DATA, PRIVACY, AND THE INDIVIDUAL

PRIVATE DATA AND PROPERTY

VERENA RISSE
TU DORTMUND UNIVERSITY

NOVEMBER 2019

PRIVATE DATA AND PROPERTY

Verena Risse

TU Dortmund University

ABSTRACT:

This contribution explores questions related to the protection and governance of private data by drawing on the analogy between privacy and property. Although this analogy is not entirely new, property and the defense of property rights remain one of the most common ways in which rights to private data are discussed in both legal and philosophical debates. Accordingly, the paper starts out by presenting theories of property generation and acquisition and adapts these to the generation and collection of private data. Similarly to property, data as such does not independently exist; it is something created by human beings. Data exists because it is measured, gathered and organized. When it comes to the acquisition of personal data, however, one cannot assume—as in the case of property—that the person who gathers data automatically owns it. Thus, it seems important to note that—unlike the case of property—generation and acquisition of data do not automatically go hand in hand. Yet, once data exists, questions of how it is protected, acquired and transmitted remain and property rights provide a helpful starting point in searching for answers.

Despite showing some conceptual similarities with property, the administration of data presents specific problems that deserve attention. The first is the problem of how to deal with the contents of privacy being a black box. In other words, how can we designate something that needs protection when respecting it entails not knowing what is inside? Unlike private property such as a house or an apartment, privacy has no natural contours or borders. Hence, it is not sufficient to refer to concepts like "solitude" or "private affairs" to designate data worthy of protection. This paper therefore sets out to discuss different options of how these borders can be drawn, ranging from generally listing classes of private data to consent-based approaches.

Second, it seems important to clarify the different roles and interests of individual, economic and public actors with regard to the acquisition and administration of private data. Whereas property can be freely traded between all three kinds of actors, who are equal *vis-à-vis* the two others, it is not clear whether this equality exists in the case of private data. Rather, it seems that individuals have a lower interest in the acquisition of data, whereas their own data is of high interest for both economic and public actors. This issue is discussed in reference to autonomy. Generally, one may assume that by acquiring additional material and financial resources or information, the owner increases her

possible courses of action and—if these add up to a reasonable number of choices—also her autonomy. This analogy has its limits in the case of private data. While the acquisition of information empowers a person to make better decisions, the acquisition of personal data does so too, but usually in a more direct way (e.g., by acquiring body data to improve one's health), or by way of learning from the results of a large study, for which the data is rarely collected and processed by private individuals.

In sum, the analogy between property and privacy allows not only to identify potential problems and tensions in the governance of private data, but also offers tools and arguments that help clarify what is distinct about both concepts.

Reference to this paper should be made as follows:
Risse, V. (2019): "Private Data and Property", *Data, Privacy and the Individual*.

INTRODUCTION

This contribution explores questions regarding the protection and administration of private data by drawing on the analogy between data and property. Property and the defense of property rights remain one of the most likely ways in which rights to private data are discussed in both legal and philosophical debates. Recent statements by the European Commissioner for Competition, Margrethe Vestager, provide an illustration. At the Web Summit, held in Lisbon in November 2018, Vestager stated: "So it's clear that our data belongs to us—and that we have the right to control what happens to it." (Vestager, 2018)

This paper sets out to explore, first, why property is an attractive concept to tackle questions of data privacy. Second, two data-specific problems will be discussed, which set data significantly apart from property rights regimes. The first is the problem of how to deal with the fact that protecting private data can require creating a black box. This problem describes the need to designate something that needs protection, whereas respecting it entails not knowing what this something is about. The second is the question of how to deal with the stark heterogeneity among the concerned private, public and economic actors who show considerably different interests and degrees of power when it comes to the acquisition, administration, and use of private data.

PRIVACY AND PROPERTY—THE GROUNDWORK

Data is omnipresent in our world, not just since the information revolution. Rulers have always collected personal information about their subjects, and sellers have always sought to understand their customers by surveying them or by getting to know them personally. What is new is the extent to which information technology permits us to gather, process, store, and analyze large amounts of personal data. Often enough, individuals themselves supply data to other actors, be they economic, public, or private, by submitting information about themselves online or by filling out forms at diverse occasions. Corporations such as Google systematically pursue data mining activities, and governments request comprehensive information from their citizens and store it in ever-growing data bases. Against this background, it seems necessary to clarify two fundamental questions regarding the processes of data production and protection: first, who owns the data at what point in the process and, second, when and how should data be protected?

Handling private data by reference to property rights can proceed in two possible ways. Either one applies property rights directly to private data, or one assumes that data is a

concept *sui generis* for whose administration property rights can still offer important insights. Both approaches rely on the assumption that the administration of private data and private property are sufficiently similar. This first section of this paper considers the most important similarities between private data and (conventional forms of) property in three different respects: first, regarding their origin and acquisition, second, regarding their character and, third, regarding their governance.

Property theories classically start out with a genealogy of property (Waldron, 2016). This means that they deal with the question of how the institution of property comes into the world initially; they seek to answer the question of how one can make others believe that one owns something if the concept of ownership has yet to be introduced. John Locke famously solves this question by offering what is called the 'first occupancy' theory. According to this view, the person who first occupies a good or land becomes the initial owner (Locke, 1988 [1689], I, para. 86; II, para. 28). Competing views, such as the one offered by David Hume, consider that there is no such thing as an original possession, but that there is constant struggle for ownership, which can only be resolved by some form of agreement (Hume, 1978 [1739], p. 490).

Closely linked to this aspect of the genealogy of the concept of property is the question of how acquirable goods originate. In the most basic way, goods are either created or found (Waldron, 2016). Goods upon which ownership can be attributed emerge either because one person forms, produces, or in some way brings them into being, or, alternatively, because goods are found, which means that a person discovers, harvests, or quarries a resource that is growing or given in an (unappropriated) habitat.

Data—and more specifically, private data—shares relevant similarities with property when it comes to its origination. This is to say, data, too, comes into being as data because it is collected or retrieved. Data in a specific sense exists because it is saved and to some degree systematized. This also explains why data is a *plurale tantum*, as one datum without additional information or context is not telling as data. For example, the datum 'blue eyes' alone has little meaning, but three pairs of blue eyes in a total sample of five gathered in a Southern European country add up to data that might allow telling a story. This constraint does not imply, however, that data only becomes normatively relevant when it is manifold or otherwise marketable. Even the collection of one single datum can pose questions of permissibility insofar the future use and combination of this datum is made possible by way of its original collection and storage. Yet unlike some cases of private property in which the value of specific goods can decrease as soon as one owns a certain number of these goods (e.g., one does not need an infinite number of washing machines), data gets more meaningful the more of it is obtained.

Despite the similarities regarding their generation, it is less clear that private data and property keep following the same route when it comes to their appropriation. In the case of property, it is generally assumed that the person who creates or finds the good in question also becomes its owner. Famously, John Locke used the formulation that by way of 'mixing one's labor with' (Locke, 1988 [1689], II, para. 27; also see Becker, 1976) a good, the good is being appropriated. This means that, for instance, the person who digs out natural resources, carves a whistle, or writes a book becomes the owner of these goods.

Who is the natural owner of private data? One might assume that it is the person who the data is about. It could be argued that a person owns the right to information about herself and thus also owns the right to all data about herself. Yet, it could be the case that someone else collected or measured the data. If one applied the analogy of property acquisition to this case, one might wonder whether this fact turns the data collector into the rightful owner. This argument seems particularly strong if one is concerned with data a person does not even know about or with data that she could not have obtained herself. Still, ascribing the status of ownership to the data collector and not the data supplier creates certain uneasiness. The reason for this seems to be that one inherently assumes that one owns one's own body and that this ownership extends to all information around one's physique and identity. The relevance of our identity is also visible in the fact that one owns one's image even if the medium on which it appears (e.g., a video tape from a police CCTV) is owned by another actor.[1]

In addition to these basic considerations regarding origination and appropriation, several other features of property seem pertinent in the context of private data.

A feature that generally distinguishes property from other entitlements and which follows from the fact that it can be appropriated is that property is tradable—if ownership can be obtained upon a good, it can generally be assumed that this ownership is transferrable to other persons. Tradability also holds true for data: once it has been collected and the proper ownership has been established, data can generally be transferred. Yet, although the transfer is technically possible, this does not necessarily mean that it is always legitimate. Whether it is morally acceptable to transfer data is related to the potential disagreement regarding ownership described above between the data supplier and the data collector. For instance, one could think of cases in which the person providing the data has precluded any secondary use of her data. Thus, a person could have agreed to the use of her data for a specific procedure (e.g., solving a problem by turning to customer service), but not for any further uses (e.g., market research or customer service employee training).

[1] The author wishes to thank an anonymous reviewer for stressing this point.

Furthermore, the protection of private property through property rights entails certain privileges for the right-holder. First, it involves the entitlement to dispose upon the good. As an owner, one can handle one's property at one's discretion, which includes using, changing or even destroying the good in question. Second, this entitlement to dispose is generally considered to be exclusive. Other actors—be they public or private—have no say in the administration of the good.

If, on these grounds, one opts for a direct application of property rights to data, the following distinctions help classify data more precisely in terms of property.

The first distinction is between tangible and intangible property. Whereas tangible property denotes physical objects upon which ownership can be acquired, intangible property refers to goods that can be owned but that have no physical substance. A table and a house are examples for tangible property; examples for intangible property are knowledge, creations, patents, and copyrights. Data, according to this description, would be classified as intangible property.

Furthermore, one can distinguish rivalrous from nonrivalrous property. Rivalrous property can only be used by one person at a time. For instance, a bike or a book can only be used (i.e., ridden or read, respectively) by one person at a time. Nonrivalrous property, by contrast, can be simultaneously used by several persons without decreasing the value for each owner. Thus, a text (displayed on different media) can be read by several persons at a time. Accordingly, data constitutes a case of nonrivalrous property because the same datum can potentially be used or processed in different contexts or settings concurrently.

In sum, this first part established that property and private data are comparable in relevant ways. This comparability makes property a suitable concept to investigate questions regarding the generation, acquisition, and administration of data. Moreover, it is possible to treat data directly in terms of property, in which case data would be characterized as intangible and nonrivalrous property.

PROBLEMATIZATIONS

Having laid out why property is generally an interesting concept to look into when thinking about the protection and administration of data, it seems necessary to also point out that the two concepts differ in relevant ways. Therefore, in what follows, two specific problems relating to the governance of private data will be discussed. These are, first, the need to create a black box around data to protect it effectively and, second, the need to deal with the stark heterogeneity of involved actors.

1. Protecting the Content of a Black Box

The protection of private data differs significantly from that of property insofar as the protection of data can require not knowing about the content of the data and—ideally—not even knowing whether the data exists (or could exist) at all. As soon as it is known that personal data exists or could potentially exist in a given context, information about that person is already generally available. In this regard, data also differs from other forms of intangible property like creations, ideas, or inventions. In the case of the latter, it is desirable and necessary that they are known, because otherwise another person could claim ownership upon these immaterial goods.

One may thus ask whether there are adequate strategies to protect private data inside a black box so that the data supplier maintains control over her data while making visible only the black box and not the data that it contains.

One analogy that comes to mind when thinking about protecting data in a block box is that of a safe deposit box at a bank. The customer who rents a safe deposit box from a bank receives a box in which she can store valuables or other objects. She fills the box shielded from view and locks it with a key, which she keeps during the time of the lease. The bank possesses a master key with which a bank clerk can enter the space in which the safe deposit is stored. Thanks to this double security system only the customer knows what is inside the box, whereas the bank's knowledge only extends to who rented the safe deposit box. Renting safe deposit boxes differs from other banking activities in a significant respect: unlike money stored in a savings account, the owner can be sure to receive the exact content of the box at the end of the lease. Conversely, the stored items are not at the bank's disposal for further investments. The customer pays for the storage of her valuables.

The case of the safe deposit box is interesting for secure data storage because of the compartmentalization. By keeping one locked box inside another locked space with two separate keys, it is ensured that the content of the safe deposit box remains a secret to everyone except the owner. Creating a space that is secure and separate from other data repositories and to which only the data owner has access could therefore be a first step to respond to the need for a black box. Note, however, that the safe deposit box only works if there already is a good that can be placed inside the box. Therefore the example works less well for cases in which data has not yet been obtained or for cases in which one does not want some information to become data in the first place, according to the considerations about origination developed above.

A first strategy to respond to the need to keep private data inside a black box would be to designate areas in which privacy must be respected and to forbid all data collection within the limits of these areas. To this aim, it does not seem sufficient to refer to concepts such as 'solitude' or 'private affairs,' which would designate the relevant areas too loosely.

Instead, the designation would need to be more precise and subject-related, referring to, for instance, all data relating to health, sexual interests, etc. One can compare these designated areas with examples such as a diary or a locked bathroom door, which also denote spaces that must not be intruded upon. Yet, even in these seemingly obvious cases one can come up with counter-examples of situations in which an intrusion is warranted, both for the public good or the good of the data subject. For example, if someone has given detailed accounts of a planned murder or suicide in her diary entries, one could find good arguments that justify reading them. Also, one must bear in mind that new areas worthy of protection could always emerge, which have not yet been designated or whose contours are inadvertently vague. As a response to this problem, Adam Moore suggests to use John Rawls' conception of an 'original position' as an indicator for sensitive data areas (Moore, 2010, pp. 144-145; Rawls, 1999, pp. 118-123). In Rawls' thought experiment of the original position, persons are placed behind a 'veil of ignorance' so that they do not know socially-sensitive facts about themselves such as race, gender, age, profession, health status or the like. While Rawls employs the veil of ignorance to prevent people from designing principles of justice that would favor certain social groups over others, the device could serve as one indicator for protectable data areas as well. Asking which personal details would have to be concealed to protect a person from disadvantaged treatment can help uncover those categories of private data that make people particularly vulnerable and therefore deserve more sensitive treatment.

A second strategy to black box private data would be to assume that no personal data can be collected or processed, unless the person who the data is about has consented to it. The consent, which would usually be given by way of an agreement before the data has been collected, serves as a lock that prevents the black box from being opened. The general idea behind consent-based approaches is that a person's rights or entitlements are not violated when she has waived her rights. In other words, by consenting to the infringement, the intervention is justified—there is no violation of rights. To be valid, consent must be given fully, freely, and voluntarily—consent must be comprehensive, and must not be obtained under pressure or coercion. In practice, consent is often expressed in contracts or other forms of agreement. Thus, for example, if person A rents an apartment to person B, B does not violate A's property right upon the apartment when she moves in within the conditions of the lease. In the specific case of the black box of data, consent could either be given in general terms (for instance, for a topical set of data or for data obtained within a designated time at a designated place) or specifically for a given datum while leaving all other data unrevealed. A difficulty with the consent-based approach is that, in practice, the costs of obtaining consent can be high for all parties involved, depending on the strategy used to obtain consent (Posner, 1981). Obtaining consent is especially burdensome if specific and differentiated kinds of consent are required—as is often the case in the context of personal data.

A third strategy would be to anonymize data at the moment of its creation or collection. This would mean that all personal data is being black boxed insofar as individual details and the identifiability of individual persons are concerned. This strategy would leave open the possibility of collecting data in diverse contexts; that is, it is not necessary to exclude certain areas from the data collection. One problem with this strategy, however, is that it only works if anonymization is sufficiently robust—if relevant data are not attributable to specific persons. Another problem is that in some contexts, such as genetics, it is necessary to obtain non-anonymous data. Protecting such data would require that the box around anonymized data be as securely closed and separately stored as the bank safe considered at the beginning of this section. This requirement is supported by different investigations in which researchers have demonstrated that even when data has been anonymized, the combination of few publicly-available reference points can allow reidentification (Montjoye et al., 2015; Narayanan and Shmatikov, 2008).

Overall, it appears that no strategy can fully respond to the need for black boxing private data. It therefore seems advisable to opt for a combination of strategies to achieve the highest possible protection in relevant circumstances.

2. The Heterogeneity of Actors

Given the stark heterogeneity among individual, economic, and public actors with regard to the acquisition, accumulation, and usage of private data, it seems important to clarify the different roles and interests of these actors and their stakes in the issue of private data administration.

Private actors are individuals whose features or traces of action are being collected as data. Economic actors tend to collect data from private actors usually with the aim of selling their products. Public actors collect and process individuals' data for the purposes of administration and governance. Even this short and sketchy enumeration shows that there is an imbalance between the provision and the use of data among these different actors.[2] Most strikingly, private actors appear primarily as data suppliers, whereas economic and public actors appear primarily as data collectors and data processors. Of course, public actors will also be interested in the acquisition of data from economic actors or from other public actors, and the same is true for economic actors, respectively. Yet both are rarely in the role of sharing their data with private actors.

This discrepancy is not necessarily unjustifiable. One can, for example, think of cases in which public actors have a significant interest in obtaining private data. Most importantly, this is true in cases in which information acquired on the basis of private

[2] Note that this does not necessarily imply that there is an imbalance in the trade of goods (one of which can be data) between these actors.

data is used to avert harm to a large number of people or to prevent a threat to a country's infrastructure or state institutions (Moore, 2010, p. 151). A case in point would be the tracking of telephone calls or e-mails in order to prevent mass murder.

Despite being potentially justifiable in given circumstances, the described heterogeneity has significant normative implications, which set the case of data apart from the one of property. This can most clearly be demonstrated in reference to autonomy. In the case of property, when acquiring additional material and financial resources or information, the owner generally increases her possible courses of action and thus her autonomy.[3] Possessing more and having more means at one's disposal creates more options for action. Having a viable amount of options to choose from is an essential part of the idea of autonomy (Raz, 1986). This does not mean, however, that every additional option increases autonomy. Instead, the actor needs an *adequate* range of options at her disposal.

Moreover, this increase in autonomy by means of acquiring additional resources is particularly true and valuable for private as opposed to public or corporate actors. This is due to the conceptual core of autonomy, which centers on projecting one's own life plan and choosing one's actions in accordance with what one values (Raz, 1986, p. 155).[4] The capacities of exercising free will and choosing from an adequate range of options are important in order to fully benefit from autonomy. Being able to enjoy this degree of freedom, however, is specific to individuals. Public actors and economic agents have a more limited scope of will formation. As (public) institutions or legal entities, these actors do not possess the ability to choose their actions freely. Instead, their courses of action are defined according to their designated purpose as institutions or corporate entities (Waldron, 2011, pp. 324-326). This difference is underscored by the fact that only the liberties of individuals are protected in the most extensive form, namely as human rights.

Increasing autonomy through the accumulation of property does not work in the same way in the case of private data. While the acquisition of information empowers a person to make better decisions, the acquisition of personal data does so in a less comprehensive way. Instead, if private data is acquired or processed by private actors, this is usually done more directly and more targeted, for instance, by measuring body or fitness data to improve one's health. Collecting other actors' data—be they private, economic, or public—influences a private actor's decision-making set-up more indirectly. Private persons can benefit, for instance, from the findings of medical studies using large amounts of individual health data, but will most likely not process the data themselves and they might not have donated the relevant data. The opposite is true for economic or public actors. Acquiring data and, in particular, personal data from private persons

[3] As was stated above, saturation can be reached as soon as one possesses a certain number of specific goods. This does not contradict the fact that having more material and financial means at one's disposal generally increases one's available courses of action.

[4] Raz refers to autonomy as obtaining '(part) authorship of one's life.'

significantly helps them refine their courses of action, although it does not seem conceptually adequate to say that these institutional actors increase their autonomy. What is problematic, however, is that the collection and processing of personal data by public and economic actors can impede the autonomy of individuals. For example, by means of price discrimination or personalized advertisements customers are presented with a modified set of options and these modifications are in favor of a choice to the benefit of the economic actor. When individuals' autonomy is at stake, the impediment is in need of justification.

In sum, private, economic and public actors show significant differences regarding their interests and power when it comes to the generation and administration of data. Although being potentially justifiable, this heterogeneity can decrease the autonomy of individuals.

CONCLUSION

This paper discussed and evaluated whether property is a helpful concept to better understand and administer private data. In the first section, relevant similarities have been found among private data and property regarding generation, characterization, and governance. Although this suggests that property rights can be a helpful reference for the discussion of private data regimes, data and property do bear crucial differences that must be taken into account. Two of these differences were discussed in the second part of the paper, namely, the need to black box some data to achieve its full protection and the question of how to deal with the heterogeneity of involved actors and their respective preferences. Although the differences between data and property were significant, the reference to property rights proved a helpful framework for discussion and offered relevant tools for the analysis of private data.

Overall, property might not be the sole concept by which to assess questions of data protection and administration, but this paper revealed that it can still offer important insights and helpful starting points for an ongoing discussion.

REFERENCES

Becker, Lawrence (1976): 'The Labor Theory of Property Acquisition,' *The Journal of Philosophy* 73 (18), pp. 653-664.

Hume, David (1978 [1739]): *A Treatise of Human Nature*, Lewis A. Selby-Bigge and Peter H. Nidditch (eds.). Oxford: Clarendon Press.

Locke, John (1988 [1689]): *Two Treatises of Government*, Peter Laslett (ed.). Cambridge: Cambridge University Press.

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

De Montjoye, Yves-Alexandre, Laura Radaelli, Vivek Kumar Singh, and Alex "Sandy" Pentland (2015): 'Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata,' *Science* 347 (6221), pp. 536-539.

Narayanan, Arvind and Vitaly Shmatikov (2008): "Robust De-Anonymization of Large Sparse Datasets", *IEEE Symposium on Security and Privacy*, pp. 111-125.

Posner, Richard (1981): 'A Reply to Some Recent Criticisms of the Efficiency Theory of the Common Law,' *Hofstra Law Review* 9, pp. 775-794.

Rawls, John (1999): *A Theory of Justice* (rev. ed.). Cambridge, MA: Harvard University Press.

Raz, Joseph (1986): *The Morality of Freedom*. Oxford: Oxford University Press.

Vestager, Margrethe (2018): 'Building a fairer digital world.' Speech delivered at the Web Summit, Lisbon, 7 November 2018, URL: https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/building-fairer-digital-world_en.

Waldron, Jeremy (2011): 'Are Sovereigns Entitled to the Benefit of the International Rule of Law?,' *The European Journal of International Law* 22 (2), pp. 315-343.

Waldron, Jeremy (2016): 'Property and Ownership,' *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL: https://plato.stanford.edu/archives/win2016/ entries/property/.