaneously construct space and subjectivity, a mode of ordering represented by a new apparatus."

[46] In making the predictive analytics of machine learning understandable and appealing to marketers, Siegel (2013, p. 20) argues that "[w]e usually don't know about causation, and we often don't necessarily care. [...] the objective is more to predict than it is to understand the world. [...] It just needs to work; prediction trumps explanation."

[47] Van Dijck (2014, p. 200), who deconstructs the ideological grounds of data-driven culture as rooted in problematic ontological and epistemological claims (so-called dataism). As part of this logic, predictive analytics reveals "a slippery slope between analysis and projection, between deduction and prediction."

[48] See, Cohen (2016), mocking the rhetoric of participation and openness in today networked data flows (s.c. the "participatory turn"), which hides the design to position surveillance outside from legal and social control.

driven businesses is no longer limited to tracking consumer behaviours in direct interactions. As discussed in the previous section, their strategy includes surreptitious tracking technologies and data partnerships with other companies, which can link and sync its corporate database and "collaborative analytics platforms" with third parties.

Looking at the scope and scale of big data practices, as well as at predictive orientations and the drive toward increasingly networked environments, social psychologist Shoshana Zuboff has labelled the present information economy the age of surveillance capitalism.[49] In her book, Zuboff takes inspiration from the work of Karl Polanyi, who observed that the advent of capitalism (the "great transformation") consisted in the market's progressive control of social spaces.[50] More to the point, Polanyi analysed the way in which capitalism, through a process of commodification, transforms land, labour, and money into products to be bought and sold on the market, where they otherwise would not have lent themselves to that use. According to Zuboff, in surveillance capitalism, the dominance of the market extends to "human experience," which is claimed "as free raw material for hidden commercial practices of extraction, prediction and sales."[51]

Surveillance capitalists are said to respond to two imperatives. The extraction imperative establishes that the algorithmic systems run by the surveillance capitalist can provide good results only insofar as significant quantities of data are injected into the system.[52] The new competition in surveillance capitalism is therefore finding new "mines" of (free) raw material which can guarantee continuous access. The extraction imperative would explain why enterprises such as Google and Facebook embark on all kinds of apparently unrelated services (e.g., search engines, game streaming, advertising, wearables, Internet access service). Such services help to funnel all forms of mediated computation into an extraction architecture, which is not interested in personal data ("in you") but only in the knowledge that can be extracted from it.

The second imperative of surveillance capitalism, which has partly replaced the first, is prediction. The commodity for trade in surveillance capitalism is no longer merely knowledge but "predictions" or "predictive models" about large subsets of people: this knowledge is used to anticipate or "get ahead" of the consumer, or it is otherwise now sold to buyers interested in doing the same. Moving from the imperative of prediction, in Zuboff's construction, surveillance capitalism finally arrives at what she calls "action economies." Predicting behaviour is aimed at intervening in the target to influence behaviour in real life. In order to achieve this, "machine processes are configured to

[49] Zuboff (2019).

[50] Polanyi (1944).

[51] This is only one of eight definitions given to the "surveillance capitalism." For a flavour of the range of ideas addressed in the book, consider the other seven: (*i*) "a parasitic economic logic in which the production of goods and services is subordinated to a new global architecture of behavioral modification"; (*ii*) "a rogue mutation of capitalism marked by concentrations of wealth, knowledge, and power unprecedented in human history"; (*iii*) "the foundational framework of a surveillance economy"; (*iv*) "as significant a threat to human nature in the twenty-first century as industrial capitalism was to the natural world in the nineteenth and twentieth"; (*v*) "the origin of a new instrumentarian power that asserts dominance over society and presents startling challenges to market democracy"; (*vi*) "a movement that aims to impose a new collective order based on total certainty"; and (*vii*) "an expropriation of critical human rights that is best understood as a coup from above: an overthrow of the people's sovereignty."

[52] See on this also, Sadowski (2019).

intervene in the state of play in the real world among real people and things."[53] The shift from predicting behaviour to modifying it will be analysed in the next chapter, as this requires looking at practices aimed at influencing consumer behaviour.

1.2.2 Algorithmic assemblage and the production paradigm

According to Zuboff, algorithmic systems, or what she calls "machine intelligence," represents the new means of production—the "new factory".[54] There is a two-way relationship between data and algorithmic systems and the whole surveillance operation. On the one hand, algorithmic systems transform raw material (consumer data) into refined products (extracted knowledge and predictions). On the other, the raw material must constantly be injected into the machine for it to work. AI allows businesses to make sense of consumer data and make decisions only to the extent that already large, varied, and updated behavioural data are available to train the system and improve learning over time. This two-way relationship creates a self-reinforcing loop where data-hungry machines spur more data collection, leading to refined predictions.

Still, the industrial-production metaphor appears especially appropriate in describing the role of algorithmic systems in the process of transforming raw data into actionable insights.

In digital environments, consumers are no longer physical bodies roaming about in shopping malls or in and out of grocery stores but a set of objective data points that flow into the database and can be stored and transferred as binary bits.[55] In this setting, algorithmic systems are supposed to reconstruct consumers' subjectivity as standardised and comparable structures that can be screened out or set up for direct intervention by marketers.[56] The consumer is converted into a digital assemblages and acquires meaning only in connection with other assemblages. The new digital reality is made up of new virtual subjects which emerge as "data doubles,"[57] consisting of attributes, personal descriptions, connections, and predictive scores reconstructed by the algorithm.[58]

However, as has long been argued in critical data studies, it is essential to be reminded that there is nothing "natural" about the inferences and predictions made about the data analysed. In fact, "everything about information is artificial,"[59] and relying on

---

[53] Zuboff (2019), p. 199.

[54] Zuboff (2019), p. 38.

[55] Negroponte et al. (1997).

[56] Zwick and Dholakia (2004).

[57] The concept shows a clear link to digital-driven buyer personas as a product of marketing creations.

[58] Haggerty and Ericson (2000, p. 606): "this assemblage operates by abstracting human bodies from their territorial settings and separating them into a series of discrete flows. These flows are then reassembled into distinct 'data doubles,' which can be scrutinised and targeted for intervention. In the process, we are witnessing a rhizomatic levelling of the hierarchy of surveillance, such that groups which were previously exempt from routine surveillance are now increasingly being monitored." Data doubles are often referred to by other labels, such as "dividuals" (Deleuze 1992) and "data shadows" (Simon 2005).

[59] Lanier (2010).

mathematical models or statistical methods does not guarantee that the knowledge created is undeniable or truthful.[60] Even when the insights derived from data are based on rigorous analytical procedures, they should not be considered neutral or objective realities. A complex mix of hidden intentions, systematic and random errors, partial information, or biased visions of the problem contributes to making this new knowledge as situated and partial as any other type of knowledge. The same algorithms developed to analyse data should be treated as contested, situated objects of inquiry.[61]

The "subjectification" by which algorithmic assemblage creates data doubles therefore needs to be carefully approached. Algorithmic systems sift through the database, creating groups, profiles, and scores, creating a digital reality where desirable "data doubles" are set apart from "data undesirable" for a specific product or marketing message.[62] In this setting, the meaning of individual consumers is of no interest to the algorithmic business or to the other actors in the value chain.[63] In other words, there is no effective guarantee that consumer profiles represent the individual self.

There is a longstanding aphorism in media theory that says, "If you are not paying for the product, or you don't know what the product is, *you* are the product."[64] The claim has a very particular meaning in the algorithmic business. Here, the traditional production paradigm which characterised the Fordist economy is supplanted by another approach.[65] The algorithmic business does not abandon the production function: it only shifts its object from producing a product to manufacturing a consumer. From this perspective, the baseline approach to algorithmic marketing is that managers are less interested in finding the best tactics for selling a product or service to consumers than in finding the best consumers that fit its current products. It is in this light that "consumers become the product." Businesses no longer spend time selecting the best campaign with which to generate demand and lure consumers into buying; rather, algorithmic systems operate by "recursive selection of the best consumer or the best groups of consumers" for specific goods. This new production paradigm is apparent in audience research, which is not concerned with finding the best advertising campaign with which to attract the attention

[60] See Boyd and Crawford (2012), who deconstruct the widespread mythology that "large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy."

[61] Barocas, Hood, and Ziewitz (2013).

[62] Pridmore and Zwick (2011) argue that the surveillance character of algorithmic assemblage does not manifest with the individualisation of identities. Instead, it is concerned with "the collection of personal information to discriminate individuals into previously categorised consumer lifestyle groups or profiles".

[63] The method of consumers protection and how big data are valued reflect what Zuboff (2019, p. 535) calls a "formal indifference" of surveillance capitalism to its population of customers and users.

[64] The phrase was already used prior to the Internet boom of the late 90s. For example, one can see nearly that exact phrase (or at least certainly the exact principle) described in Serra and Weyergraf (1980) on TV advertising. The interview references a short film titled *Television Delivers People*, made by the American video artist in 1973, who claimed that "The Product of Television, Commercial Television, is the Audience."

[65] Zwick and Dholakia (2013).

of consumers but is to find the best consumers, the ones who are most likely to be attracted by a particular product.[66]

As we will see in greater detail in the next chapter, problems may therefore arise when knowledge about data doubles is used to sort consumers into different categories and is packaged into corresponding consumer offers.

1.3 Commercial surveillance and the struggle for privacy

If the new market practice portends a new socio-technical framework where businesses constantly monitor, analyse, and score consumer behaviours, we might expect privacy laws to step in. At least in Western democracies, resistance to surveillance has traditionally been mobilised under the concept of privacy.[67] Yet privacy laws turn out to have little force when it comes to most contemporary commercial surveillance.

To be sure, the EU has regulations in place that govern commercial data practices. Particularly significant among these regulations, having outstripped any other attempt to ensure respect for privacy and data protection in the EU and beyond, is the General Data Protection Regulation.[68] However, this regulation is based on theoretical premises that appear insufficient to prevent commercial surveillance in the algorithmic business. To understand why, it is necessary to have a closer look at EU privacy legislation.

1.3.1 Introducing consumer privacy

Unlike other legal systems whose approach was much more market-oriented, the EU installed barriers against surveillance by proceeding from a fundamental-rights perspective.[69]

In 1949, following the atrocities of the Second World War, the Council of Europe was established, bringing together the states of Europe to promote the rule of law, democracy, human rights, and social development. To this end, it adopted the European Convention on Human Rights (ECHR) in 1950, which entered into force in 1953. The ECHR contained Article 8, "Right to Respect for Private and Family Life," which reads as follows:

---

[66] Pridmore and Zwick (2011, p. 122) speak of "consumer brands."

[67] Holvast (2007).

[68] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ L 119, pp. 1–88 (GDPR for short).

[69] For example, on the differences between the EU and US approaches to granting individuals rights to personal information, see Paul M. Schwartz (2003).

> 1. Everyone has the right to respect for his private and family life, his home and his correspondence.
>
> 2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

The theoretical background lies in the US development of the right to be left alone, as absorbed into the EU via the famous 1956 judgment of the German Constitutional Court which recognised the right to privacy as a fundamental right of the person.[70]

Subsequently, the right to privacy was reconceptualised, mainly in response to advancements in technology and to commercial developments.[71] In particular, with the rise of digital technologies and the electronic marketplace, privacy lawyers realised that there emerged qualitatively new possibilities for a state to intrude on the life of an individual. The Council of Europe thus adopted the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data.[72] It did so as a part of the Article 1 mandate, which reads as follows:

> The purpose of this Convention is to secure in the territory of each Party for every individual, whatever his nationality or residence, respect for his rights and fundamental freedoms, and in particular his right to privacy, with regard to automatic processing of personal data relating to him ("data protection").[73]

The shift from privacy as the right to be left alone to the right to data protection made its way into legislation. In 1995 the EU adopted the Data Protection Directive,[74] which, in keeping with the obligations set forth in Convention 108, sought to harmonise data-protection legislation across all Member States.

In 2000 the EU enacted the Charter of Fundamental Rights,[75] which later became a part of the EU's primary law. In it is a new fundamental right to "protection of personal data," which sits alongside the right of privacy:

> 1. Everyone has the right to the protection of personal data concerning him or her.

---

[70] Warren and Brandeis (1890). The paper is considered the seminal essay in the history of privacy, by American and privacy lawyers alike.

[71] Among the many theories of privacy, Tavani (2008) singles out three. The first is Warren and Brandeis's traditional notion of privacy, ascribed to the category of "physical privacy," which is the freedom from physical intrusion. The second and third are "decisional privacy" and "psychological privacy," concerned with protection from interference in important life decisions and the protection of one's intimate thoughts.

[72] Council of Europe, Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Jan. 28, 1981, ETS. No. 108.

[73] Ibid., Article 1.

[74] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281, pp. 31–50 (also Data Protection Directive)

[75] Charter of Fundamental Rights of the European Union, OJ C 326, 26.10.2012, pp. 391–407.

> 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified
>
> 3. Compliance with these rules shall be subject to control by an independent authority.[76]

The fundamental right to privacy and the fundamental right to data protection were unified in the GDPR. Approved in 2016 and entered into force in May 2018, the GDPR provided a general framework on data-protection law in the EU, putting forward the idea of informational self-determination.[77] The basic idea is that individuals enjoy privacy if they can set up private zones in which their personal information is held and if they can limit access to it from others, and that and if they consent to personal data disclosure, they can retain control over such data throughout its processing.[78] Both concepts operate on the assumption that individuals have a choice: they can control (decide) if and to what extent they want to restrict access to personal information. The GDPR is a general framework applicable to natural and legal persons that process personal data. It provides individuals with a series of ex-ante transparency rights before controlling access to personal data and ex-post actionable rights. It also imposes several obligations on data collectors to ensure that personal data are processed in accordance with citizens' fundamental rights.

The GDPR has genuinely been one of the most outstanding achievements for individual privacy in this century. However, when assessed against the backdrop of data-driven marketing, its regulatory approach has two significant shortcomings that make it largely inadequate to reduce commercial surveillance. The first relates to legal policy; the second is conceptual.

1.3.2 Consenting privacy away

An important reason for the hands-off approach to consumer data collection in current surveillance practices is that most of these practices are governed by a longstanding legal paradigm, the so-called "notice and consent".

The fundamental concern in developing data-protection principles and legislation was to foster an appropriate balance between individuals' privacy and the free-flow of information. The idea is that personal data should be collected and processed only when necessary and proportionate to the achievement of a legitimate aim or when individuals have freely expressed their informed consent. Indeed, the right to privacy self-determination aims to guarantee that individuals are free to construct the person they wish to become, have autonomy of action and thought, and can decide whether to give third parties access to their private life.

---

[76] Ibid, Article 8.

[77] Rouvroy and Poullet (2009).

[78] This is clearly stated in the preamble of the GDPR, Recital 7 ("Natural persons should have control of their own personal data") and in Recital 39 ("Personal data should be processed in a manner that ensures appropriate security and confidentiality of the personal data, including for preventing unauthorised access to or use of personal data and the equipment used for the processing").

Based on these considerations, the "privacy self-management" approach has been developed to provide individuals with the ability to decide whether to consent to disclosing their data. Crucially, the approach assumed that those who are appropriately notified of the reason, context, and purpose of their data collection, processing, or disclosure would be able to decide whether to *freely* consent to such activities. We can see this approach at work when the GDPR requires the processing of personal data to have a legal basis. Central among such legal bases is the consent of the individuals concerned.[79] In keeping with Article 7, the consent requirement has been strengthened in the form of consent that needs to be "freely given, specific, informed and unambiguous."[80] In order for consent to be informed, the data controller is required to provide the data subject with many essential pieces of information. These include the identity of the controller or the processor, the purpose of the processing, and the reason that explains why the processing is done.

Although, in theory, such a regulatory approach seems to help individuals exercise control over their personal data, in practice it provides limited protection in the context of contemporary commercial surveillance practices.

So, on the one hand, the GDPR rests on an intuitive assumption about individuals making privacy choices and deciding whether to express consent: not only should they be duly informed in making those choices, but they are presumably in a position where they *can* be so informed (with all information they need to that end). But, on the other hand, this assumption has extensively been proven to be not only fallacious,[81] but also *de facto* extremely difficult to put into practice.[82] Studies have shown that consumers faced with lengthy and complex privacy documents tend to completely disregard them. This information overload poses a threat to individuals' ability and motivation to scrutinise the key details that are necessary to make informed privacy decisions. Moreover, the need to give consent in real time, often through elementary actions (such as ticking a box or pressing a button), prompts people to neglect privacy statements and resort to simple heuristics in order to have seamless access to online interactions and services. Also prompting people to give consent is the inherent human tendency to focus on certain and proximate advantages, rather than on the uncertain, remote, and nebulous risks attendant on the processing of their data.

Even when someone is willing to read privacy policies, these documents are so numerous and complex as to stymie any effort of the average consumer or layperson to understand what they actually mean. Thus, a study has found that the average Web user would have

---

[79] Article 6 GDPR.

[80] Recital 32 GDPR: "Consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement."

[81] Indeed, the simpler the law makes the consent procedure, the less users are likely to understand what they are being asked to consent to. But then the more meaningful you make the consent procedure, by providing sufficient information about your data processing, the more time-consuming and demanding it will be for users to go through all that information in forming an informed judgment about giving their consent to such processing.

[82] Barocas and Nissenbaum (2009).

to spend 8 hours a day for 76 days every year to read the privacy policies of the websites they visit. And even if they did have that time, privacy notices often use terminology the layperson cannot understand, in that they refer to different technologies, data-collection mechanisms, business activities, and legal concepts.[83] On top of that, privacy policies often present information ambiguously and vaguely, using terms that do not convey specific information.[84] Privacy policies contain abstract terms that do little, if anything, to give meaning to the idea of informed consent (as when we are told that data will be collected for "personalization" or shared with "third-party business affiliates"), nor does the language or terminology help when it seems straightforward and concrete (e.g., collecting "data about you").[85]

More and more often, refusing to agree to data collection is not a real option for consumers. Consumers know that even if they had sufficient knowledge about the collection and use of their data, they would likely end up consenting, since refusing would mean not being able to accomplish the computer-mediated activity at issue (as when interested in a certain product or a particular service).[86] After all, it makes no sense to devote time and energy to reading complex policies and trying to figure out the ways in which our data will be used if the subsequent deliberation always ends with a "yes" to the processing under conditions unilaterally established by the controller. Consumers nowadays can hardly avoid consenting to privacy policies without opting out of much of modern life: all websites, whether commercial or otherwise, use them. In many cases, consumers are required to consent, either because offers and services that are not based on invasive digital tracking are not available or because not participating would lead to serious social, economic, cultural disadvantages in life.[87] In addition, such a forced choice is the baseline scenario when a service is provided under conditions of quasi-monopoly, or when market pressures lead the majority of players to converge on more stringent surveillance practices, as in the case of data-driven marketing.

The problems with notice and consent are even more visible in the context of big-data surveillance, considering the specific socio-technical operation.[88] Big data make it

---

[83] According to a study carried out in Sherman (2008), a proper understanding of the meaning of privacy documents would require the IQ of an average PhD.

[84] Legally speaking, such practices would run contrary to Article 12 GDPR, which requires the information given in privacy policies be "concise, transparent, intelligible and easily accessible form" and be stated in "clear and plain language."

[85] Reidenberg et al. (2016).

[86] Koops (2014), p. 251: "often, there is little to choose: if you want to use a service, you have to comply with the conditions—if you do not tick the consent box, access will be denied."

[87] This social fact could naively be regarded as a privacy paradox. On the one hand, individuals seem to be concerned about their informational privacy and to want to protect it. On the other hand, they voluntarily disclose information online and rarely exercise their data-protection rights. However, an alternative explanation is that people may want privacy and understand its importance, and they are even aware of the rights they have, but in the current data-driven marketplace, using and sharing data has become such an integral part of their life that their "choice" is not a real one. Such an explanation calls into question whether the decisions made by individuals when sharing data are indeed paradoxical or instead simply reflect their needs in a highly "datafied" society.

[88] See Mantelero (2014), who calls for a general acknowledgement of the "transformative approach of Big data."

technically impossible for the entity collecting the data to provide meaningful information on the type of data collected or its purposes before the act of processing, except by referring to broad categories and making vague statements. As mentioned in the previous section, the essence of information strategies based on big data is to capture or record any activity as a preventive measure and then determine how to use the information for marketing purposes. In data-driven marketing, consumer information will naturally be used to provide commercial offers and improve services. However, such information is arguably not enough to ensure informed consent to personal-data processing. There are still many unknowns: what will activity data show? How sensitive will this information be? What conclusions will data-analysis algorithms come to? How will the analysis be incorporated into their interest profiles? How will this profile be used for advertising? These questions are difficult to answer precisely because the processing activities employed in the data-driven market do not have an established end goal. Nor does it have a predefined hypothesis for which it only collects a statistically significant sample. A consumer who agrees to data processing will not be able to know how her data are being processed and used.[89]

Aside from the inherent flaws of notice and consent, there is also the problem that the model of individual consent as a means of managing privacy fails to reflect the collective interests connected to contemporary privacy choices.[90] When the average consumer is fully informed and aware of the costs and benefits of disclosing her data and allows companies to collect and analyse her personal information, her choice affects not only her privacy but that of everyone in her cohort. The personal data of a particular consumer can be used to build profiles that can be applied to the entire group, including those who did not consent to the processing. Thus, all members of the group who have similar health issues, social conditions, or psychological attitudes are potentially affected. As soon as the system is provided with data (predictors) about them, further information can be inferred based on the profile, even though they did not consent in the first place to any processing of personal data. Individual consent does not consider the externalities of the processing, namely, the extent to which other individuals and societal arrangements are affected, beyond individual privacy.

Given the above-mentioned criticisms, the notice-and-consent mechanism has in practice evolved into a contractual pay-off model, consisting in trading consent to personal-data

[89] On the difficulty of reconciling privacy principles with big data, see also Tal Z. Zarsky (2016). Among other things, big data would be incompatible with the principle of data minimisation, which under Article 5 GDPR, requires that "personal data shall be […] adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed." Apparent conflicts emerge with the principle of purpose limitation, requiring that "personal data shall be […] collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes" (Article 5(1)(b) GDPR). On the possibility of reconnecting AI and big-data applications with the principle of EU data protection though extensive interpretation, see Sartor (2020b).

[90] See Baruh and Popescu (2017), who argue that the predominant privacy-protection regimes underplay privacy as a collective good. In particular, they show how the two common individual privacy strategies—i.e., not consenting to privacy policies and completely relying on market-provided privacy protections—may result in fewer privacy options available to society at large. On a similar note, Mantelero (2016) explains that the atomistic approach to data protection no longer holds in mass-predictive data analysis. In such a new context, privacy interests have an additional layer consisting of in "the collective dimension of data protection, which protects groups of persons from the potential harms of discriminatory and invasive forms of data processing."

exploitation in exchange for the ability to access online services.[91] As such, it seems that, rather than curtailing big-data surveillance, the notice-and-consent model contributes to creating the conditions to transform the privacy of users' data into a commodity that can be analysed and traded for services. We will consider this point in more detail when addressing the concept of digital vulnerability (Chapter 7).[92]

1.3.3 The economic value of personal information

The second reason why the effectiveness of data protection laws is limited in contemporary commercial surveillance practices—that it is relevant to address here—concerns its conceptual approach to defining personal information.

On the fundamental-rights approach, data protection law is concerned with personal information to the extent that such information identifies or makes identifiable data subjects as natural persons. This option is reflected in the definition of personal data contained in the GDPR:

> "personal data" means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person [...][93]

The notion of personal data is not limited to data that identify natural persons but also includes those data that render them identifiable. The GDPR addresses identifiability in Recital 26, where it is cast in terms of the conditions under which a piece of data not explicitly linked to a person still counts as personal data, in that there is still the possibility of identifying the person concerned. This possibility rests on the availability of technological means which can provide such re-identification:

> To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly.
>
> To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the

[91] The assumption that consent is adequately informed based on acceptance of online service contracts turns out to be "the biggest lie on the Internet."

[92] Hull (2015) argues that the failure of notice and consent to protect privacy, while at the same time making it easier to exploit personal data, may be the fruit of specific ideological choices. The notice-and-consent model consecrates the belief that, despite the scope of fundamental rights, privacy can only be treated in terms of an individual's economic choice to disclose information.

[93] Article 4(1) GDPR.

> amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.[94]

One of the crucial points about commercial surveillance is that companies do not need to identify individuals to use the data economically, but only need to be able to make assessments about their behaviour.[95] The meaning of what an individual is does not concern big-data companies as much as the behavioural implications the data about them promise. Companies only need to make an assessment about groups of consumers by sorting them into segments, or classes, and attributing predictive scores to understand how they are likely to behave in response to a particular commercial initiative. There needs to be an ability to signal-out consumers to address them, as by assigning cookies to their browsers or an ID to their device, but from a marketers' point of view, personal information is workable even if consumers remain anonymous.[96]

The status of datasets containing non-identifiable data is somewhat ambiguous in current data-privacy law.[97] Since data protection is mainly concerned with the risk connected to personal information, the current privacy regime favours those data practices where personal data are collected so that the risk of identification is mitigated. Examples are anonymised and pseudo-anonymised data.

Anonymised data are personal data collected or processed so that the data subject is not or no longer identifiable in an irreversible way. Recital 26 GDPR clearly states that the regulation itself does not apply to the processing of anonymous information, provided that careful engineering and constant monitoring are ensured for the data to remain deidentified. Therefore, data controllers and processors do not have to abide by the principles and rules of data protection.

The other category of de-identified data is pseudonymised data. Pseudonymised data are defined as data requiring additional information to identify the data subject. Such information, however, is held separately and securely through the use of technical and organisational measures by which to "ensure that the personal data is not attributed to an identified or identifiable natural person."[98] Pseudo-anonymised data fall within the scope of application of the GDPR. Therefore, data controllers are required to have a legal basis for their processing, complying with the duty to inform data subjects and to respect their rights. However, the GDPR allows for a meaningful relaxation of the purpose-specification principle, since Article 6(4) permits the processing of pseudo-anonymised data for uses beyond the purpose for which the data were initially collected.[99]

---

[94] Recital 26 GDPR.

[95] On this point, see Paul M. Schwartz and Solove (2011), who caution that the restricted notion of personal information risks leaving the behavioural data markets untouched, even if they probably rank among the most privacy-invasive spaces.

[96] Andrew and Baker (2019).

[97] On whether behavioural data in targeted advertising can be covered by data-protection legislation, see also Zuiderveen Borgesius (2016).

[98] Article 2(5) GDPR.

[99] Article 6(4)(e): "Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives

For companies subject to the GDPR's new restrictions, anonymisation and, to a lesser degree, pseudo-anonymisation has become an appealing option. With anonymisation, they are no longer bound to comply or fully comply with the EU data-protection principles and rules. Indeed, legally speaking, many marketers prefer to collect or process anonymised or pseudo-anonymised data because they know that the less personal data they process about real consumers, the more they can analyse their behaviour and make assessments about them and trade these knowledge products. While pseudonymised and anonymised data provide additional protections for informational privacy, they do not diminish the risk of surveillance. In fact, given the chosen legislative options, they seem to facilitate it. Thus, paradoxically, the very de-identification "safeguards" designed to shield individuals from extraction of their private data institutionalise the exchange of their behavioural data, which despite non identifying the individual are those that companies value the most.[100]

A further issue regarding the conceptualisation of personal information, and thus the actual protective scope of privacy legislation in current surveillance markets, concerns the status of inferred data. As noted above, marketers infer new information about data subjects by applying algorithmic models to their personal data. The key issue, from a data-protection perspective, is whether the inferred information should be deemed new personal data, distinct from the data from which it has been inferred. Suppose, for instance, that a consumer's propensity to buy certain goods or the consumer's likelihood to have a certain personal characteristic is inferred from his or her online activity. Is the inferred presumed behaviour or personality type a new item of personal data? And is it so even when the inference is only probabilistic? If the inferred information counts as new personal data, the use of automated inferences would trigger all the consequences that the processing of personal data entails under the GDPR: the need of a legal basis, the conditions for processing sensitive data, the data subject's rights (such as the right of access, the right to rectification, and the right to erasure).

Nowhere does the GDPR mention the term "inferred data" or "inference." The rights it mentions that pertain to profiling do not count for much: they include the right to object to profiling under certain conditions (Art. 21) and the right not to be subjected to automated individual decision-making (Art. 22). If data subjects request any information about any profiling they may be under, data controllers will have to comply, while also providing meaningful information about the logic involved and its consequences for the

referred to in Article 23(1), the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia: […]. (e) the existence of appropriate safeguards, which may include encryption or pseudonymisation." This is complemented by Recital 29, providing that "in order to create incentives to apply pseudonymisation when processing personal data, measures of pseudonymisation should, whilst allowing general analysis, be possible within the same controller when that controller has taken technical and organisational measures necessary to ensure, for the processing concerned, that this Regulation is implemented, and that additional information for attributing the personal data to a specific data subject is kept separately."

[100] Andrew and Baker (2019). The authors conclude that "its reach will be limited if law makers fail to understand the broader behavioral data ecosystems they seek to regulate […] the law creates the space for a behavioral data market in which commercial self-interest is likely to flourish. In this way, whether deliberately or otherwise, the GDPR will make it possible for the behavioural data market to continue to function unencumbered."

data subject.[101] There is an extensive debate on how far this "right to explanation" actually goes.[102] However, few argue that there is an obligation for data controllers to disclose (1) the actual algorithms used, (2) the actual weighting of the data subject's data, and (3) data about other data subjects used in the profiling. Without such information, it is impossible for data subjects to check whether data about them is being inferred correctly. Companies may not be very keen to share algorithms and profiles, as these can be considered trade secrets of vital interest, giving them a competitive edge. These companies may also suggest that profiles are corporate secrets because they may, via reverse engineering, enable disclosure of their analyses and software.[103]

Even if the GDPR is not explicit on this point, it has been argued that data inferred through profiling should be considered personal data when they are ascribed to an identified or identifiable natural person.[104] In this connection, we need to distinguish the general correlations captured by the learned algorithmic model from the results of applying that model to the description of a particular individual. Consider, for instance, a machine learning system that has learned a model (e.g., a neural network or a decision tree) from a training set consisting of consumers' online activities and propensity to buy a certain product. In this example, the system's training set may consist of personal data about consumers, including their identifiers, the data collected about them, personal information in their profiles, browsed webpages, and items purchased. The learned algorithmic model no longer contains personal data since it links any possible combinations of possible input values (predictors) to a corresponding likelihood of buying (target). The correlations embedded in the algorithmic model are not personal data, since they apply to groups and classes, that is, to all individuals sharing similar characteristics. We can view them as group data, concerning the set of such individuals (e.g., those who are assigned a higher likelihood of buying that particular product, since they have a browsed certain webpage, have certain age, or included some basic interest in their shopping profiles, etc.). Suppose that the algorithmic model is then applied to the input data consisting in the description of a new consumer so as to determine that person's likelihood of buying that product. In this case both the description of the consumer and the likelihood attributed to him or her by the model count as personal data, the first being collected data, and the second inferred data.

The view that inferred data can be deemed personal data was endorsed by the Article 29 WP, being implied in particular by the broad concept of personal data adopted in Opinion 4/2007.[105] The view of extending protection to inferred data also seems to be presupposed

---

[101] Art. 13(2)(f), 14.2(g), and 15(1)(h).

[102] Among others, see Wachter, Mittelstadt, and Floridi (2017); Malgieri and Comandé (2017); Veale and Edwards (2018); Kaminski and Malgieri (2019).

[103] Hildebrandt and Rouvroy (2011), p. 23.

[104] On the possibility of considering inference as personal data under the GDPR, see extensively Sartor (2020a), p. 40ff.

[105] Article 29 Data Protection Working Party, Opinion 4/2007 on the Concept of Personal Data, adopted on 20 June 2007. In particular, the opinion follows that idea that in order for data to count as "related" to an individual, there needs to be a "content," "purpose," or "result" element. The first criterion means that information qualifies as personal data when it relates to a person, i.e., it is about that person, as when "the information contained in a company's folder under the name of a certain client clearly relates to him." Under the second criterion ("purpose"), an item of information counts as personal insofar as it is used to

by the Article 29 WP statement that in case of automated inference (profiling), data subjects have the right to access both the input data and the (final or intermediate) conclusions automatically inferred from such data.[106]

However, even if inferred data is personal data, there may still be practical issues with the rights of data subjects.[107] For instance, the right to rectification requires data subjects to show that the data are wrong, but it is impossible for data subjects to do so unless they can analytics tools and data about other data subjects used in the analysis. Obviously, data subjects may object to the profiling altogether, but this may be too rigorous. Data subjects may also consider transferring their data to other data controllers that offer more transparency on their profiling processes. This can be done via the right to data portability, under which data subjects have the right to receive personal data about them in a structured, commonly used and machine-readable format. However, according to the Article 29 Working Party, this right does not include inferred data, as this data is limited to personal data which a data subject has provided to a controller.[108] In fact, a data controller may further limit the right to data portability by inferring data while deleting the original data on which the inferences are based, even if this is done in reversible ways.

The ambiguity of the legal status of pseudo-anonymised and anonymised data and of the legal status of inferences is one relatively clear indication that data-protection law fails to grasp the commercial value of personal information. However, the limitations of the GDPR are understandable. Data-protection law reflects a fundamental-rights approach to privacy which enables the law to control extensive monitoring only insofar data poses a threat to individuals and render them identifiable in the real world. After all, the GDPR, and more generally data-protection legislation, focuses on personal data, and not on personal information or knowledge that can be exploited for economic purposes.

Tacking stock of the previous two sections, the focus on EU data protection on singular pieces of personal identifiable information and the right of individuals to control or restrict access through consent to that information risk losing sight of the larger societal power relations between data-driven companies and individuals in the digital economy. The individual-centred approach to informational privacy emphasises the fundamental right and responsibility of individuals to control their information, and indeed assumes that they can do so. But if privacy is conceptualised purely as individual choice, it misses the significant moves in contemporary society that go beyond the ability of individuals to make conscious decisions about their own informational self.

---

evaluate, treat in a certain way, or influence the status or behaviour of an individual. Under the "result" criterion, even without a "content" or "purpose" element, data can be considered personal whenever their use is likely to have an impact on a person's rights and interests.

[106] Article 29 Data Protection Working Party, Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679, adopted on 3 October 2017 as last revised and adopted on 6 February 2018: "In addition to general information about the processing, pursuant to Article 15(3), the controller has a duty to make available the data used as input to create the profile as well as access to information on the profile and details of which segments the data subject has been placed into" (p. 17).

[107] Wachter and Mittelstadt (2019).

[108] Article 29 Working Party, Guidelines on the Right to Data Portability (2016), p. 242.

References

Anderson C (2008) The end of theory: The data deluge makes the scientific method obsolete. Wired url: https://www.wired.com/2008/06/pb-theory/

Andrejevic M, Gates K (2014) Big data surveillance: Introduction. Surveillance & Society12(2), pp. 185–196

Andrew J, Baker M (2019) The General Data Protection Regulation in the Age of Surveillance Capitalism. Journal of Business Ethics 168(3), pp. 565–578

Balducci B, Marinova D (2018) Unstructured data in marketing. Journal of the Academy of Marketing Science 46(4), pp. 557–590

Ball K, Di Domenico ML, Nunan D (2016) Big data surveillance and the body-subject. Body & Society 22(2), pp. 58–81

Barnhard A (2020) A True End-to-End ML Example: Lead Scoring. url: https://towardsdatascience. com/a-true-end-to-end-ml-example-lead-scoring-f5b52e9a3c80

Barocas S, Hood S, Ziewitz M (2013) "Governing algorithms: A provocation piece" "Governing Algorithms" conference (paper presentation), May 16-17, 2013

Barocas S, Nissenbaum H (2009) "On notice: The trouble with notice and consent" Proceedings of the Engaging Data Forum: The First International Forum on the Application and Management of Personal Electronic Information

Bartsch M, Dienlin T (2016) Control your Facebook: An analysis of online privacy literacy. Computers in Human Behavior 56, pp. 147–154

Baruh L, Popescu M (2017) Big data analytics and the limits of privacy self-management. New media & society 19(4), pp. 579–596

Boyd D, Crawford K (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society 15(5), pp. 662–679

Brynjolfsson E, Hu YJ, Rahman MS (2013) Competing in the age of omnichannel retailing. MIT Cambridge

Burgess C, Burgess M (2020) The New Marketing: How to Win in the Digital Age. SAGE

Camilleri MA (2020) The use of data-driven technologies for customer-centric marketing. International Journal of Big Data Management 1(1), pp. 50–63

Chen S (2018) "Estimating customer lifetime value using machine learning techniques" Data mining ed. by C Thomas IntechOpen, pp. 17–34

Christl W, Spiekermann S (2016) Networks of Control – A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy. Report url: https://crackedlabs.org/en/networksofcontrol

Clarke R (1988) Information technology and dataveillance. Communications of the ACM 31(5), pp. 498–512

– (2019) Risks inherent in the digital surveillance economy: A research agenda. Journal of Information Technology 34(1), pp. 59–80

Cohen JE (2016) "The Surveillance-Innovation Complex: The Irony of the Participatory Turn" The Participatory Condition in the Digital Age ed. by D Barney et al. University of Minnesota Press, pp. 207–226

Constantiou ID, Kallinikos J (2015) New games, new rules: big data and the changing context of strategy. Journal of Information Technology 30(1), pp. 44–57

Davenport T (2014) Big data at work: dispelling the myths, uncovering the opportunities. Harvard Business Review Press

– (2018) From analytics to artificial intelligence. Journal of Business Analytics 1(2), pp. 73–80

Davenport T, Harris J (2017) Competing on analytics: Updated, with a new introduction: The new science of winning. Harvard Business Press

Deleuze G (1992) Postscript on the Societies of Control. Routledge

Diaz Ruiz CA, Kjellberg H (2020) Feral segmentation: How cultural intermediaries perform market segmentation in the wild. Marketing Theory 20(4), pp. 429–457

Erevelles S, Fukawa N, Swayne L (2016) Big Data consumer analytics and the transformation of marketing. Journal of Business Research 69(2), pp. 897–904

Fisher E, Mehozay Y (2019) How algorithms see their audience: media epistemes and the changing conception of the individual. Media, Culture & Society 41(8), pp. 1176–1191

Forbrukerrådet (2020) Out of Control. How consumers are exploited by the adtech industry - and what we are doing to make it stop. Report url: https://www.forbrukerradet.no/undersokelse/noundersokelsekategori/report-out-of-control/
Fuchs C (2011) Web 2.0, prosumption, and surveillance. Surveillance & Society 8(3), pp. 288–309
Gandy Jr OH (1993) The Panoptic Sort: A Political Economy of Personal Information. Westview
Haggerty KD, Ericson RV (2000) The surveillant assemblage. The British journal of sociology 51(4), pp. 605–622
Hildebrandt M (2019) Privacy as protection of the incomputable self: From agnostic to agonistic machine learning. Theoretical Inquiries in Law 20(1), pp. 83–121
Hildebrandt M, Rouvroy A (2011) Law, human agency and autonomic computing: the philosophy of law meets the philosophy of technology. Routledge
Holvast J (2007) "History of privacy" The History of Information Security: A Comprehensive Handbook ed. by KMM de Leeuw, J Bergstra Elsevier, pp. 737–769
Hull G (2015) Successful failure: what Foucault can teach us about privacy self-management in a world of Facebook and big data. Ethics and Information Technology 17(2), pp. 89–101
Jenkinson A (1994) Beyond segmentation. Journal of Targeting, Measurement and Analysis for Marketing 3(1), pp. 60–72
Kaminski ME, Malgieri G (2019) Algorithmic Impact Assessments under the GDPR: Producing Multilayered Explanations. Available at SSRN 3456224
Koops B-J (2014) The trouble with European data protection law. International Data Privacy Law 4(4), pp. 250–261
Krasmann S (2020) The logic of the surface: on the epistemology of algorithms in times of big data. Information, Communication & Society, pp. 1–14
Lanier J (2010) You are not a gadget: A manifesto. Vintage
Lomborg S, Bechmann A (2014) Using APIs for data collection on social media. The Information Society 30(4), pp. 256–265
Lyon D (1994) From big brother to electronic panopticon. The Electronic Eye: The Rise of Surveillance Society, pp. 57–80
– (2007) Surveillance studies: An overview. Polity
– (2010) "Surveillance, power and everyday life" Emerging Digital Spaces in Contemporary Society ed. by P Kalantzis-Cope, K Gherab-Martin Palgrave Macmillan UK, pp. 107–120
Malgieri G, Comandé G (2017) Why a right to legibility of automated decision-making exists in the general data protection regulation. International Data Privacy Law
Mantelero A (2014) The future of consumer data protection in the EU Re-thinking the "notice and consent" paradigm in the new era of predictive analytics. Computer Law & Security Review 30(6), pp. 643–660
– (2016) Personal data for decisional purposes in the age of analytics: From an individual to a collectivedimension of data protection. Computer Law & Security Review 32(2), pp. 238–255
Mayer-Schönberger V, Cukier K (2013) Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt
McAfee A, Brynjolfsson E (Oct. 2012) Big Data: The Management Revolution. Harvard Business Review url: https://hbr.org/2012/10/big-data-the-management-revolution
Mitchell A (2010) The rise of volunteered personal information. Journal of Direct, Data and Digital
Marketing Practice 12(2), pp. 154–164
Montgomery AL et al. (2004) Modeling online browsing and path analysis using clickstream data.
Marketing science 23(4), pp. 579–595
Morrison R (2015) Data-driven organization design: Sustaining the competitive edge through organizational analytics. Kogan Page Publishers
Negroponte N et al. (1997) Being digital. Computers in Physics 11(3), pp. 261–262
Obar JA, Oeldorf-Hirsch A (2020) The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. Information, Communication & Society 23(1), pp. 128–147
Park YJ (2013) Digital literacy and privacy behavior online. Communication Research 40(2), pp. 215– 236
Polanyi K (1944) The great transformation. Beacon Press
Popov A, Iakovleva D (2018) Adaptive look-alike targeting in social networks advertising. Procedia Computer Science 136, pp. 255–264
Pridmore J, Zwick D (2011) Marketing and the rise of commercial consumer surveillance. Surveillance & Society 8(3), pp. 269–277

Prince C et al. (2021) Are We Living in Surveillance Societies and Is Privacy an Illusion? An Empirical Study on Privacy Literacy and Privacy Concerns. IEEE Transactions on Engineering Management, pp. 1–18
Privacy Choice (Mar. 2010) Credibility Gap: What does Ghostery really see?. Post by Jim Brock url: http://blog.privacychoice.org/2010/03/04/credibility-gap-what-does-ghostery-really-see/ (visited on 12/12/2021)
Procter R, Voss A, Lvov I (2015) Audience research and social media data: Opportunities and challenges.
Participations: Journal of Audience Reception Studies 12(1), pp. 470–493
Reidenberg JR et al. (2016) Ambiguity in Privacy Policies and the Impact of Regulation. The Journal of Legal Studies 45(2), pp. 163–190
Ridley-Siegert T (2015) Data privacy: What the consumer really thinks. Journal of Direct, Data and Digital Marketing Practice 17(1), pp. 30–35
Rouvroy A, Poullet Y (2009) "The right to informational self-determination and the value of self-development: Reassessing the importance of privacy for democracy" Reinventing data protection? Ed. by S Gutwirth et al. Springer, pp. 45–76
Sadowski J (2019) When data is capital: Datafication, accumulation, and extraction. Big Data & Society 6(1)
Sartor G (Apr. 2020a) New aspects and challenges in consumer protection. Study PE 648.790 European Parliamentary Research Service, Policy Department for Economic, Scientific and Quality of Life Policies
– (Sept. 2020b) The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. Study No. PE 641.530 European Parliamentary Research Service, Scientific Foresight Unit (STOA)
Schaub F et al. (2016) "Watching them watching me: Browser extensions impact on user privacy awareness and concern" NDSS workshop on usable security, pp. 1–10
Schneider H et al. (2017) "Your data, your vis: Personalizing personal data visualizations" IFIP Conference on Human-Computer Interaction Springer, pp. 374–392
Schwartz PM (2003) Property, privacy, and personal data. Harv. L. Rev. 117, p. 2056
Schwartz PM, Solove DJ (2011) The PII problem: Privacy and a new concept of personally identifiable information. NYUL rev. 86, p. 1814
Serra R, Weyergraf C (1980) Richard Serra: Interviews, Etc., 1970-1980. Hudson River Museum
Sherman E (Sept. 2008) Privacy policies are great – for PhDs. CBS News 4 url: https://www.cbsnews.com/news/privacy-policies-are-great-for-phds/ (visited on 12/12/2021)
Siegel E (2013) Predictive analytics: The power to predict who will click, buy, lie, or die. John Wiley & Sons
Simon B (2005) The return of panopticism: Supervision, subjection and the new surveillance. Surveillance & Society 3(1)
Sipior JC, Ward BT, Mendoza RA (2011) Online privacy concerns associated with cookies, flash cookies, and web beacons. Journal of Internet Commerce 10(1), pp. 1–16
Strycharz J, Ausloos J, Helberger N (2020) Data Protection or Data Frustration? Individual Perceptions and Attitudes Towards the GDPR. European Data Protection Law Review 6(3), pp. 407–421
Tavani HT (2008) "Informational privacy: Concepts, theories, and controversies" The Handbook of Information and Computer Ethics ed. by KE Himma, HT Tavani John Wiley & Sons, pp. 131–164
Toffler A (1980) The third wave. William Morrow (US)
Van Dijck J (2014) Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. Surveillance & society 12(2), pp. 197–208
Van Otterlo M (2013) A machine learning view on profiling. Privacy, Due Process and the Computational Turn - Philosophers of Law Meet Philosophers of Technology. Abingdon: Routledge, pp. 41–64
Veale M, Edwards L (Apr. 2018) Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling. en Computer Law & Security Review 34(2), pp. 398–404 ISSN 02673649 doi 10.1016/j.clsr.2017.12.002 url: https://linkinghub.elsevier.com/retrieve/pii/S026736491730376X (visited on 11/29/2021)
Wachter S, Mittelstadt B (2019) A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. en Columbia Business Law Review, pp. 494–619 url: https://osf.io/mu2kf (visited on 11/29/2021)
Wachter S, Mittelstadt B, Floridi L (2017) Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law 7(2), pp. 76–99
Walker KL (2016) Surrendering information through the looking glass: Transparency, trust, and protection. Journal of Public Policy & Marketing 35(1), pp. 144–158

Warren SD, Brandeis LD (1890) The right to privacy. Harvard law review, pp. 193–220
Wood DM, Ball K (2013) Brandscapes of control? Surveillance, marketing and the co-construction of subjectivity and space in neo-liberal capitalism. Marketing Theory 13(1), pp. 47–67
Zarsky TZ (2016) Incompatible: The GDPR in the age of big data. Seton Hall L. Rev. 47(4), pp. 995–1020
Zuboff S (2019) The age of surveillance capitalism: the fight for the future at the new frontier of power. Profile Books
Zuiderveen Borgesius FJ (2016) Singling out people without knowing their names – Behavioural targeting, pseudonymous data, and the new Data Protection Regulation. Computer Law & Security Review 32(2), pp. 256–271
Zwick D, Dholakia N (2004) Consumer subjectivity in the Age of Internet: the radical concept of marketing control through customer relationship management. Information and Organization 14(3), pp. 211–236
– (2013) "Strategic database marketing: customer profiling as new product development" Marketing Management Routledge, pp. 481–496