4 Situating Personal Information: Privacy in the Algorithmic Age

Jens-Erik Mai

Introduction

Informational privacy is often understood as the ability or right to have control over one's personal information. In fact, as discussed in the Introduction to this volume, there is a growing concern that platforms use personal information in ways that compromise their users' right to privacy. In his classic definition, Westin stated that privacy is "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (Westin 1967, 7). This basic idea of a tight link between informational privacy and personal information is echoed in today's two major approaches to informational privacy: the *access* approach in which privacy is about the ability to "limit or restrict others from information about" oneself (Tavani 2008, 141) and the *control* approach in which privacy is the ability to have "control of personal information" (Solove 2008, 24). These notions of informational privacy turn on the basic assumption that *informational privacy is about the protection of personal information*. In this chapter, I will discuss and problematize this assumption and explore the notion of personal information as it is formed in the age of algorithms, datafication, and big data.

The notion of "information" often assumes that information somehow represents, relates to, corresponds to, or points to particular people, places, or things in the world. This use of "information" follows a tradition in the philosophy of information that takes data and information to be reified entities that can be manipulated and subjected to abstract, rational analysis and that exist independent of context, situation, time, and place. In this sense, information *just is*—or as information philosopher Tom Stonier

(1990, 21) stated, "*Information exists*. It does not need to be *perceived* to exist. It does not need to be *understood* to exist. It requires no intelligence to interpret it. It does not have to have *meaning* to exist. It exists."

I will argue, however, that personal information is in fact created in contexts, in situations, in use, and via conversations and constructions about the significance of the information. As such, *information doesn't simply exist*. The understanding of information I use here is based on semiotic communication theory that places the construction of meaning as central to understanding information (cf., e.g., Fiske 2011). It follows that information is best understood as a semiotic sign in communicative practices.

In the following section, I will review how a few central privacy theories have used the notion of information. Next, I present three cases that illustrate how data and information are collected, processed, and used in the algorithmic age. In the final section, I outline a theoretical framework for understanding information as a sign, and I discuss the consequences of such an approach for informational privacy.

To Control Information

The idea that information *just is* is most clearly articulated in the tradition that considers informational privacy to be concerned with property rights over information—that information has thinglike characteristics and is owned by specific people. Moore (2010, 5) conceptualizes privacy as the "right to control access to and use of physical items, like bodies and houses, and information, like medical and financial facts" and defends a "control over access and use" (ibid., 5) definition of privacy. The claim to privacy is a claim to "control access to places and ideas" (ibid., 23) which can be "written, recorded, spoken, or fixed in some other fashion" (ibid., 23). The basic idea is that we have a property right to our personal information and the condition of privacy is one of "voluntary seclusion or walling off" (ibid., 26) and making "personal information inaccessible" (ibid., 26). This idea was most clearly articulated by Murphy (1996, 2383–2384), who asserted about personal information that "such information, like all information, is property" and therefore the basic question is simply to determine "[w]ho owns the property rights to such information" (ibid., 2384).

Solove (2008, 29) criticizes the control theory for being "too vague, too broad, or too narrow." It is too vague because it fails to define the various

types of information in play in privacy situations; it is too broad because it fails to accommodate the relational nature of information and the fact that not all personal information is private; and lastly it is too narrow because it fails to account for privacy situations that are not informational in nature. Instead, Solove proposes that we "identify the various types of information and matters that are private" (ibid., 67) and determine those types of information in which intimacy or sensitivity is involved. The traditional approach has been to classify information as "public or private under the assumption that these are qualities that *inhere* in the information [emphasis added]" (ibid., 69). Solove proposes—from a grounding in "philosophy of pragmatism" (ibid., 46)—a contextual approach to privacy with a focus on specific privacy problems that there is a need to address, and he proposes that instead of defining privacy from a general perspective, the aim should be to start with practical problems that privacy theory and practice ought to solve and let those contexts define the privacy problems at hand and whether information is public or private.

Nissenbaum (2010) is likewise critical of the control approach to informational privacy. She argues that the control approach relies on the notion of a distinction between a public realm and a private realm and the ability to master the location of the information. Nissenbaum suggests that the social norms governing a given situation should provide the context to understand the privacy issues at stake. As such, Nissenbaum's focus is on the contextual, social norms that determine whether some information belongs in the public realm or in the private realm—the basic idea being that the information itself can be moved from the public realm to the private realm without having an effect on the information itself; the same information can exist in the public realm and in the private realm. Rubel and Biava (2014, 2424) are also critical of the control approach to informational privacy and suggest that informational privacy is better understood as a relation between three entities: "some person or persons, some domain of information, and some other person or persons." The point is that nothing is said beforehand about the nature of these relations, merely that those are the relations to be analyzed and understood in a given privacy situation or context. The nature of the relations then defines the particular privacy issue at stake and makes it possible to compare various privacy situations.

Common for these—and other similar—conceptualizations is that they take a pragmatic approach and aim to give voice to the contextualization

of privacy issues, arguing that there are norms, specifics, or matters within a given situation that define and limit the privacy issues at play. However, they do not explicitly consider how the notion of information might be affected by their pragmatic, contextual, situational interpretation stance—they maintain the basic idea that information *just is*, that information is an entity that can be manipulated and moved between different spheres or domains.

Being Pragmatic

One route to conceptualize the role of information in informational privacy theories is to determine the "various types of information or matters that are private" (Solove 2008, 67) and which other types of information or matters are not private and to use this distinction to draw a line between privacy and non-privacy. A common approach is to draw this line based on the "intimacy or sensitivity" (ibid., 67) of the information in question; information that by nature is intimate or sensitive is personal information, while other types of information are not.

Solove (2008) suggests that a better approach is to look at the purposes for which the information is used to analyze the privacy issues involved; "information is public or private depending upon the purposes for which people want to conceal it and the uses that others might make of it" (ibid., 69). In other words, the focus ought to be on the nature of the purpose for which the information is applied. This is similar to Nissenbaum's (2010, 3) call for "contextual integrity" in which "finely calibrated systems of social norms, or rules, govern the flow of personal information."

Both Solove and Nissenbaum—as well as other privacy scholars—have been influenced by the idea that the particular situation or context is of importance to understanding the privacy issues at play. What is not discussed is the status or conceptualization of information; except for the use of notions such as sensitive, personal, intimate, or private about the information and the acknowledgement that it can take different forms (written, recorded, spoken, etc.), nothing is said about the nature of information involved in informational privacy. Solove and Nissenbaum seem to hold "the self constant" and ignore "the problem of the evolving subjectivity" (Cohen 2012, 20). That is, while they recognize the importance of the specific situation or context, their analyses are limited to understanding the

privacy situation as specific and unique and to acknowledging that it is a mistake to assume that information by nature is either private or public.

In this chapter I augment these contextual approaches to privacy by arguing that information itself is a contextual, situational, and pragmatic construct. As such, the present chapter builds upon and expands the work of Nissenbaum, Solove, and others who have advocated a contextual approach to privacy. This chapter augments their work by providing a contextual understanding of personal information.

The concept of information is tricky, and it is used in several different ways with various connotations. As Agre once noted,

> Computers are frequently said to store and transmit information. The term information, though, conceals a significant ambiguity. On one hand, information can be defined (as per Shannon and Weaver) as a purely mathematical measure of information and information-carrying capacity, without regard for the content. On the other hand, information is information also about something. (Agre 1994, 107)

In this sense, the scholars discussed above can be said to use a notion of information where information has an information-carrying capacity. What is of concern for these scholars is not so much the content of the information as the information itself; in such conceptualizations, information *just is*.

However, as Agre noted, information can also be conceptualized as information about something; as Westin (1967, 7) originally noted, privacy is the claim to determine the "extent information about them is communicated to others." To reuse Warren and Brandeis's (1890, 214) famous question—"What is the thing which is protected?"—is informational privacy concerned with protecting the stuff called "information" or with protecting the state of affairs that the information is about? The answer to that question will lead to different kinds of informational privacy theories. Therefore, we need to ask "What is information?" But first, let us pause to clarify the interrelation between information and personal information.

Information and Personal Information

As a way to avoid the larger philosophical discourse about the notion of information, one could suggest that information and personal information should be regarded as separate notions that are not related (figure 4.1). In

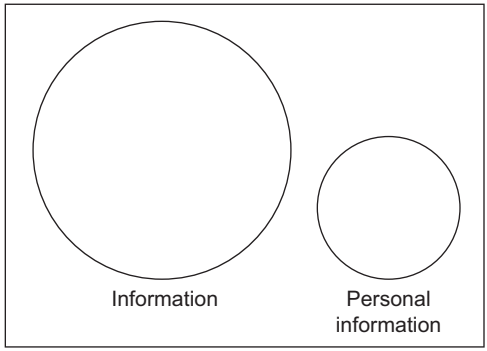

Figure 4.1

this sense, information and personal information are different in nature; they do not share characteristics or have properties in common. In such an understanding, information and personal information would be different concepts, and one would conceptualize personal information as a unique ontological unit. This understanding would allow one to define personal information as a specific set of ontological concepts that operate in specific and unique ways, independent of discussions of the notion of information as such. My sense is that this is not an attainable understanding and that personal information in many ways behaves in the same way as information per se and entails the same challenges. In this chapter, I will therefore regard information and personal information as the same conceptual unit.

We could consider personal information to share characteristics with information per se—and it could be argued that personal information is in fact just a small subset of information (figure 4.2). That is, we could single

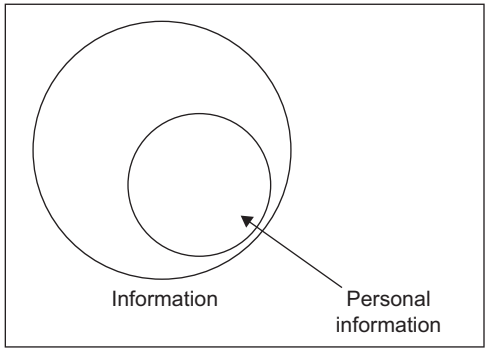

Figure 4.2

out personal information as a special kind of information—as a subset of information—defining it as "data about an individual that is identifiable to that individual—for example, my genetic code, my video preferences, my sexual preference, my credit history, my eye color, my income" (Murphy 1996, 2383) or as "any information relating to an identified or identifiable natural person" (Data Protection Working Party 2007, 4).

In this chapter, however, I will argue two things about this relation between information and personal information. First, the distinction is not as sharp as figures 4.1 and 4.2 might imply. As we will see in the next section of this chapter, the flow, use, and production of information in digital environments can blur the distinction between information about individuals (personal information) and information about all other things—in fact, if we were to maintain the distinction between information and personal information, there would probably be very little information in the digital information environment that would not be personal information (figure 4.3). As Cohen (2012) has noted, the idea of designating some information as "intimate" or "sensitive" and of specific interest in a privacy context is problematic considering today's data and information practices:

> Although privacy law purports to recognize a . . . principle, that . . . operates primarily to protect small islands of concededly "intimate" or "sensitive" information and correspondingly small enclaves of acknowledged physical seclusion. In an age of distributed information processing, moreover, even those islands are eroding. (Cohen 2012, 248)

In the algorithmic age, most information can be used in the construction of personal profiles—and thus the distinction between information and

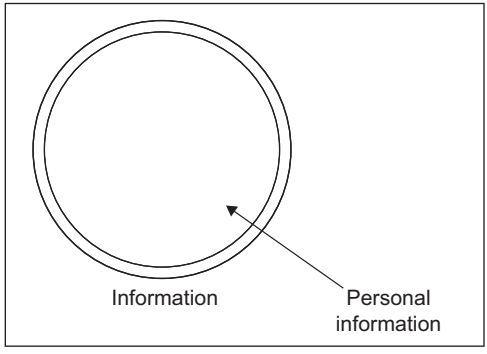

Figure 4.3

personal information becomes less clear, as lots of seemingly casual and informal information is used in the algorithmic creation of personal data profiles. There would, of course, still be some information that we would not regard as personal information, such as tree rings or fossils—until they are picked up by an individual and entered into a personal sphere at which point tree rings and fossils might in fact also become personal information, as pointed out by Van Hoboken (Joris van Hoboken, comments to author, January 12, 2017).

My second argument about the relations between information and personal information is that to understand the nature of information and appreciate the challenges of constructing a theory of informational privacy, we need to conceive of information as signs in a communicative process. What is interesting about personal information is that it is stuff that means *something about something*, and we ought to be concerned with the situations in which that meaning is produced. In other words, information is not best conceptualized as footprints of a state of affairs but as signs in a communicative process (Mai 2016b).

For a better understanding of informational privacy, we need a more granular understanding of the concept of information, as well as a more nuanced language to discuss it. With these in place, we can explore and articulate different approaches to what informational privacy is about and the avenues one might take to address current challenges. However, before digging into theories of information, let us first explore a few cases of information and data in the wild.

Three Examples of Datafication

I present three examples below in which there is an intersection between the collection, processing, and use of personal information and privacy. These examples will demonstrate how the control-oriented approach to privacy is too weak and insufficient in today's algorithmic age.

The first example is the application of algorithms in basically any and all digital platforms today, which has resulted in algorithms playing "an increasingly important role in selecting which information is considered relevant to us" (Gillespie 2014, 167). Such algorithms increasingly dominate our daily interactions with social media; they are an integral component of search engines, recommender systems, online newspapers, social

networking platforms, and so forth. The use of algorithms and the degree to which they "black box" (Pasquale 2015) information processes and decisions has been widely discussed. However, as Burrell (2016, 2) argues, "what is *new* in this domain is the more pervasive technologies and techniques of data collection, the more vast archives of personal data including purchasing activities, link clicks, and geospatial movement." Although these algorithms are already nestled "into people's daily lives and mundane information practices" (Gillespie 2014, 183), there is still very little known about the "ways in which users know and perceive that algorithms are part of their 'media life'" (Bucher 2017, 31). One challenge, of course, is that people rarely understand—*because it is black boxed*—which information about them has been collected and how that information is processed and used to generate the personalized services offered by platforms. As Gillespie (2014, 173–174) has noted,

> The most knowable information (geo-location, computing platform, profile information, friends, status updates, links followed on the site, time on the site, activity on other sites that host "like" buttons or cookies) is a rendering of that user, a "digital dossier" or "algorithmic identity" that is imperfect but sufficient. What is less legible or cannot be known about users falls away or is bluntly approximated.

Sophisticated Internet users might be able to determine that personal information is being collected in their daily interaction with various digital media; many people are aware that whatever data and information they themselves provide is being collected, and that hardware information about their devices is collected as well. But, as Gillespie notes, that is rarely enough information to provide detailed personalized services—much is approximated in the construction of personal profiles. Bucher (2017, 34) gives the example of Shannon, a professional career counsellor in her forties, who has blogged about Taylor Swift and then starts receiving "Facebook ads for products that younger people might like." The reason for this might be that Facebook's algorithm has recomputed her age given the information about Taylor Swift and makes the assumption that Shannon is younger than she actually is. Shannon is rather relaxed about this and finds that it is "rather amusing" though she often finds "Facebook ads to be 'slightly offensive as they make assumptions about me, which I don't like to think are true'" (ibid., 34).

Bucher's example is a good illustration of Cohen's (2012, 26) point that "raw information . . . is not terribly useful to anyone"; the fact that Shannon

has blogged about Taylor Swift says very little about Shannon without an understanding of the contextual situation in which Shannon has generated this information. The meaning of the information that is entered into the algorithms and personal data profiles, therefore, is best understood within the context of the "origins, purposes, and effects of [the] socially situated processes of sorting and categorization" (ibid., 26). The challenge, as Burrell (2016, 1) has pointed out, is that "the inputs themselves may be entirely unknown or known only partially"; once the information has been collected, it has been separated from the context in which it was produced and now interpreted and used in new contexts and therefore "is both already desiccated and persistently messy" (Gillespie 2014, 170).

The second example I will present here is the widely used but very illustrative Target case. The Target case illustrates the point that the distinction between ordinary, everyday information and sensitive information is not as clear-cut as sometimes imagined. The case was first reported in the *New York Times* by Charles Duhigg (2012), who told the story of a father who went to a Target store complaining that his daughter had received coupons for maternity clothing and baby products. Like many other chain stores, Target assigns each customer a unique "Guest ID number." This allows Target to identify and track each customer, provide customers with personalized services, and offer customers coupons that are tailored to their shopping interests. One set of customers of interest are pregnant women. Through research and data mining, Target assigns each customer a "pregnancy prediction" score, representing the likelihood that the customer is pregnant. Target is even able to calculate an expected due date. Target had found that they could assign the pregnancy score fairly precisely by analyzing the purchasing history for approximately twenty-five products, such as "quantities of unscented lotion," "supplements like calcium, magnesium and zinc," "scent-free soap," "extra-big bags of cotton balls," "hand sanitizers," and "washcloths" (Duhigg 2012). It is valuable for Target to know that a customer might be pregnant because this is a "watershed moment for couples" (Mayer-Schönberger and Cukier 2013, 57) in which they might change their shopping habits and develop new brand loyalties.

It could be argued that the daughter had control of the information regarding her purchases of the twenty-five products used to calculate her pregnancy prediction score that she provided to Target through her membership card, and that she benefitted from the transaction through coupons

and discounts. The challenge is that it is not misuse of personal information that in itself creates a privacy problem; rather, the issue is that Target is able to produce sensitive personal information through predictive analysis of the young girl's purchasing history. As Inness (1992, 58) argues, "[I]t is the *intimacy* of this information that identifies a loss of privacy." The fact that Target knows a customer's purchase history for, say, washcloths might not feel like a violation of their privacy, especially if the information is traded for a discount on washcloths. However, most people would probably consider knowledge about their health—such as whether or not they are pregnant—too personal and intimate to share with Target. This example demonstrates that in the age of big data, predictive analysis, algorithms, and machine learning, the challenges to informational privacy are less about the control of personal information and more about what companies know about their customers. The focus has "shifted from concerns about revealing information about oneself to others to concerns about the new insights that others can generate based on the already available data" (Mai 2016a, 199).

The last example I will present here is Cheney-Lippold's (2011) Quancast case. Quancast builds profiles of Internet users based on their interactions with a range of sites and platforms. These profiles are constructed based on the users' seemingly meaningless data collected as they go about their daily activities; the data about these interactions flows "into rigid database fields as part of the subsumption implicit in data mining" (ibid., 169). One element in the process is the establishment of a user's gender:

> As a user travels across these networks, algorithms can topologically striate her surfing data, allocating certain web artifacts into particular, algorithmically-defined categories like gender. The fact that user X visits the web site CNN.com might suggest that X could be categorized as male. And additional data could then buttress or resignify how X is categorized. As X visits more sites like CNN.com, X's maleness is statistically reinforced, adding confidence to the measure that X may be male. As X visits more sites that are unlike CNN.com, X's maleness might be put into question or potentially resignified to another gender identity. (Cheney-Lippold 2011, 169–170)

The data collected about an individual user is analyzed and checked against other statistical analyses to ultimately establish the gender of the user. However, as new data is collected and analyzed, the gender of the user might change, and, as such, "gender becomes a vector" (ibid., 170) that is

de-essentialized and constructed on a purely digital basis merely to develop marketing information about the users, which Quancast can then sell. The construction of gender within the context of Quancast becomes decontextualized from the social and cultural construction of gender; for Quancast the category of "gender" is merely a category "embedded in the logic of consumption" (ibid., 171) and is divorced from real, actual, and lived gender experiences.

Cheney-Lippold's (2011) Quancast example shows that data is collected to establish personal information about an individual, for example, "X is male" (ibid., 170). This establishment of maleness, however, has little to do with X as a human being; it is merely for marketing purposes. X's gender is established through predictive and inferential analyses of X's interactions with various sites and platforms, and X's gender is established through algorithmic analyses based on information that might not in the first place be regarded as personal information.

A common question raised for these three examples is whether they involve the notion of privacy. It is my sense that we could explore and read the cases from the perspective of privacy and discuss in which ways and to what extent the individuals' personal information was used in improper ways, and whether they had control over their personal information in the situations. I think most of us would agree that there is at least some degree of privacy involved, and that the individuals involved are at risk at least to some degree, given specific interpretations of the cases and the data practices involved. However, I would submit that we cannot meaningfully understand and analyze the privacy issues at play in these scenarios using the framework of control of personal information. In fact, I believe applying a control approach would cause us to misread the privacy concerns at stake in these examples for two reasons. The first reason is that the three cases demonstrate that the issue here has less to do with the data provided by the users/customers, and everything to do with the new insights and information the companies produce about them. The users/customers have no direct control over this new information and will, in most circumstances, not know that the information even exists. On the other hand, the fact that such information is produced and exists currently provides a flourishing personal data economy, as outlined by Zuboff in the first chapter of this volume.

The second reason for this potential misreading of the privacy concerns in the examples is that we have mistakenly based the idea of informational privacy on a specific understanding of information. When informational privacy is reduced to the challenge of controlling information, then it necessarily objectifies information, turning it into a reified entity that can be managed, that can be manipulated, and to which we can restrict access. If we regard information as a sign, as *something about something*, then informational privacy becomes concerned with the fact that something is known about someone and how we might act in such a situation. Information then becomes less central to informational privacy, because we cannot control and regulate the interpretation and production of meaning. Rather we must "control" the situations in which meaning is ascribed: the use of data.

At this point we need to return to and address the basic question for this exploration of the conceptual foundation of informational privacy, namely, *what is information?*

Information as Signs

One way to answer the question "What is information?" is to list all the things that one considers to be information: books, numbers, addresses, health records, phone bills, names, DNA codes, computer programs, weblogs, credit card statements, purchasing records, e-mails, likes on Facebook, and so on—a list of everything in the world that we consider to be informative. Such a list might end up containing almost everything in the world: cows, cups, and coffee are quite informative, but "if anything is, or might be, informative, then everything is, or might well be, information" (Buckland 1991, 356). This "extensional approach" (Furner 2016, 289) to determining the answer to the question of what is information has been labeled "information realism" (Fox 1983, 17) in the sense that it is an ontological quest for finding and listing the things in the world that are informative and therefore are information.

Another approach to understanding the nature of information is to adopt what Furner (2016, 289) calls an "intentional approach," in which the "properties that something must have" (ibid., 289) to be treated as information are identified. These might be that the information must be true to be information, that it can be used to generate knowledge, that it

can be used as a vehicle for communication, that it represents some state of affairs, or some combinations of such sets of properties. One dominant approach along these lines is advocated by Floridi (2008, 2010), who has developed a philosophy of information for "fields that treat data and information as reified entities, that is, stuff that can be manipulated (consider, for example, the now common expressions 'data mining' and 'information management')" (Floridi 2010, 20). Floridi defines information as "meaningful independent of an informee" (ibid., 22) and thereby constructs a notion of information that does not rely on a knowing subject. Dretske (2008) takes a similar approach to understanding information; he suggests that information "is independent of what we think or believe. It is independent of what we know" (ibid., 31) because information is "answers to questions" (ibid., 29) and "not just any answers . . . [but] true answers" (ibid., 29). Information cannot simply be understood in terms of meaning—as in what the information is *about*—because, as he explains, "[M]eaning is fine. You can't have truth without it" but "information, unlike meaning, has to be true" (ibid., 29).

In this sense information exists before human activities, language, and thought. It is true, and it has no agency. Cohen (2012, 20) has criticized these approaches to information within the context of informational privacy for accepting "liberal individualism's commitments to immateriality and disembodiment" and, as such, creating a construct of information that "appears to be the ultimate disembodied good, yielding itself seamlessly to abstract, rational analysis" (ibid., 20). These conceptualizations of "information" follow a tradition of conceptualizing "information" as what has been called "natural information," following Grice's (1957, 1989) distinction between natural and nonnatural meaning.

Natural and Nonnatural

Grice uses "natural" and "nonnatural" meaning to distinguish between statements that entail their meaning and statements that mean through conventions. As an example, Grice (1957, 213) uses the statement "Those spots mean (meant) measles" to illustrate statements that entail meaning; if someone utters that sentence, we would rightly expect that there is an actual correlation to a state of affairs in which someone has the measles. In other words, one cannot say, "Those spots meant measles, but he hadn't got measles" (ibid., 213); if someone has these particular spots, then he or she

has measles. There is no room for uncertainty and debate. Grice's notion of natural meaning is a notion of meaning where a statement entails a given state of affairs; it is true and objectively correct.

If we regard information to be natural, as in natural meaning, then information is regarded as a state of affairs. In this approach, information is viewed as that which reflects what is actual and true—and, as such, information can be manipulated and analyzed to gain knowledge about people's actual affairs, interests, and intentions. For people to enjoy freedom from surveillance and enjoy privacy, they should merely be able to control the flow of the material representations of information. If one controls when, how, and to what extent others have access to the material information, then one enjoys privacy.

Under this tradition, the fact that Shannon writes about Taylor Swift *entails* a particular age, the fact that a girl shops for particular products at Target *entails* that she is pregnant, and the fact that an Internet user visits certain web sites *entails* a particular gender.

Contrast that with the example of nonnatural meaning, which Grice (1957, 214) gives as "Those three rings on the bell (of the bus) mean that the bus is full." This utterance does not in the same way *entail* a particular state of affairs. One could very reasonably, "go on and say, 'But it isn't in fact full—the conductor made a mistake'" (ibid., 214). In this instance, someone has the intention of communicating something; there is a human agent present who can be correct or incorrect in his or her understanding of the actual state of affairs. There is no direct entailment between the statement and the fact that the bus is or is not full. The meaning of the statement is based on conventions—we have agreed that three rings mean that the bus is full. There is room for uncertainty, debate, and interpretation.

Grice's concept of nonnatural meaning is those situations in which meaning emerges out of use and in context. It is the type of meaning that is based on conventions and intentions; a statement means something because we have agreed on the correct usage of that statement. Similarly, with information, information is created and used in particular contexts and situations, and we can only understand information within those boundaries. In other words, to assign meaning and significance to Shannon writing about Taylor Swift, a girl shopping for particular products at Target, and an Internet user visiting CNN.com outside those particular contexts is, to use Gillespie's (2014, 174) words, "bluntly approximated." To

suggest that these people ought to be able to protect their informational privacy by using their abilities to control their personal information would be a mistake.

Within philosophy of information, Grice's work on the distinction between natural and nonnatural meaning is used to develop the notions of natural information and nonnatural information (Søe 2016). While natural information only allows for true, objectively correct information, nonnatural information allows for mistakes, misinterpretations, and misuse—and hence opens the way for concepts such as (personal) misinformation, that is, unintended misleading information, and (personal) disinformation, that is, information which is intentionally misleading (Søe 2016). Following Grice's conceptualization of meaning, and his grounding in semiotic thinking, we could articulate the core properties of information by stating that it functions as a vehicle used in the production and exchange of meaning. That is, information is a sign (Mai 2013).

If we regard information as a sign, and personal information as nonnatural information—then we arrive at a significantly different understanding of informational privacy. Information gains meaning through interpretations with specific contexts, situations, and usages. In this conceptualization, informational privacy cannot be reduced to the flow and control of information but must be understood in the sense of what is known and knowable about the individuals in question.

However, what is known and knowable about individuals is always constructed at an interpretive distance and is always one take on the meaning of what is being said and done. So, let us now consider practices that involve the production and use of information about individuals.

Situating Informational Practices

Big data has been associated with a particular ideology or belief system about the world, that is, specific understandings of how humans operate in the world and how people make sense of the world and communicate—this ideology can be labeled "data behaviorism" (Rouvroy 2013, 143) or "dataism" (van Dijck 2014, 198). Common for much work in big data is the belief in "the objective quantification and potential tracking of all kinds of human behavior and sociality through online media technologies" (ibid., 198), and "prediction is the hallmark" (Ekbia et al. 2015, 1529) for work in big data. However, in practice, it may be difficult to attain the goal of

objectivity and disinterestedness, because "big data does not arrive in the hands of analysts ready for analysis" (ibid., 1531); data has to be cleaned or conditioned to be usable, which involves deciding which attributes and variables to keep and which to ignore. In fact, in a study of working data scientists, Carter and Sholler found that,

> disinterest may not always be possible, especially for workers in business settings that involve contact with a client. Analysts might aspire to objectivity but be forced by circumstances to recognize their own positioning and the role of communication in data analysis. (Carter and Sholler 2016, 2317)

So, even if "dataism presumes *trust* in the objectivity of quantified methods" (van Dijck 2014, 204), in reality, this trust does not hold up.

The drive behind big data is a belief in the decontextualization of data—that data has meaning beyond particular situations and that more data will lead to more meaning and better understanding. The hope is that ultimately, we can do away with theories, perspectives, and interpretation. As Chris Anderson (2008) wrote, in a small piece in *Wired* where, even before big data was in vogue, he foresaw a future where it would be possible to harvest massive amounts of data, "With enough data, numbers would speak for themselves" (Anderson 2008). This is a hype, of course, but a dangerous hype that "leads to the withering away of interpretation—not through the actions of a cabal, but through a sociologic excluding from the archive all data which is not big" (Bowker 2014, 1797) and which, "thanks to relatively simple algorithms allowing, on a purely inductive statistic basis, to build models of behaviors or patterns, without having to consider either causes or intentions" (Rouvroy 2013, 143).

The challenge is that both the individuals interacting with platforms producing personal information and the analysts trying to make sense of their data are "real embodied people [who] do not experience 'information' in the abstract; [but] rather . . . through the lens of embodied perception" (Cohen 2012, 33). When people interact with digital platforms and write about Taylor Swift, shop at Target, or surf various websites, they produce data. This data can be used to infer something about them: that they are of a certain age, that they are pregnant, or that they are a certain gender. These are categories. Big data analysis, machine learning, and predictive analysis are ways to categorize—techniques to place people, objects, and phenomena into categories. Until recently, categories were thought to be unproblematic containers:

> They were assumed to be abstract containers, with things either inside or outside the category. Things were assumed to be in the same category if and only if they had certain properties in common. And the properties they had in common were taken as defining the category. (Lakoff 1987, 6)

However, scholarship across the social sciences and humanities as it has played out at least since the *linguistic turn* in the middle of the twentieth century (Rorty 1967) has found that categories, meaning, and understanding are embodied experiences. It is now widely accepted that categories are constructions; categories are not objectively *in* the world. They are constructed to make sense of and explain the world.

Even for simple concepts or categories—such as "dog"—there is no objective and universal meaning or sense. A dog is a *pet*; it is sometimes *owned* by people as their *property*; it is a *mammal*, a *guard*, a *dish*; it sometimes lives in *houses* and other times in *herds*; and it may even be a *show dog*. The notion of dog does not have a life or a meaning per se. There is nothing that is *the* concept dog or *the* properties that define dog:

> Our concept of the dog involves a lot more than a list of dog-like properties—it involves knowledge of how the dog operates in the world and how it is related to other things in that world. Concepts are not isolated entities. In order to grasp a concept, we require not only definitional, but also encyclopedic knowledge. (Bryant 2000, 59)

To understand the notion of dog, we need to be part of the usage and situations in which the concept is used. If we do not know the concept, we might look it up in a dictionary, but "'dictionary words' . . . must be defined in terms of other dictionary words" (Eco 1984, 50). Dictionaries and dictionary-like definitions require encyclopedic knowledge of the culture, the context, and the languages in which they are used.

There is no such thing as *raw* data, just as there is no such thing as *the* meaning of a word: "*The meaning of a word is its use in the language*" (Wittgenstein 1958, §43). Similarly, with categories, we understand the world not merely through individual concepts "but also in terms of *categories* of things," and if we change those categories, "we change our understanding of the world" (Lakoff 1987 9).

We might attribute meaning and properties to the facts that Shannon *writes about* Taylor Swift, that a young girl who *is shopping* for particular products at Target is pregnant, or that *people create* particular patterns of surf history as they visit various websites. However, that meaning and those

properties will only be one of many possible interpretations, one particular take on what the data could show or tell us about the people who engaged in activities and thereby produced data. Hence, personal information is *signs* open for interpretation, analysis, and (mis)use.

Conclusion

It is commonly suggested that informational privacy seeks to protect people's personal information, the basic idea being that people's ability to enjoy privacy is tied to their ability to control the flow of their personal information. However, the notion of personal information has remained largely unmarked and undertheorized in the privacy literature, leaving behind an understanding of personal information in which information *just is*.

Privacy scholars have tended to focus on *informational control* as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (Westin 1967, 7)—when in fact the core challenge is the processes by which personal information is created. If one focuses on the term "information," one might start to list all the things we consider personal information and then argue that people have property rights over those things, that they ought to be able to control the flow of those things, or that they should give consent for the use of such things—and then believe that once those property rights, control, or consent issues are in place, then people can enjoy privacy. However, the challenge that privacy ought to be concerned with is the situation in which meaning is created; thus we must ask, "What does it mean that others have information *about* individuals, groups, or institutions? And when does a certain use of information constitute a privacy intrusion?"

Digital platforms and services utilize big data analyses, predictive analysis, algorithms, and machine learning to produce (personal) information about individuals. The next major challenge for informational privacy theory is thus to develop a foundation that recognizes the complex and often opaque processes by which that (personal) information is produced. As Van Hoboken argues in this volume, this might lead to the position that regulation should turn away from regulating the *collection* of personal information altogether and instead regulate the *use* of (personal) information. In fact, people might only be able to enjoy informational privacy once digital

platforms and services' business practices are based on ethical foundations in which their practices are truly virtuous. Privacy cannot be limited solely to an individual, liberal right but should be expanded to an expectation of how society allows individuals to be treated by companies. The next generation of informational privacy theory should establish and make explicit the internal goods, norms, and standards of the algorithmic age, which should include as necessary components "justice, courage, and honesty" (MacIntyre 1981, 191).

References

Agre, Philip E. 1994. "Surveillance and Capture: Two Models of Privacy." *The Information Society* 10 (2): 101–127.

Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired* 16 (7).

Bowker, Geoffrey C. 2014. "The Theory/Data Thing." *International Journal of Communication* 8 (2043): 1795–1799.

Bucher, Taina. 2017. "The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms." *Information, Communication & Society* 20 (1): 30–44.

Buckland, Michael. 1991. "Information as Thing." *Journal of the American Society for Information Science* 42 (5): 351–360.

Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding the Opacity of Machine Learning Algorithms." *Big Data & Society* January–June 2016: 1–12.

Bryant, Rebecca. 2000. *Discovery and Decision: Exploring the Metaphysics and Epistemology of Scientific Classification*. Cranbury, NJ: Associated University Presses.

Carter, Daniel, and Dan Sholler. 2016. "Data Science on the Ground: Hype, Criticism, and Everyday Work." *Journal of the Association for Information Science and Technology* 67 (10): 2309–2319.

Cheney-Lippold, John. 2011. "A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control." *Theory, Culture & Society* 28(6): 164–81.

Cohen, Julia E. 2012. *Configuring the Networked Self: Law, Code, and the Play of Everyday Practices*. New Haven, CT: Yale University Press.

Data Protection Working Party. 2007. *Opinion 4/2007 on the Concept of Personal Data*. Article 29, WP 136. Brussels: European Commission.

Dretske, Fred. 2008. "Epistemology and Information." In *Handbook of the Philosophy of Science, Philosophy of Information*, vol. 8, edited by Pieter Adriaans and Johan van Benthem, 29–47. Amsterdam: Elsevier.

Duhigg, Charles. 2012. "How Companies Learn Your Secrets." *New York Times*, February 16, 2012. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

Eco, Umberto. 1984. *Semiotics and the Philosophy of Language*. London: Macmillan.

Ekbia, Hamid, Michael Mattioli, Inna Kouper, G. Arave, Ali Ghazinejad, Timothy Bowman, Venkata Ratandeep Suri, et al. 2015. "Big Data, Bigger Dilemmas: A Critical Review." *Journal of the Association for Information Science and Technology* 66 (8): 1523–1545.

Fiske, John. 2011. *Introduction to Communication Studies*. 3rd ed. London: Routledge.

Floridi, Luciano. 2008. "Trends in the Philosophy of Information." In *Handbook of the Philosophy of Science, Philosophy of Information,* vol. 8, edited by Pieter Adriaans and Johan van Benthem, 113–131. Amsterdam: Elsevier.

———. 2010. *Information: A Very Short Introduction*. Oxford: Oxford University Press.

Fox, Christopher J. 1983. *Information and Misinformation: An Investigation of the Notions of Information, Misinformation, Informing, and Misinforming*. Westport, CT: Greenwood Press.

Furner, Jonathan. 2016. "'Data': The Data." In *Information Cultures in the Digital Age*, edited by Matthew Kelly and Jared Bielby, 287–306. London: Springer.

Gillespie, Tarleton. 2014. "The Relevance of Algorithms." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, 167–193. Cambridge, MA: MIT Press.

Grice, H. Paul. 1957. "Meaning." In *Studies in the Way of Words*, edited by H. Paul Grice, 213–223. Cambridge, MA: Harvard University Press.

———. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Inness, Julie C. 1992. *Privacy, Intimacy, and Isolation*. New York: Oxford University Press.

Lakoff, George. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

MacIntyre, Alasdair. 1981. *After Virtue: A Study in Moral Theory*. Notre Dame, IN: University of Notre Dame Press.

Mai, Jens-Erik. 2013. "The Quality and Qualities of Information." *Journal of the American Society for Information Science and Technology* 64 (4): 675–688.

———. 2016a. "Big Data Privacy: The Datafication of Personal Information." *The Information Society* 32 (3): 192–199.

———. 2016b. "Personal Information as Communicative Acts." *Ethics and Information Technology* 18 (1): 51–57.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.

Moore, Adam D. 2010. *Privacy Rights: Moral and Legal Foundations*. University Park: Pennsylvania State University Press.

Murphy, Richard S. 1996. "Property Rights in Personal Information: An Economic Defense of Privacy." *Georgetown Law Journal* (83): 2381–2417.

Nissenbaum, Helen. 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford University Press.

Pasquale, Frank. 2015. *Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Rorty, Richard, ed. 1967. *The Linguistic Turn: Recent Essays in Philosophical Method*. Chicago: University of Chicago Press.

Rouvroy, Antoinette. 2013. "The End(s) of Critique: Data Behaviourism versus Due Process." In *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, edited by Mireille Hildebrandt and Katja de Vries, 143–167. Oxon, UK: Routledge.

Rubel, Alan, and Ryan Biava. 2014. "A Framework for Analyzing and Comparing Privacy States." *Journal of the Association for Information Science and Technology* 65 (12): 2422–2431.

Solove, Daniel J. 2008. *Understanding Privacy*. Cambridge, MA: Harvard University Press.

Stonier, Tom. 1990. *Information and the Internal Structure of the Universe: An Exploration into Information Physics*. London: Springer.

Søe, Sille Obelitz. 2016. "The Urge to Detect, the Need to Clarify: Gricean Perspectives on Information, Misinformation and Disinformation." PhD diss., University of Copenhagen, Faculty of Humanities.

Tavani, Herman T. 2008. "Informational Privacy: Concepts, Theories, and Controversies." In *The Handbook of Information and Computer Ethics*, edited by Kenneth Einar Himma and Herman T. Tavani, 131–164. Hoboken, NJ: Wiley.

Van Dijck, José. 2014. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance & Society* 12 (2): 197–208.

Warren, Samuel D., and Louis D. Brandeis. 1890. "The Right to Privacy." In *Information Ethics: Privacy, Property, and Power*, edited by Adam D. Moore, 209–225. Seattle, WA: University of Washington Press.

Westin, Alan F. 1967. *Privacy and Freedom*. New York: Atheneum.

Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. New York: Macmillan.

HUMAN RIGHTS IN THE AGE OF PLATFORMS

EDITED BY RIKKE FRANK JØRGENSEN

FOREWORD BY DAVID KAYE

Human Rights in the Age of Platforms

Information Policy Series

Edited by Sandra Braman

A complete list of the books in the Information Policy series appears at the back of this book.

Human Rights in the Age of Platforms

Edited by Rikke Frank Jørgensen

Foreword by David Kaye

The MIT Press
Cambridge, Massachusetts
London, England

The Open Access edition of this book was published with generous support from Knowledge Unlatched and the Danish Council for Independent Research.

Knowledge Unlatched

This book was set in Stone Serif and Stone Sans by Jen Jackowitz. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Jørgensen, Rikke Frank, editor.
Title: Human rights in the age of platforms / edited by Rikke Frank Jørgensen.
Description: Cambridge, MA : The MIT Press, [2019] | Series: Information policy | Includes bibliographical references and index.
Identifiers: LCCN 2018049349 | ISBN 9780262039055 (hardcover : alk. paper)
Subjects: LCSH: Human rights. | Information society. | Information technology--Moral and ethical aspects.
Classification: LCC JC571 .H7695266 2019 | DDC 323--dc23 LC record available at https://lccn.loc.gov/2018049349

10 9 8 7 6 5 4 3 2 1

Contents