ie | CENTER FOR THE GOVERNANCE OF CHANGE

Telefónica | DIGITAL POLICY LAB

# DATA, PRIVACY AND THE INDIVIDUAL

## PRIVACY MATTERS.

Carissa Véliz

MARCH 2020

# Contents

# Introduction

The first few years of the 21st century were characterised by a progressive loss of privacy. Two phenomena converged to give rise to the data economy: the realisation that data trails from users interacting with technology could be used to develop personalised advertising, and a concern for security that led authorities to use such personal data for the purposes of intelligence and policing.

In contrast to the early days of the data economy and internet surveillance, the last few years have witnessed a rising concern for privacy. As bad data practices have come to light, citizens are starting to understand the real cost of using online digital technologies. Two events stamped 2018 as a landmark year for privacy: the Cambridge Analytica scandal, and the implementation of the European Union's General Data Protection Regulation (GDPR). The former showed the extent to which personal data has been shared without data subjects' knowledge and consent and many times for unacceptable purposes, such as swaying elections. The latter inaugurated the beginning of robust data protection regulation in the digital age.

Getting privacy right is one of the biggest challenges of this new decade of the 21st century. The past year has shown that there is still much work to be done on privacy to tame the darkest aspects of the data economy. As data scandals continue to emerge, questions abound as to how to interpret and enforce regulation, how to design new and better laws, how to complement regulation with better ethics, and how to find technical solutions to data problems.

The aim of the research project *Data, Privacy, and the Individual* is to contribute to a better understanding of the ethics of privacy and of differential privacy. The outcomes of the project are seven research papers on privacy, a survey, and this final report, which summarises each research paper, and goes on to offer a set of reflections and recommendations to implement best practices regarding privacy.

# Why the Ethics of Privacy and Differential Privacy?

**Given the context of ubiquitous data collection coexisting with increasing concern for privacy and the regulation of personal data, privacy experts are faced with the practical challenges of, first, establishing what good data practices look like, and second, proposing ways of securing those high standards. To answer the first challenge, it is important to have a good grasp on the ethics of privacy: what is privacy, why is it important, what is at stake in its loss, and how does it relate to other goods and values?**

Ethics grounds good regulation and is an important complement to good laws. Good regulation is inspired in ethics: it is desirable to legalise what is morally right and ban what is gravely morally wrong. That is why companies with better ethical standards often are ahead of the law and have less trouble complying with future legislation. But laws are and should be limited. They establish minimal requirements for fairness. If we ban everything that is wrong, we risk becoming a police state in which all infractions are punished by law. Consider medical ethics. Not everything that is morally wrong (e.g., a doctor engaging in consensual sexual relations with their patients) is or should be illegal. Ethics goes beyond the law in that it identifies moral issues, analyses the kind of society that is most conducive to leading a good life, and makes recommendations accordingly. In a sense, ethics is more ambitious. Good laws lead to an orderly communal life and to fairness in society. Ethics, done well, promotes the wellbeing of all.

Once we have a sense of the ethics of privacy, we have to come up with reliable ways of protecting privacy. Among the many technical solutions proposed, differential privacy stands out as one that has garnered much attention and enthusiasm in the academic community. First put forward by Cynthia Dwork in 2006,[1] differential privacy aims to protect privacy while allowing researchers and companies to analyse sensitive data. Differential privacy adds precisely enough mathematical noise to the results of statistical queries from a database to mask the details of every individual in the database, but little enough to make sure that the results retain sufficient accuracy. Roughly, a differentially-private query mechanism is one in which you could subtract the data of any one individual without altering the answers that the database yields, and thus you could not infer any piece of sensitive information about any one individual.

What makes differential privacy so attractive is that it works irrespectively of how much information an attacker has. One of the privacy challenges of the digital age is that attackers can often get hold of multiple databases to infer information about people who are in them. Differential privacy's key guarantee is that any privacy breach that occurs as a result of combining databases could also have occurred without the results of differentially privacy queries.

Another advantage of differential-privacy is that it allows us to mathematically quantify the privacy loss, or, more precisely, the greatest possible information that an attacker could gain, through epsilons—the closer to zero epsilons, the stronger the privacy protection. Of particular interest to institutions and companies intent on preserving privacy is the possibility of differential privacy as a constraint on algorithms collecting data.


In other words, a differentially-private mechanism can be used at the time of data collection so that there is no such thing as an 'original' database with private information that could be compromised.

Establishing ethical data practices and developing technical ways of protecting sensitive data are two fundamental pillars to tackle the privacy challenges that we face. Scandals like that of Cambridge Analytica are already shaking the trust that people had placed in companies and institutions, and, even more concerning, democracy itself. Future data scandals will likely bring down companies, and they could challenge the legitimacy of democracies. The time is now to make sure we avoid such preventable disasters.

ESTABLISHING ETHICAL DATA PRACTICES AND DEVELOPING TECHNICAL WAYS OF PROTECTING SENSITIVE DATA ARE TWO FUNDAMENTAL PILLARS TO TACKLE THE PRIVACY CHALLENGES THAT WE FACE.

1 Cynthia Dwork, "Differential Privacy," in *International Colloquium on Automata, Languages, and Programming. Automata, Languages and Programming,* ed. Michele Bugliesi, et al. (Venice: Springer, 2006).

# PERSPECTIVES ON THE ETHICS OF PRIVACY AND DIFFERENTIAL PRIVACY

# Privacy

Kevin Macnish's (University of Twente) paper constitutes an introduction to how legal scholars and philosophers have thought about privacy in the past century.[2] The paper uses historical and contemporary examples to illustrate some of the most pressing ethical concerns regarding privacy.

Legal scholarship on privacy began in 1890, when United States judges Samuel Warren and Louis Brandeis published their article 'The Right to Privacy.' Warren and Brandeis were worried with how technology—the concern back then focused on the development of photography—was pushing the limits of what used to be considered private. They claimed that the right to privacy was an instance of the more general right 'to be let alone.'

Philosophical debates on privacy took off in 1975, with Judith Jarvis Thomson arguing that the right to privacy could be explained away appealing to other, more fundamental rights, like the right to property and self-ownership.

Most people today do not subscribe to these views of privacy. As research on privacy has become more developed, nuanced, and precise, two strands of views have emerged. According to the first strand, defended by people like Anita Allen, privacy is a matter of *access* to things deemed private (such as personal information, or the naked body). The second strand, defended by people like Julie Inness and Alan Moore, argues that having privacy is rather having *control* over what we deem private. Macnish sides with access theories of privacy, arguing that sometimes we may lose control of our private information without losing privacy (for example, when we lose our diary but no one ever reads it).

Another important account within academic debates is that of Helen Nissenbaum, who argues that what is important about respecting privacy is keeping the content of what is private in the context it was supposed to be in—medical data in the doctor's office, and private conversations between friends. Privacy is violated when information is shared in inappropriate contexts: when our doctor sells our medical information to advertising companies, or our friends share our private conversation on Twitter. Macnish goes on to argue that the value of privacy resides in its ability to protect individuals and to contribute to democratic freedoms. The interests we have, both as individuals and as a society, in securing privacy support recognising privacy as a basic right.

Finally, Macnish ends his paper by refuting two popular arguments regarding privacy: that it is opposed to security, and that, if you have done nothing wrong, then you should not fear the diminishment of your privacy.

Regarding the first view, Macnish points out that we partly value privacy because it keeps us safe. Furthermore, he emphasises that not all people lose the same amount of privacy—some groups in society, such as immigrants, are targeted more frequently than others, and may not gain much in security through this privacy loss. Finally, he calls attention to the fact that people may prefer privacy to security.

Macnish then refutes the second view, first, by reminding us that even if we have nothing to hide today, tomorrow we may have something to hide about today (for instance, if a totalitarian regime were to come to power and disapprove of something we have done or of who we are). Second, he notes that it is naïve to trust authorities (whether private or public) with sensitive information. Abuse is always possible. Third, Macnish reminds us that there are many things human beings wish to keep private even when they have done nothing wrong.

2 Kevin Macnish, 'Privacy,' *Data, Privacy, and the Individual*. Madrid: Center for the Governance of Change, IE University, 2019. www.ie.edu/cgc/research/data-privacy-individual

# The Ethics of Data Acquisition

One of the central ethical difficulties facing any data acquisition procedure is balancing the rights of data subjects against the potential benefit that data collection and analysis can offer. Alfred Archer, Nathan Wildman, Huub Brouwer, and Amanda Cawston (the Tilburg Centre for Logic, Ethics, and Philosophy of Science) provide a critical overview of various data acquisition models, determining and assessing the ethical issues each raises.[3]

The first model is the **opt-in donation model,** wherein data is acquired via individual philanthropy, without any kind of incentive or compensation. The opt-in model fares well in protecting data subjects' rights because individuals have complete freedom to not donate their data without fearing penalties or disadvantages. There is no worry, then, regarding unfairness, coercion, or exploitation. The main disadvantage of this model is that it is unlikely to be able to collect large data sets that can be representative of the population.

The second alternative is the **compulsory model,** in which all citizens are compelled to surrender their data, regardless of whether they want to share their data. The advantage of acquiring data from all citizens is the high quantity and diversity of data, but the model faces objections of violating autonomy (roughly, people's ability and right to lead their lives as they wish, in accordance with their values and without being subjected to coercion or manipulation) and privacy, and it risks engaging in unfairness and exploitation.

The two previous models prioritise the fulfilment of one of the two goals—gathering sufficient data or respecting data subjects' right—at the expense of the other. A third model that does a better job balancing both goals is the **opt-out model,** according to which everyone is presumed to give their consent for data collection and analysis unless they state otherwise. Concerns about this model include that subgroups of the population might disproportionally opt-out, influencing how representative the data is, people not being sufficiently informed about what data is being collected and how it may be used in the future, and mechanisms for opting-out being too complicated.

The final model assessed, and the one that the authors argue is most promising, is the **market model,** according to which people donate their data in exchange for some sort of compensation or incentive. Although this model has the potential to generate large data sets and make both parties in the exchange better off as a result of the data transfer, ethical concerns are present here as well. Market models may worsen inequalities by providing a greater incentive to poorer individuals to donate their data. Furthermore, if the incentive is too high, it might be an offer that data subjects cannot reasonably refuse, and if it is too low, data subjects might be the victims of unfairness or exploitation.

Even though the market model seems to be the best model to reach an appropriate balance between the goals of protecting data subjects and using their data, businesses and governments need to find ways of addressing the ethical challenges that arise in this framework.

3 Alfred Archer, Nathan Wildman, Huub Brouwer, and Amanda Cawston, 'The Ethics of Data Acquisition,' *Data, Privacy, and the Individual.* Madrid: Center for the Governance of Change, IE University, 2019. www.ie.edu/cgc/research/data-privacy-individual

# Private Data and Property

Closely related to the ethics of data acquisition, Verena Risse's (TU Dortmund) paper explores questions related to the protection and governance of private data by drawing on the analogy between privacy and property.[4] A common proposal to respect people's rights while using their data is to treat personal data as property, and compensate data subjects accordingly. It is unclear, however, to what extent personal data can be equated to property.

Risse starts out by presenting theories of property generation and acquisition. Both data and property are created by human beings. Data comes into being when human experiences are measured, collected, and analysed. Unlike property, however, it is not obvious that data is owned by whoever collects and analyses it.

Despite the intuitiveness of treating data like property, the analogy falls short in two respects, argues Risse. The first problem is what could be dubbed 'the black box problem.' Property can be easily delimited, and the contours made public—it is relatively straightforward to tell where a house ends and record those limits in a public registry while respecting the right to privacy. Not so with private data. In order to respect privacy, the content of what is private must not be known. What ought to remain private should never become data in the first place, or should only become data if it can be strongly anonymised or blacked box some other way.

At the moment, argues Risse, there does not seem to be a satisfactory way of black boxing private data. One could think that relying on consent might be a way out of the problem, but Risse argues that appropriate consent (free, voluntary, and specific) is too burdensome for all parties involved.

> 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.

The second problem with equating property and data is the heterogeneity among individual, economic, and public actors with regard to the acquisition, accumulation, and usage of private data. Property can be traded between all three kinds of actors as equals. Not so with data. Individuals have a lower interest in acquiring data than economic and public actors who are increasingly dependent on data. The acquisition of personal data empowers economic and public actors much more than it empowers individuals. Although the analogy between property and privacy allows important insights into the nature of both concepts, the differences are important enough to think that it is a mistake to treat personal data as if it were private property.

4 Verena Risse, 'Private Data and Property,' *Data, Privacy, and the Individual.* Madrid: Center for the Governance of Change, IE University, 2019. www.ie.edu/cgc/research/data-privacy-individual

# Informed Consent

In this paper, Kevin Macnish (University of Twente) explores the importance of consent in the collection and processing of personal data.[5] After an overview of what consent is and different kinds of consent, Macnish focuses on the debate as to whether consent is justified because it helps to respect autonomy, a view defended by Tom Beauchamp and James Childress, or whether it is grounded on limiting harm, a view defended by Onora O'Neill and Neil Manson.

The debate is relevant for the collection of personal data. If consent is important for autonomy, then the implication is that we should ask consent for all uses of personal data. If, on the other hand, consent is important only when there is a risk of harm, we may not need to ask consent for some uses of personal data (when there is no risk of harm, or perhaps a small risk of harm). Macnish ends up defending the autonomy justification for consent, although he recognises that such a stance comes at a cost: in some cases, asking for consent may undermine data collection that would have desirable consequences for individuals and society.

Macnish then considers the value of Jay Katz's proposal of consent as a mutual decision-making process. While Macnish argues that joint decision-making is not synonymous with consent, it may have a place when discussing risky choices.

Risk assessment involves two elements: the likelihood of harm, and the severity of the possible harm. One of the biggest challenges of risk assessment is the subjectivity involved when it comes to determining what levels of risk are acceptable for what kind of trade-offs in other goods (e.g., when it comes to exposing people's privacy, what level of risk are we willing to accept in exchange for the goods that data collection and analysis can provide?). First, people are notoriously diverse in how comfortable they feel about risk, some people being more risk averse than others. Second, the people making the decisions about risk are not necessarily the people who will bear the brunt of harms if things go wrong.

This asymmetry is standard with data collection. Data subjects are the ones who are risking their privacy, while risky decisions about the management of their data are often made by private companies and governments who have no skin in the game: they have everything to gain from exploiting data subjects' personal data and nothing to lose if things go wrong and data is misused (except perhaps a loss of reputation).

Macnish suggests that the power to remedy the asymmetry lies with policy-makers and courts who can impose costs on those collecting data, and guarantee fair compensation to data subjects who might be harmed through data collection.

A further solution to the challenges of making ethical decisions about risk is to engage in participatory Technology Analysis. Sven Ove Hansson and Helene Hermansson suggest that all stakeholders should be involved in discussing the possible impacts of technology in our lives. This proposal is consistent with that of Katz. In sum, Macnish proposes that when making decisions about risk, stakeholders should engage in a joint decision-making process that emphasises the value of consent as a tool to protect autonomy.

5 Kevin Macnish, 'Informed Consent,' Data, Privacy, and the Individual. Madrid: Center for the Governance of Change, IE University, 2019. www.ie.edu/cgc/research/data-privacy-individual

# Privacy, Autonomy, and Personalised Targeting

While previous papers have been concerned about the ethics of how personal data is collected, Karina Vold and Jess Whittlestone (University of Cambridge) are concerned with how personal data is used.[6] In particular, they examine the ethics of targeting ads and services to individuals.

Vold and Whittlestone first explore the connection between privacy and autonomy, which, they note, is rarely taken into account in contemporary policy discussions. When others have access to personal information about a subject, they can use that information to influence them. If such influence is exerted in surreptitious or manipulative ways, autonomy can be all the more jeopardised.

Although privacy and autonomy have always been connected, technological changes regarding data collection—the kind and amount of personal data collected, who holds and has access to the data, and how the data is used—make this link more important than ever. Asymmetries in data access create asymmetries in power. Vold and Whittlestone find particularly problematic the asymmetry between tech companies and users because, unlike governments, CEOs are not elected representatives and the goal of companies may not be aligned with public interests.

One of the most common uses of personal data is personalised targeting—using data to customise content and interventions. Personalised targeting can have advantages for both companies and consumers, if it helps improve services, but it can also lead to ethical problems. Vold and Whittlestone argue that the main concern is that personalised targeting can be manipulative—it is hidden from view, likely to be deceptive, and likely to be misaligned with users' interests. If the targeted person is not informed (and perhaps reminded) that they are receiving targeted information, they may act as if they were having access to a much more representative view of whatever state of affairs the context refers to.

The paper ends by offering some guidelines to ethical personalised targeting. First, personalised targeting should be consistent with people's values and interests. Second, it should be transparent. Third, companies should ask for users' consent for the collection and use of their data. Fourth, personalised targeting should not attempt to restrict information or choices in a way that (knowingly) misrepresents reality. Finally, personalised targeting should not make use of particularly sensitive personal data—for instance, information about people's vulnerabilities.

PERSONALISED TARGETING CAN BE MANIPULATIVE—IT IS HIDDEN FROM VIEW, AND OFTEN DECEPTIVE AND MISALIGNED WITH USERS' INTERESTS.

6 Karina Vold and Jessica Whittlestone, 'Privacy, Autonomy and Personalised Targeting,' *Data, Privacy, and the Individual.* Madrid: Center for the Governance of Change, IE University, 2019. www.ie.edu/cgc/research/data-privacy-individual

# Differentially-Private Data Sets: Methods, Limitations and Mitigation Strategies

Data set releases are often proposed as one way to address some of the ethical concerns related to institutions holding too much data, and hence too much power. Releasing data sets allows data to be available for secondary use and analysis. Data set releases, however, threaten the privacy of data subjects who might not have consented to such a release or who might have consented without realising the privacy risks they were signing up for.

THE PURPOSE OF THE GDPR IS NOT TO THWART DATA ANALYSIS BUT RATHER TO MAKE SURE THAT DATA ANALYSIS IS COMPATIBLE WITH THE PRIVACY OF DATA SUBJECTS.

One of the ways in which privacy can be better protected when releasing data sets is through the use of differential privacy. In this paper, Jordi Soria-Comas (Catalan Data Protection Authority) weighs the advantages and disadvantages of differential privacy in the context of data set releases.[7]

Perhaps the most attractive aspect of differential privacy is that it can offer protection regardless of whether intruders have access to other data sets. This is a significant advantage in the digital age. Institutions releasing data sets have to take into consideration, not only the personal information that can be gleaned from the particular data set being published, but also how that data can be aggregated and linked with other publicly available data sets (e.g., census data, Twitter data, etc.), further jeopardising data subjects' privacy.

Soria-Comas goes on to review the two main approaches used in generating differentially private data sets: histograms, and record aggregation and masking. Unfortunately, both methods lead to information loss. There are two common strategies to mitigate the loss of information incurred when using differential privacy: to increase the privacy budget, and to use a relaxed version of differential privacy.

The privacy budget is the amount of information the system will allow to offer—the more queries are allowed, the more information a researcher can get from a database, and the more privacy is risked.

Using large privacy budgets, argues Soria-Comas, renders differential privacy meaningless. Relaxing privacy guarantees, the author argues, is more promising to diminish the information loss, as long as privacy guarantees can still be meaningful, even if reduced. He ends the paper by describing some methods to relax differential privacy while maintaining some privacy guarantees.

7 Jordi Soria-Comas, 'Differentially-Private Data Sets: Methods, Limitations, and Mitigation Strategies,' *Data, Privacy, and the Individual.* Madrid: Center for the Governance of Change, IE University, 2019. www.ie.edu/cgc/research/data-privacy-individual

# Formal Versus Empirical Approaches to Data Anonymity

Most research on data anonymity focuses on methods with formal guarantees of anonymity, such as differential privacy. In this paper, Paul Francis (Max Planck Institute for Software Systems) argues that computer scientists should be open to and encouraged to work on empirical data anonymisation mechanisms in addition to formal ones—in much the same way that researchers work on both formal and empirical approaches to cryptography.[8]

Since the first paper on differential privacy in 2006, this method has been promoted as a practical and guaranteed private solution to data anonymity. After Apple, Google, and Uber announced their using differential privacy to protect users' privacy, media articles have generally portrayed differential privacy in an optimistic light. Given all the virtues of differential privacy, one might expect for it to be used widely. In fact, the opposite is true. The vast majority of institutions use weak ways of protecting privacy (e.g., removing personally identifying information like names and addresses). The reason for why differential privacy is not more widely used, argues Francis, is that differential privacy is not practical when configured with the strong level of anonymity that is necessary for the method to be meaningful. In other words, when differential privacy makes anonymity strong, data usability is lost because only a handful of queries can be made before the data must be made unavailable in order to preserve privacy.

If differential privacy seriously diminishes data usability, one might wonder why companies like Apple use it. According to Francis, it is questionable whether Apple uses differential privacy in a meaningful way.

The company and external researchers who have reversed engineered Apple's use of differential privacy disagree on how strongly private those methods are. Differential privacy can give a very accurate measurement of anonymity, but mathematical proofs are based on assumptions (e.g., will the data analysed be correlated with other data?), and when there is disagreement about the assumptions, there will be disagreement about how strongly private a particular tool is. In short, there are no international standards to label something differentially private, and no industry oversight, which results in companies claiming to use differential privacy without there being any certainty as to how private that data is.

In addition to continuing to research and invest in formal methods of privacy protection, and establishing standards for what counts as differentially private, Francis argues that we should also focus on using and developing empirical methods of ensuring anonymity. In contrast to formal methods, empirical methods do not offer a mathematical proof that can show how strongly protecting of privacy a tool is. Empirical methods, however, can be put to the test. By encouraging white-hat attacks on privacy tools, and offering generous remuneration to whoever manages to reidentify individuals in a given data set, we can get a sense of how strong our privacy protections are. Perhaps combining both formal and empirical methods for anonymising data can get us closer to better protecting privacy.

8 Paul Francis, 'Formal Vs Empirical Approaches to Data Anonymity,' *Data, Privacy, and the Individual.* Madrid: Center for the Governance of Change, IE University, 2019. www.ie.edu/cgc/research/data-privacy-individual

# PUBLIC VIEWS ON PRIVACY

# Survey

Siân Brooke (University of Oxford) and Carissa Véliz (University of Oxford) conducted an online survey of 1,107 people, mostly Americans and Europeans, about their views on privacy.[9] Among the many highlights of the survey, the following stand out.

**Privacy-related negative experiences online are extremely common.** The average respondent had more than one bad privacy-related experience. Almost a quarter of participants reported having had experienced an unauthorised access to their online account, 22% have had their credit card number stolen or have experienced bank fraud or unauthorised purchases from their account, and 10% have been victims of spyware. The great majority of respondents (92%) report having had at least one privacy breach.

**Table 1:** Experiences regarding privacy (All respondents)

| EXPERIENCE | PERCENT |
|---|---|
| Unauthorised access to my online account | 23% |
| Credit card number stolen / bank fraud / unauthorised purchases from your account | 22% |
| Being charged more for a product or service than other people | 10% |
| Someone using spyware on me | 10% |
| Someone impersonating me | 8% |
| Private emails or messages posted online without my consent | 7% |
| Public shaming online (people targeting me and shaming me for something I did or wrote, or for who I am) | 6% |
| Private images or videos posted online without my consent | 6% |
| Doxxing (private information posted online, such as my address) | 4% |
| Other (Free Text) | 2% |

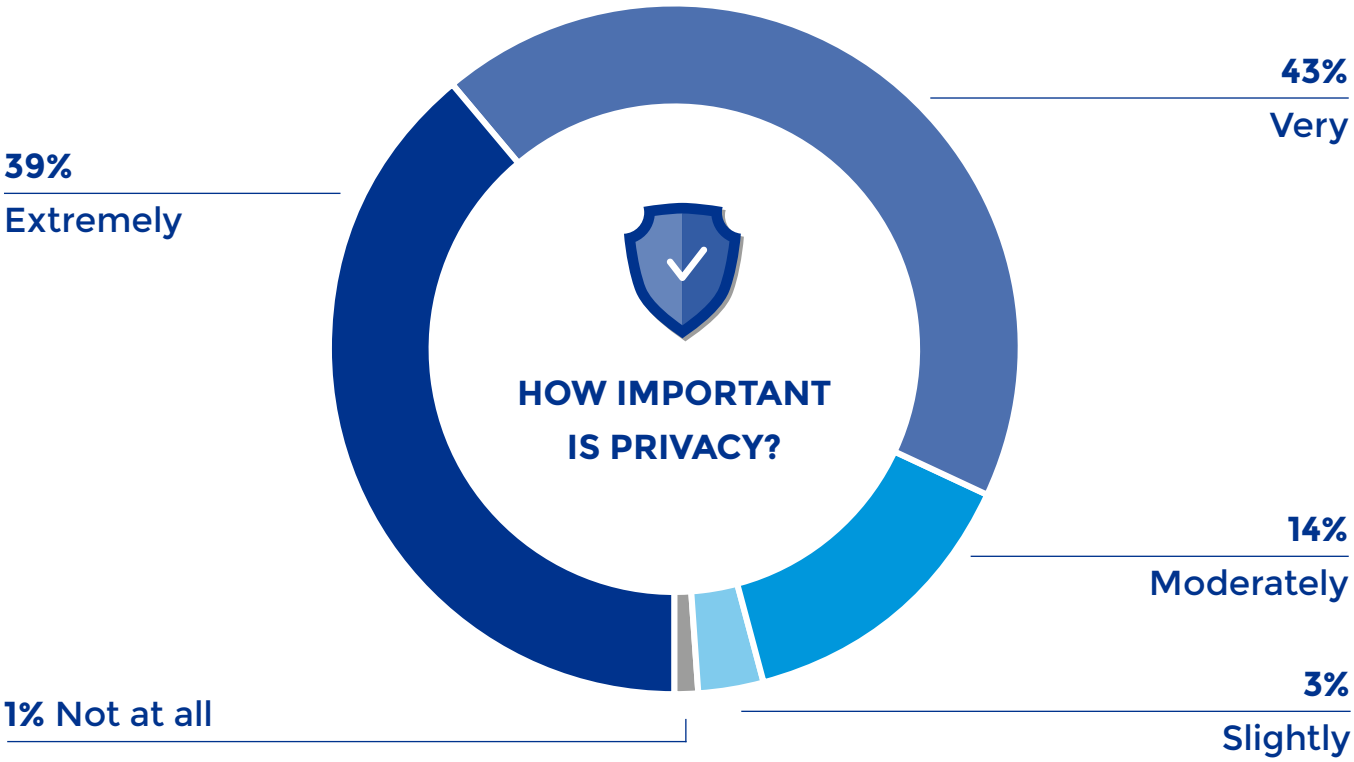
9 Siân Brooke and Carissa Véliz, 'Views on Privacy. A Survey,' *Data, Privacy, and the Individual*. Madrid: Center for the Governance of Change, IE University, 2019. www.ie.edu/cgc/research/data-privacy-individual

EVEN THE COMPANIES WITH THE HIGHEST RATING FOR TRUST ONLY REACH THE HALFWAY MARK OF OUR SCALE.

**People care about privacy.** Across continents, age, gender, and levels of education, people overwhelmingly think privacy is important. An impressive 82% of respondents deem privacy extremely or very important, and only 1% deem privacy unimportant. Similarly, 88% of participants either agree or strongly agree with the statement that 'violations to the right to privacy are one of the most important dangers that citizens face in the digital age.' There was no statistically significant difference across any demographic variable. One of the interesting implications from this result is that, against popular belief, young people do not seem to think privacy less important than their seniors.

**People are worried about their privacy, and they value privacy not only instrumentally, but also as a good in itself.** Respondents overwhelming expressed concern about their privacy. People's first concern when losing privacy is the possibility that their personal data might be used to steal money from them. Interestingly, in second place in the ranking of concerns, people report being concerned about privacy because 'Privacy is a good in itself, above and beyond the consequences it may have.' In other words, while privacy is important for people insofar as it can instrumentally protect them from certain harms, it is also extremely important for them in an intrinsic way that is unrelated to risks or harms.

**Figure 1:** Importance of privacy (All respondents).



Other reasons for being concerned about privacy (in order of importance to participants) include: personal data being used for impersonation in a way that could affect participants' credit ratings; the risk of personal data being used to tarnish their reputation; personal data being used by personal or professional enemies to hurt participants; personal data being misused by governments; personal data being used to discriminate against participants; the lack of privacy changing other people's behaviour in undesirable ways; the lack of privacy limiting free speech; and the lack of privacy leading to negative changes in participants' own behaviour.

**Not all uses of personal data by companies are thought to be unacceptable**. Respondents think that the most acceptable use of personal data by companies is to develop new products, followed by personalising ads. The most unacceptable uses of personal data by companies are thought to be selling it to third parties, influencing voting, and engaging in price discrimination. Curiously, while people are neutral regarding companies using personal data to personalise advertisements, they tend to disapprove of companies using this data to influence purchases. Given that personalised ads are a way of influencing purchases, more research is necessary to investigate what kinds of influences people disapprove of.

**Table 2:** Companies' use of personal data (All respondents)

| USE OF PERSONAL DATA | MD (MEDIAN) |
|---|---|
| Sell to Third Parties | 2 (Disagree) |
| Personalise Ads | 3 (Neutral/Undecided) |
| Price Discrimination | 1 (Strongly Disagree) |
| Develop New Products | 4 (Agree) |
| Investigate Prospective Employees | 3 (Neutral/Undecided) |
| Investigate Current Employees | 2 (Disagree) |
| Predict Behaviour | 2 (Disagree) |
| Influence Purchases | 2 (Disagree) |
| Influence Voting | 1 (Strongly Disagree) |

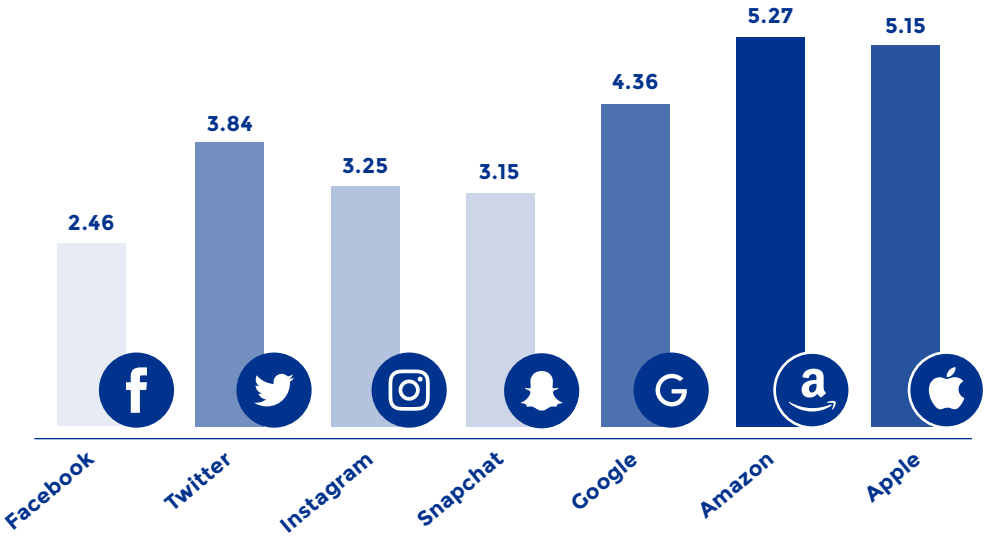



**Figure 2:**
Trust in Companies (Europe)

**People tend to feel that they cannot trust companies and institutions to protect their privacy and use their personal data in responsible ways.** From the Big Tech companies, Facebook is thought to be the most untrustworthy, and Apple and Amazon the most trustworthy. Even Apple and Amazon, however, get a low mean trust score (5,15 and 5,27 respectively, out of 10). That Amazon is trusted slightly more than Apple regarding privacy is surprising, given Apple's efforts in this area. Also somewhat surprisingly, Americans rate each company lower in trustworthiness than Europeans.

People do not trust other companies and institutions either. The least trusted institution are governments, followed by Internet Service Providers. The most trusted institutions are banks (with a mean of 6,69 out of 10) and healthcare providers (with a mean of 6,71).

**The majority of people believe that governments should not be allowed to collect everyone's personal data.** Most respondents (55%) think that governments should only be allowed to collect the data of criminal suspects, as opposed to everyone's data, with Europeans being slightly more likely than Americans to find some uses of bulk data collection acceptable.

The most acceptable purpose for governments collecting citizens' personal data is thought to be to catch criminals of serious crimes. Even then, only 29% of respondents find this use acceptable. Governments' use of personal data to prevent serious crimes was deemed acceptable by 24% of respondents. Interestingly, only 16% of people think it is acceptable to use this data for making sure that citizens are paying their taxes. Similarly, only 11% think that bulk data collection should be used to catch criminals of petty crimes.

**Privacy is thought to be a right that should not have to be paid for.** Angela Wineger and Cass Sunstein carried out a survey in which they asked people, on the one hand, how much would companies have to pay them per month to access their personal data, and on the other hand, how much would they be willing to pay per month to delete all their personal data from all parties that hold it. They found that participants were willing to pay $5 a month to delete their data, but asked $80 to allow companies to access their data. The authors ventured that perhaps a reason why people would be willing to pay such a low amount to delete their data is that they might think privacy is a right, the implication potentially being that they think they should not have to pay for something that they are owed as a matter of right.

Winegar and Sunstein's survey only involved Americans. They hypothesised that data privacy might be more likely to be considered a right in Europe.[10]

10 A.G. Winegar and Cass R. Sunstein, "How Much Is Data Privacy Worth? A Preliminary Investigation," *Journal of Consumer Policy* 42, no. 425-440 (2019).


Our survey connects with this literature by asking these questions to both Americans and Europeans, and by including the option of responding that privacy is a right for which people should not have to pay. Like Winegar and Sunstein, we found that people were willing to pay much less per month ($14, median) for deleting their data than what they would demand companies for them to access their personal data ($450, median).

Furthermore, our survey confirmed that the great majority of people who are not willing to pay anything to delete their data think that privacy is a right that should not need to be paid for: 73% of participants would pay nothing because privacy is a right, against 9% of participants who would pay nothing because they are not worried about online platforms holding their personal data. Only 19% of participants were willing to pay for their data to be deleted.

As Winegar and Sunstein hypothesised, Europeans were more likely than Americans to think that, when it comes to paying for one's data to be deleted, privacy is a right that should not have to be paid for (76% of Europeans thought so, and 68% of Americans). In line with this difference, Americans seem to be willing to pay nearly twice as much as Europeans to have their personal data deleted, and they ask for 150% of what Europeans demand for access to their personal data, both of which suggest that Americans are more open to seeing privacy as something that can be monetised.

**Table 3:** Pay to delete personal data: Region

| PAY TO DELETE PERSONAL DATA | ALL REGIONS | | EUROPE | | AMERICA | |
|---|---|---|---|---|---|---|
| | FREQUENCY | % | FREQUENCY | % | FREQUENCY | % |
| Nothing. Privacy is a right and I don't think we should need to pay for it. | 803 | 73% | 477 | 76% | 288 | 68% |
| I would pay a specified amount. | 205 | 19% | 109 | 17% | 90 | 21% |
| Nothing. I'm not worried about online platforms holding my personal data. | 96 | 9% | 42 | 7% | 48 | 11% |
| **Total** | **1104** | | **628** | | **426** | |

# DISCUSSION AND RECOMMENDATIONS

# Privacy Matters

**Our research unveiled a few overarching lines of thought that bear implications for good data practices.** Both ethicists and the public agree that privacy is a right that deserves strong protection. When 88% of people think that violations to the right to privacy are one of the most important dangers that citizens face in the digital age, governments and companies have good reason to take privacy seriously. People care about privacy, and they are rightfully unsatisfied with how their data is being used by both companies and governments. Institutions wanting to regain people's trust need to better protect citizens' privacy.

Our research has revealed that most people (92%) have already had a bad experience online related to privacy. This number is bound to go up in coming years.

One of the challenges for privacy experts has been to convey the importance of something that is often intangible until it is too late—that is, until the negative consequences of a privacy breach are felt.

As more people get acquainted with the harms and risks of identity theft, public exposure, and other privacy-related experiences, it will become more obvious why privacy is not a frill, but something necessary to have a well-functioning society.

At a minimum, then, privacy is important because citizens value it above and beyond its consequences; because the lack of it leads to harms such as theft, public exposure, and discrimination; and because it protects both security and democratic freedoms. Whenever sensitive data gets collected and stored, it is likely to be abused in some way at some point in time.

# Personal Data Is Not Like Property

Privacy is very important, for both individuals and societies, but from this conclusion it is not obvious to know how we should treat personal data in order to respect privacy. A very popular proposed solution is to treat personal data as property—to allow people to sell or trade their personal data. Given that capitalist societies are highly respectful of private property, it is intuitive to think that honouring personal data as property is respectful of privacy. Our research suggests this is not so.

The first difference between property and private data is the black box problem. Unlike physical property that can be easily delimited and its contours shared publicly, some private information should never become data in the first place, such that it is not always easy to establish beforehand what ought never to be caught in data collection (particularly given how data gets aggregated to then make sensitive inferences).

Second, while property is traded between all kinds of actors as equals, when it comes to data, individuals have a lower interest in acquiring it than companies and governments who know how to analyse it. Market models of data, therefore, increase inequality between individuals and institutions, and amongst individuals as well, as they provide a greater incentive to poorer individuals to sell their data, thereby turning privacy into a luxury product.

Third, while people who own a property such as a house have the moral authority to sell it, individuals do not have the moral authority to sell their personal data because that data contains sensitive data about other people.

Consider the genetic data that we share with our kin, including distant relatives. Similarly, your list of contacts includes the personal data of many people who have not consented to the trade of their data.[11] Individuals are not the rightful owners of their personal data in the way that they are the rightful owners of their private property.

Further support for not treating personal data as property comes from the results of our survey, which suggest that people, and Europeans in particular, tend to think that privacy is not the kind of thing that should be for sale.

The most direct implication of the differences between property and personal data is that models that incentivise individuals to trade their data are ethically flawed. Personal data is not only something that belongs to individuals—there is a collective aspect to privacy that make individual models fall short of being realistic solutions to the protection of privacy.

INDIVIDUALS ARE NOT THE RIGHTFUL OWNERS OF THEIR PERSONAL DATA IN THE WAY THAT THEY ARE THE RIGHTFUL OWNERS OF THEIR PRIVATE PROPERTY.

11 Carissa Véliz, "Privacy Is a Collective Concern," *New Statesman*, 22 October 2019.

# Privacy, Power, and Autonomy

Many of the ethical challenges that stem from personal data collection and use are related to asymmetries of power. When it comes to data, individuals are systematically disadvantaged with respect to companies and governments. Companies and governments have access to much more data than individuals, and they have much greater capacity to analyse the data in ways that lead to their advantage, and not necessarily to the advantage of data subjects. Data practices are more often than not inscrutable to individuals, which makes them vulnerable to invisible abuses of power that end up further disadvantaging them. Furthermore, while companies and governments are the actors who are making decisions about risky practices, individuals are the ones who are exposed to the greatest risks.

In other words, powerful institutions are gambling with data by engaging in risky practices while individuals are footing the bill when things go wrong.

As long as these asymmetries of power remain unaddressed, it will be difficult to achieve fair and ethical data practices. Companies should do what they can to empower users to be able to protect their data, thereby respecting people's autonomy. Making default settings privacy-conservative is an effective way to better protect privacy. Both companies and governments can also empower individuals by informing them thoroughly about what happens to their data, and never knowingly misrepresenting reality. The more a company wants to know about its users, the more it should be willing to give information about itself to its users. Policy-makers must make sure that the interests of data controllers and processors are aligned with those of data subjects.

If companies and governments get an advantage out of data, data subjects must get just as much or more of an advantage. If data subjects risk harms, companies and governments must risk just as much or more.

# Investing in Privacy and Ethics

One of the biggest challenges for companies and governments wanting to analyse data safely is anonymisation. According to the European legislation, personal data ceases to be personal once it becomes anonymised. But research has shown time and again that reidentification is possible with supposedly anonymised data.[12] Part of the problem is that we are uncertain about what kinds of technology might be developed in the future to reidentify data that is considered strongly anonymous today.

Differential privacy stands out as a promising possible solution to the problems of anonymising data. Our research suggests, however, that much more work needs to be done to make differential privacy—or a similar technical solution—practicable for businesses and governments. At the moment, putting in practice differential privacy is complicated, and adequate standards for privacy are controversial. The challenge is to insert precisely enough mathematical noise to make data as private and safe as possible without distorting statistical results. What this means in practice is that, first, databases often have to be much larger to get similar results, and second, the number of queries one can make to a database while maintaining high privacy protection will be limited. Differential privacy is most attractive as a way to collect data, such that there is never an 'original' database of sensitive information about individuals in the first place.

A promising avenue is to combine differential privacy with other privacy-protecting tools that may be empirical approaches. The disadvantage of empirical approaches to protect privacy is that we cannot be sure of how strong they are. We can invite people to try to hack our systems, but as long as getting to our data is more profitable or attractive than whatever is paid to people to test systems, there will always be a risk that our privacy protection is not good enough.

The current limits of technical solutions to protect privacy suggest, first, that we ought to be investing more in developing better privacy tools. And, second, given the existing data risks, that we ought to be investing more in developing better digital ethics.

IF WE INVESTED A FRACTION IN PRIVACY AND ETHICS AS WE ARE INVESTING IN DEVELOPING ARTIFICIAL INTELLIGENCE, WE WOULD HAVE MANY MORE REASONS TO BE OPTIMISTIC ABOUT THE FUTURE OF THE DIGITAL AGE.

12 Y. A. de Montjoye et al., "Unique in the Crowd: The Privacy Bounds of Human Mobility," Sci Rep 3 (2013); Y. A. de Montjoye et al., "Identity and Privacy. Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata," *Science 347,* no. 6221 (2015); Bradley Malin and Latanya Sweeney, "Determining the Identifiability of DNA Database Entries," *Proceedings, Journal of the American Medical Informatics Association* (2000).

# Data Principles

**The following data principles are not exhaustive, and no amount of principles can replace the judgment of an adequate ethics committee when assessing the ethics of privacy in a particular context. Nevertheless, these principles can serve as a start to developing better data practices.**

1. Data collection and analysis is only justifiable if it is necessary to fulfil a valuable objective for society and individuals
2. Data subjects whose data is being collected and analysed should be the main beneficiaries of such data collection and analysis
3. The more information data subjects share, and the more sensitive it is, the more they should benefit from the collection and analysis of their data
4. Collect as little data as possible, and collect the least sensitive data possible
5. Do not infer sensitive information from non-sensitive information without explicit and meaningful consent from data subjects
6. Have a plan to delete data, and store data for as little time as possible
7. Keep data as safe as possible during storage and analysis
8. Use technological tools that are privacy-protecting (e.g. differential privacy, homomorphic encryption, etc.)
9. Do not sell or share data with third parties without the explicit and meaningful consent of data subjects
10. Establish a chain of responsibility for data (who is responsible for what, and who is going to take care of what if something goes wrong)
11. Inform data subjects of your data practices (what data you will collect, for how long it will be stored, what kind of analysis will you perform on it, what it will be used for, etc.) before you collect their data
12. Allow data subjects to easily access, download, and visualise the data that you collect from them
13. Allow data subjects to easily delete their data, as well as to contest or modify information held about them, and to withdraw their consent for further data collection, analysis, or sharing
14. Allow data subjects to easily contact a Data Protection Officer (or equivalent) with any concern or data request they might have
15. Make sure data subjects are aware of the risks involved when giving up their data
16. Offer acceptable choices to data subjects. Instead of imposing take-it-or-leave-it policies, allow data subjects to negotiate giving up less data or no data and receive less functionalities or perks. Some people prefer their privacy over other benefits, and this choice should be respected
17. Ask consent from all relevant data subjects whose data is being collected or analysed. If one individual consents to sharing their data, but the data in question contains personal information about other people, consent is needed from those other relevant individuals in order to keep and use that data.
18. Default options should be privacy-conservative. Data subjects should have to opt-in to share their data

# CONCLUSION

The data economy caught most citizens unawares. Defenders of digital technologies managed to persuade society, for a relatively short period of time, that privacy was not relevant anymore, or not as relevant as it used to be. Experience is teaching us otherwise. Both as individuals and as a society, we are having to relearn the value of privacy through suffering the bad consequences that ensue from privacy losses.

One of the most important lessons of the past decade is that privacy is not only not contrary to security, but is part and parcel of it. To keep citizens safe, we have to ensure their privacy. And to protect privacy, we need to ensure the security of our digital systems.

In the early stages of the data economy, the collection and exploitation of personal data constituted a competitive advantage that allowed the likes of Google and Facebook to dominate online markets. The ripening of the digital age, however, has brought with it an increase in the frequency and sophistication of cyber-attacks, a techlash caused by users feeling exploited and betrayed, and the progressive regulation of the data economy, the GDPR being only the beginning of a worldwide trend. In this context, the exploitation of personal data is increasingly becoming a liability. The business race in the near future will not be about who can better collect and exploit our personal data, but about who gets to protect our privacy.

# TEAM OF RESEARCHERS


**DIRECTOR:**

Carissa Véliz, University of Oxford

**RESEARCHERS:**

— Kevin Macnish, University of Twente
— Nathan Wildman, Tilburg University
— Alfred Archer, Tilburg University
— Huub Brouwer, Tilburg University
— Amanda Cawston, Tilburg University
— Verena Risse, TU Dortmund University
— Karina Vold, University of Cambridge
— Jess Whittlestone, University of Cambridge
— Jordi Soria-Comas, Catalan Data Protection Authority
— Paul Francis, Max Planck Institute for Software Systems



**RECOMMENDED CITATION:**

Véliz, Carissa, *Data, Privacy & The Individual.*
Madrid: Center for the Governance of Change, IE University, 2020