Privacy and Autonomous Systems

Jose M. Such
Department of Informatics
King's College London
Strand, London, UK
jose.such@kcl.ac.uk

Abstract

We discuss the problem of privacy in autonomous systems, introducing different conceptualizations and perspectives on privacy to assess the threats that autonomous systems may pose to privacy. After this, we outline technical and legal measures that should be put in place to mitigate these threats. Beyond privacy threats and countermeasures, we also argue how autonomous systems may be, at the same time, the key to address some of the most challenging and pressing privacy problems nowadays and in the near future.

1 Introduction

Autonomous systems are becoming mainstream in practice, with the widespread introduction of autonomous vehicles, drones, desktop and smartphone personal assistants and so on. A crucial and very important issue is how this introduction affects human values and principles. In particular, we focus in this paper on the impact that autonomous systems may have on privacy. Protecting users privacy is not only essential to respect the Universal Declaration of Human Rights but also to serve as a first-line defense to mitigate cybercrime and other illegal activities that leverage the data obtained due to privacy breaches, such as online discrimination, phishing, identity theft, cyber scams, cyberstalking, cyberbullying, etc. Privacy is indeed an integral part of manifestos created by the AI community to research and ensure ethics and values in AI systems, such as the Asilomar AI principles[1].

While there have been some previous discussions, investigations, and reviews on the issue of how AI-equipped and autonomous systems may affect privacy in the past [van Blarkom *et al.*, 2003; Chopra and White, 2007; Such *et al.*, 2014], recent developments in AI and autonomous systems and their adoption in practice have brought new challenges or materialized those that had been largely abstract. Also, previous studies of privacy related to autonomous systems departed from a particular and narrow notion or concept of privacy, but there is a lack of a deeper understanding of the plurality of privacy and how autonomous systems could be a threat for it, as well as the measures that should be put in place for privacy-respecting autonomous systems.

In this paper, we provide a stepping stone towards understanding the interplay between privacy and autonomous systems. We first introduce the different meanings and conceptualizations of privacy for a primarily AI audience, and then we outline the threats that autonomous systems may pose to privacy. After this, we focus on technical and regulatory measures that could mitigate those threats. Finally, we also argue how autonomous systems, beyond just a threat, could actually be in turn the solution, or part of it, to some of the most pressing privacy challenges in an increasingly cyber-physical and hyper-connected world.

2 Privacy

Privacy does not have an agreed definition and it may have different meanings to different people, communities, and cultures [Acquisti *et al.*, 2015]. We shall, therefore, cover the most well-known and used conceptualizations in privacy literature, often neglected from the AI community, and what they mean in practice when assessing the challenges that autonomy could bring to privacy.

Privacy as Confidentiality This conceptualization of privacy comes more from the computer security field, in which privacy is usually associated with the concept of confidentiality. In turn, confidentiality is usually defined as a security property of a system that ensures the prevention of unauthorized reading of information [Stamp, 2006]. That is, only authorized users can have access to particular data, and when this property does not hold, then there is a confidentiality breach, and hence, a privacy breach if data is personal. Confidentiality is normally achieved through encryption, authentication and authorization technologies and services.

Privacy Notice & Choice This conceptualization of privacy has its roots in the privacy theories and experiments of Westin [1967], who defined privacy 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"*, which later on evolved to the information self-determination right [Rannenberg *et al.*, 2009], which plays a fundamental role in many privacy regulations and laws. This view of privacy emphasizes the importance of offering proper notice for the collection and secondary use of

[1] https://futureoflife.org/ai-principles/

personal data and pertinent opt-in mechanisms, and its implementation attempts usually include privacy policies and controls.

Privacy as Boundary Regulation This conceptualization of privacy comes more from a social science angle and has its roots in the theories of Altman [1975]. Privacy is defined in these theories as an interpersonal boundary regulation process, whereby individuals manage how they interact with others, continuously negotiating the information they reveal/conceal to/from others. This view of privacy is fundamentally associated to the management of social relationships and socialization, and the process of relationship evolution. This nature of privacy has particularly influenced the HCI communities to develop systems that allow users to manage those boundaries in computer-mediated communications [Palen and Dourish, 2003].

Privacy as Contextual Integrity Nissenbaum [2009] suggested privacy to be contextual, in the sense that information flows of personal information could be seen as appropriate or not depending on the context where these flows happen. In particular, she argues that each context has a number of *"finely calibrated systems of social norms, or rules, [which] govern the flow of personal information"*. This conceptualization of privacy, though limited in terms of the type and definition of norms considered, is akin to the notion of norm widely used in the autonomous systems literature [Criado *et al.*, 2011], which is one of the ways for restricting autonomy that could help to achieve privacy-respecting autonomous systems, as we discuss later on.

Privacy as a Plural Concept Very recent conceptualizations of privacy acknowledge all the definitions above, resorting to and embracing the *plurality* of privacy [Mulligan and Koopman, 2015; Acquisti *et al.*, 2015]. In particular, the emerging field of *engineering privacy* [Gürses and del Alamo, 2016] precisely considers this plurality to understand and operationalize privacy, as well as to systematize technologies developed under different views of privacy to turn into an engineering task the process of developing privacy-respecting technologies. We argue that any modern analysis about the privacy threats autonomous systems could pose ought to be conducted from this plural perspective of privacy.

3 Threats to Privacy

We now discuss the threats that autonomous systems may pose to privacy. Acknowledging the plurality of privacy, we focus on the information-related activities that can be performed when it comes to personal data based on two well-known taxonomies [Spiekermann and Cranor, 2009; Solove, 2006]. Each of these activities can raise privacy concerns and may do so differently according to the conceptualization of privacy, and other aspects such as how the activities are performed, what type of data is involved, who uses/handles/sees the data, personal and cultural factors, etc. Our aim is to highlight *when* and *where* a privacy breach may happen when dealing with personal information and hence privacy-respecting mechanisms should be considered. We later on focus on the specific mechanisms and how they could be used to design privacy-respecting autonomous systems.

3.1 Information Collection

Information collection refers to the process of gathering data about an individual. This is clearly relevant to most autonomous systems as they indeed collect information about individuals. For instance, autonomous cars, drones, and personal assistants collect a large amount of information about users whether running in a smartphone, a desktop or as standalone device. Information collection occurs in two main ways [Solove, 2006]: *surveillance*, when systems automatically collect data; and *interrogation*, when systems actively ask the user to provide the data. Examples of surveillance include an autonomous car collecting time-stamped location, distance travelled, routes taken, destination [Rho, 2017]; drones equipped with cameras may record people, e.g., there was a recent public outcry at drones in the US allegedly recording girls sunbathing in bikinis, occurring multiple times with individuals even shooting some of the drones down[2]; smartphone personal assistants gathering browsing habits, photographs, videos, etc. Examples of interrogation include autonomous vehicles asking for destination, personal assistants asking for the next song to play, etc.

Note that as with the other information activities, information collection may be *malicious*, e.g., drones that would purposely spy on people, but it can also be *non-malicious* or accidental, e.g., it could just be that a drone records all the information in-flight to better help recognize the terrain and it happens to accidentally record people around. Importantly, both may have privacy implications.

Finally, challenges may emerge depending on the degree of autonomy, e.g., a fully-autonomous system may at some point decide to change the information it collects, and/or decide to collect new information that it was not collecting before or supposed to be collecting at design time.

3.2 Information Processing

Information processing refers to the use or transformation of data that has already been collected. One of the types of information processing activities is *aggregation* [Solove, 2006], which consists of different pieces of collected data put together to depict an individual in a more detailed way. It also refers to the synergies that emerge by putting together all this information. Reasoning, learning, or inferring new information about an individual are examples of those synergies. In extreme cases, aggregation can lead to a situation whereby the system actually knows more about an individual than the individual herself, which has been coined the *inverse privacy* problem [Gurevich and Wing, 2016]. Autonomous systems may indeed perform aggregation activities. For instance, in autonomous cars, one may have together location, distance travelled, routes taken, time stamps, etc., that can be used to get a detailed profile of travelling patterns and habits, to prove someone was at a particular place, to predict future destinations and moves [Rho, 2017], etc.

Another information processing activity is *identification* [Solove, 2006], whereby an individual is uniquely identified

[2] http://www.slate.com/articles/technology/future_tense/2016/05/drone_privacy_is_about_much_more_than_sunbathing_teenage_daughters.html

and sometimes linked to a real-world identity. For instance, Jaguar has already developed a way to identify the owner of a car by recognizing face and gait[3]. However, this would also open up the door to cars recognizing other passengers within the car or in other cars, pedestrians that happen to walk passed the car whether parked or in motion, giving rise to a potential legion of *little brothers* that could monitor where people go and what they do. This may be worse than current CCTVs, as they are not in every single place, they do not usually have capabilities to identify people *online*, and they are relatively static compared with autonomous systems such as autonomous vehicles and drones, which could actively *follow* potential people of interest.

Secondary use of information is also a critical information processing activity [Spiekermann and Cranor, 2009]. In particular, information gained or inferred about an individual can be used for segregation and discrimination purposes, which are regarded as very privacy intrusive. Importantly, autonomous systems may automatically segregate and discriminate, which could make these intrusions even more privacy invasive [O'Neil, 2016]. For instance, a health-care robot could automatically decide not to apply a particular treatment or look after an individual if she has unhealthy lifestyles, or prioritize patients that have healthier lifestyles. Beyond specific decisions taken, it may just be humans *feeling* discriminated by an unexpected behavior based on personal data, such as their ethnicity or skin color[4].

We can also envision scenarios where more than one of the information processing activities could be conducted together. If an autonomous system, by way of *aggregation*, knows a lot about us from information collected and inferred (even things we may not be conscious about), and can link it to us by way of *identification*, what would stop it from actually *secondarily using* that against us in a strategic way? An example, there are already discussions about the morality of persuasive systems, particularly about whether these systems should have the capacity of lying and manipulating to persuade us [Guerini *et al.*, 2014]. The more an autonomous system knows about us (even more without us being aware of it), the better it may tailor persuasive strategies and the more manipulable we may be by those strategies.

3.3 Information Management

This refers to both information collected and information processed, and it relates to how collected and/or processed data is stored and managed. One of the issues here potentially leading to privacy breaches is *insecurity* [Solove, 2006], which relates to the concept of privacy as confidentiality stated above, i.e., personal data should be kept in a secure way and only authorized individuals should be able to access it. Otherwise, it would be easy for a potential attacker —whether script kiddie or more sophisticated attacker like a black hacker or a hacker sponsored by a nation-state — to break into the systems and gather personal data. Therefore, *secure* autonomous systems seem crucial here. However, the complexity and high-connectivity of autonomous systems suggests that they may even have more potential to be hacked. An example are autonomous cars, with their multiple components communicating within and with other autonomous cars, opening up a large amount of security challenges [Lima *et al.*, 2016]. Interestingly enough, these vulnerabilities could actually be exploited by other autonomous systems beyond human hackers.

The other privacy-related issue with information management is *exclusion* [Solove, 2006], where the individual has not a voice and cannot influence in any way the data that is stored about her, even when it may not be accurate — e.g., data may be collected erroneously, and data processing, particularly probabilistic inferences and machine learning, is prone to potential errors. Therefore, the decisions made by an autonomous system may just be wrong, and a human being might suffer the consequences, e.g., an autonomous taxi denying a ride for an individual believed to be somebody else who does not have enough money left in their taxi account or who did not pay her last ride. Beyond modifying data, there are also cases in which personal data is to be erased, including the *right to be forgotten* that is recognized by law in the EU at least.

3.4 Information Dissemination

Information dissemination refers to the transfer of collected and/or processed data to other parties. Some autonomous systems may not be that autonomous when it comes to information flows and they may be restricted architecturally. For instance, all personal assistants in the market from big players like Apple's Siri, Google's Google Assistant, Microsoft's Cortana, and Amazon's Alexa collect information from the smartphone or smart device such as a user query, but actually conduct information processing in a back-end remote facility, and it is only after a reply is generated, that it is sent back to the smartphone. Note that data transfers from the smartphones to the processing facilities are encrypted, which makes it difficult for a malicious third-party attacker to sniff what is being transmitted. However, the point is that data that may initially seem to be only in the smartphone will be with the company developing the autonomous system too, so beyond the autonomous system itself, corporations developing them become a potential privacy threat too.

Information dissemination may also occur *cross-cutting* organizational boundaries. In autonomous vehicles, an example is vehicle-to-infrastructure communication, which would allow tracking the location of individual vehicles and their owners by whoever is running the infrastructure [Lima *et al.*, 2016]. Another example of the negative effects this may have for privacy, collected/processed personal data may be transferred to or used by marketing or targeted/behavioral advertising companies, e.g., will we end up in a situation in which an autonomous car brings us home via a sponsored route through some shops believed to be of our interest instead of via the fastest route?

Autonomous systems may also disseminate data and share it with *other* autonomous systems, and collaborate or orchestrate themselves for a privacy invasive practice. For in-

[3] https://www.nfcworld.com/2016/11/28/348741/jaguar-files-facial-gait-recognition-system-patent-vehicle-entry/

[4] African Americans labelled as "gorillas" by automated taggers http://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/

stance, autonomous systems could bring targeted advertising to unprecedented levels. Indeed, there are already companies specialized in drone-based advertising[5]. One could imagine an orchestrated and highly-targeted advertising approach, whereby coordinated autonomous systems are able to identify an individual and with all the personal data collected and processed about her target the individual without many constraints on place or time[6]. This could become a nightmare when compared with unsolicited phone calls, spam, or other intrusions we currently face in our day-to-day life.

4 Privacy-respecting Autonomous Systems

Given the information-related activities that can threaten privacy, and the examples we gave about how autonomous systems can perform all of them, the question is, how can autonomous systems be designed to respect privacy? We give some particular examples, beginning with approaches and technologies coming from the privacy-enhancing technologies and engineering privacy fields, then moving to technologies that are being developed by the AI and autonomous systems communities, and finalizing by highlighting the importance of adequate regulatory frameworks.

4.1 Privacy by Design

Privacy by design, again, does not have an agreed meaning, and it usually means a set of principles to policy-makers and a number of privacy-enhancing technologies (PETs) to scientists and engineers [Danezis *et al.*, 2014]. The privacy engineering field, however, seems to be moving to a consistent set of building blocks to the design of systems, including: *"privacy-engineering methods to systematically capturing and addressing privacy issues during development, management, and maintenance; and privacy-engineering techniques to accomplish privacy-engineering tasks or activities"* [Gürses and del Alamo, 2016].

There have been 8 main design strategies coupled with their respective design patterns that have been proposed to achieve privacy by design [Danezis *et al.*, 2014] — refer to this publication for further details: 1) *minimizing* data collected by using select-as-you-collect, and anonymization and pseudonymization design patterns; 2) *hiding* data by using encryption (when in transit or when at rest), traffic hiding techniques (onion routing), etc.; 3) *separating* personal data as much as possible by means of distributed approaches; 4) *aggregating* data to process it at the highest level of aggregation and with the least possible detail in which it is still useful by using the k-anonymity family of techniques or differential privacy; 5) *informing* in a transparent way the subjects of the system by having adequate interfaces and detecting potential privacy breaches; 6) providing *control* to users over data by using techniques such as user-centric identity management, end-to-end encryption, etc.; 7) *enforcing* privacy policies by appropriate access control mechanisms; and 8) *demonstrating* the compliance with privacy policies by activities such as logging and auditing.

[5] http://www.dronecast.com/

[6] This would, in a way, be similar to the futuristic advertising approach portrayed in the science fiction film Minority Report.

These design strategies and patterns may not provide a drop-in solution for the development of privacy-respecting autonomous systems. Instead, multi-disciplinary approaches that integrate the task of engineering privacy from the beginning in the design and development of autonomous systems seem utterly essential and a very exciting research field. Also, privacy by design and some of its building blocks and techniques also have their limitations. For instance, a limitation with encrypted data is that it is very difficult to conduct computations on it. Even homomorphic encryption schemes provide limited computable functions over encrypted data, which may not be enough for complex tasks typical of autonomous systems such as reasoning, planning, inferring, or learning. An alternative may be trusted computing, which was already successfully used for private multi-party learning [Ohrimenko *et al.*, 2016].

4.2 Transparency

Although embedded in some of the strategies for privacy by design (particularly informing and demonstrating) we choose to talk now separately about transparency to highlight the aforementioned challenges that make privacy by design not applicable as a drop-in solution and it has to be considered together with the particular AI techniques employed. Transparency is indeed a concept that has been argued as fundamental for both privacy [Mulligan, 2014] and autonomous systems [Wortham *et al.*, 2016] separately. For privacy, transparency enables users to know how a system works and how their data will be used, and is most often associated with some of the privacy by design strategies mentioned above like informing, controlling and demonstrating. For autonomous systems, it has been argued that transparency would allow a better human understanding of autonomous systems and the decisions they take, which would foster trust. However, what does transparency actually mean in an autonomous systems context? For a long time, making the code open source was seen as a way of transparency in software, but for very complex systems such as autonomous systems, opening up the source code may not even be enough to know how the system works [Mulligan, 2014]. For instance, even for expert researchers, it may be rather difficult to prove whether a particular autonomous system is biased to segregate or discriminate people based on their personal data [O'Neil, 2016]. A promising research field in this particular case is that of fair computations, including fair machine learning [Dwork *et al.*, 2012], which actually aims to maximize the utility of the classification tasks subject to a particular fairness constraint, like that users should not be discriminated based on their membership to a particular group of users.

There may also be some tensions between privacy and transparency in autonomous systems, particularly when transparency means knowing the current state of reasoning of an autonomous system and the data underpinning that state. For instance, some studies have started to devise ways in which transparency could be provided *at runtime*, such that the *reasoning* process of an autonomous system is apparent to the humans around it [Wortham *et al.*, 2016]. This may be acceptable unless the data that might be shown or inferred due to the current reasoning status of an autonomous system is

personal, then there may be a privacy threat to be considered.

4.3 Normative Privacy

In terms of technologies that have been designed more from the autonomous systems field that could be used to mitigate the privacy threats already outlined, we highlight here norms and normative systems — other technologies coming from the autonomous systems field and their potential role are also discussed later on. Norms and Normative systems have been extensively studied in recent years particularly as a way of limiting the autonomy of autonomous systems to adhere to acceptable behaviors [Criado *et al.*, 2011]. Norms are usually defined formally using deontic logic to state obligations, prohibitions, and permissions, but other modalities such as commitments and other formalizations have also been considered. Norms could be used to define privacy-respecting behaviors, e.g., norms could define acceptable information collection, processing, management and dissemination, so that other non-acceptable behaviors would be prohibited. This could also extend to other modalities such as obligations or commitments to report purpose of information activities and stick to the purpose stated. Permissions could also be very important, e.g., you may want the location of the autonomous car to be shared with emergency services in the event of an accident or another emergency. Norms have the added benefit they can be used to govern the appropriate information flows considering the whole socio-technical spectrum, from non-autonomous to autonomous systems to human users [Singh, 2013], and they could be very useful as a *common language* for humans and autonomous systems which could foster transparency and accountability in turn.

Norms have indeed been used recently as a building block for privacy in socio-technical systems [Kafalỳ *et al.*, 2016; Murukannaiah *et al.*, 2016]. However, in these cases privacy requirements were known in advance from the stakeholders or the application designer specified them. The challenge then arises for the cases where no previous expectations of norms exist and/or eliciting them is non-trivial. Some researchers started using crowd-sourcing approaches to elicit the privacy norms that should be in force in particular domains. For instance, Shvartzshnaider et al. [2016] crowdsourced information sharing norms in an educational environment. However, a challenge that emerges is that there is hardly a norm for which there will be perfect consensus. Therefore, which norms to adopt and what they mean for the privacy of those who do not approve them needs to be studied, particularly as privacy is known to be very much personal, influenced by socio-economic and cultural factors, together with personality traits [Acquisti *et al.*, 2015].

4.4 Regulatory Frameworks

Privacy by design, transparency, and/or privacy norms could help to at least mitigate privacy threats that autonomous systems may bring with them, but what would the incentive be for developers and vendors of autonomous systems to actually use these technologies to design and develop autonomous systems? We argue that appropriate regulatory frameworks and requirements ought to be in place so that developers and vendors are required to design and develop privacy-respecting autonomous systems. There are some voices, however, that warn against regulations to push for and ensure privacy by design and similar approaches, as they claim that it may not be ethical not to give the choice to users to decide whether or not they actually want privacy preservation, in the sense that privacy by design could be too paternalistic [Pagallo, 2012]. Nevertheless, evidence shows that privacy is *malleable*, as interfaces and systems can be / and have sometimes been designed to make users disclose more information they would normally do [Chang *et al.*, 2016; Acquisti *et al.*, 2015], and users find it very difficult to make meaningful and appropriate privacy decisions under the uncertainty arising from information asymmetries and lack of transparency [Acquisti *et al.*, 2015]. Precisely, most of the privacy by design strategies emphasize activities to inform, give control, and demonstrate privacy practices to users, so they can actually make their own decisions freely. Regulations world-wide are moving or already moved to privacy by design, including Canada[7] and the EU[8], and the U.S. Federal Trade Commission encourages companies to apply it[9]. Finally, regulations are also important in terms of the data ecosystems that could emerge from the mass adoption of autonomous systems, particularly in terms of the threats mentioned above regarding information processing and dissemination, and the impact they may even have in autonomy itself, e.g., an autonomous car could be taking a sponsored route pushed down to it from vendors or advertising companies instead of autonomously deciding the best one.

5 Privacy-enhancing Autonomous Systems

So far, we have discussed how autonomous systems could be a threat to privacy and the measures to minimize the threat. However, autonomous systems may not only be a potential *threat* but also as a potential *solution* to privacy challenges. This is what we call *privacy-enhancing* autonomous systems. Note, we assume that privacy-enhancing autonomous systems should be *privacy-respecting* as well, i.e., they should be designed in such a way they respect privacy in the first place. Next, we give two particular examples of types of privacy-enhancing autonomous systems, but others may be possible.

5.1 Privacy Personal Assistants

The boundaries between the cyber, the physical, and the social are being blurred and we live in the era of hyper-connectivity, where it is becoming almost impossible for human users to make meaningful decisions and have control over their personal information [Acquisti *et al.*, 2015], and that is only worsening with the ubiquitousness of systems surrounding us, from the Internet of Things to autonomous systems themselves. We argue that precisely having privacy personal assistants could be at least part of the answer for

[7] https://www.ipc.on.ca/privacy/protecting-personal-information/privacy-by-design/

[8] http://www.eugdpr.org/

[9] https://www.ftc.gov/news-events/press-releases/2012/03/ftc-issues-final-commission-report-protecting-consumer-privacy

resolving these privacy challenges. We now discuss two avenues where privacy personal assistants could be of help.

The first avenue is privacy assistants that help users navigate the trade-offs and negotiate sharing decisions [Baarslag *et al.*, 2016], e.g., there is a myriad of entities or apps that request access to users' data, and being able to negotiate consent is extremely complex, so a privacy assistant could actually simplify the process of negotiating consent, especially if real access patterns are made apparent to the user later on, allowing reactive mechanisms to adjust information sharing dynamically. Negotiations may also be about data that is co-owned by multiple users, e.g. photos in social media in which multiple users are depicted and should all have a say on who access the photos [Such and Rovatsos, 2016]. Beyond negotiation, privacy assistants can also help a group of users reach a privacy decision by means of recommendations, e.g., on the optimal sharing policy for co-owned data based on their individual preferences [Such and Criado, 2016] and the reasons of their preferences [Fogues *et al.*, 2017]. Privacy assistants have the added benefit that they can learn the preferences of users over time, as sometimes users are not even (fully) aware of their preferences [Acquisti *et al.*, 2015].

Also, there may be cases where due to negligence or malice applications and devices surrounding us may be invading our privacy without us even noticing it. Therefore, privacy assistants that help users detect when privacy violations happen, such as [Kökciyan and Yolum, 2016], are of out-most importance. Beyond detecting and reporting privacy violations, privacy assistants could pro-actively seek to protect users. There are examples of applications that have been specially designed to counter privacy threats. For instance, there are scripts to turn off any cameras left in an AirBnB house, so that the person renting the house makes it impossible for the owner to spy on her[10]. One can easily imagine a privacy assistant that pro-actively carries out this type of actions or more sophisticated ones to protect users' privacy.

5.2 The Privacy of Autonomous Systems

There is also a very interesting relationship between the concept of privacy and autonomy that is yet to be explored thoroughly from an autonomous system point of view. After all, privacy in humans is actually related to the concept of their autonomy and agency to make decisions on information revelation and management. From that point of view, an autonomous system could make its own *autonomous* decisions about *its data*, and when that data is communicated to others. To this aim, autonomous systems should be well-equipped with disclosure decision-making mechanisms. These decision-making mechanisms could consider well-known concepts studied in the autonomous systems field such as trust and reputation [Pinyol and Sabater-Mir, 2013] to select and only share data with trustworthy privacy-respecting interaction partners, and normative reasoning [Criado *et al.*, 2011] to consider salient information flow and activity norms to decide whether a specific action is appropriate or not in the particular context/domain. Mechanism design, which has also been extensively studied in the autonomous systems literature, could play an important role as well to determine and/or influence the disclosure decision-making mechanisms of individual autonomous systems, especially those approaches to mechanism design that consider privacy concepts such as differential privacy [McSherry and Talwar, 2007].

The privacy of autonomous systems could have very positive effects for the privacy of individuals. An autonomous system may be protecting the privacy of an individual by protecting its own privacy. For instance, an autonomous car that does not reveal its current position to others could, at the same time, protect the privacy of those individuals travelling within the car. The privacy of autonomous systems could also refer to autonomous systems making their own decisions independently from the vendor, in a sense that no further control or interaction particularly when it comes to the data it holds could actually be determined by the policies or objectives of the vendor. This would be a mitigation for undesired information disseminations to the vendor and the posterior use or further dissemination by the vendor itself. Importantly, the privacy of autonomous systems could have a negative impact on individuals' privacy as well, e.g., an autonomous system that for some reason decides not to share personal data of a user with the user herself. This is why we argue that in order to exploit the full potential of autonomous systems as privacy-enhancing technologies, autonomous systems ought to be privacy-respecting in the first place.

Also, where autonomous systems are to interact/co-operate with humans, then the privacy of autonomous systems may be an enabler for human-like relationships and interactions, as privacy is known to play a crucial role in human relationships — see privacy as boundary regulation above. Therefore, an expected privacy behavior could help towards more humanized interactions between autonomous systems and humans. This may be particularly interesting in the domain of personal assistants, and smartphone assistants such as Apple's Siri indeed attempt to go beyond the simple query/answer paradigm to more human-like conversations, though they are still far from it. This could also be important in teams of humans and autonomous systems working together towards an objective [Jennings *et al.*, 2014].

6 Conclusion

We discussed the threats autonomous systems may pose to privacy and examples of methods, technologies, and regulations to ensure privacy-respecting behaviors. We also discussed how autonomous systems could actually help address privacy problems, and should not only be seen as a potential threat to privacy. Future research in autonomous systems should consider these two very exciting avenues we discussed, i.e., privacy-respecting and privacy-enhancing autonomous systems. We particularly expect and highly encourage multi-disciplinary collaborations between AI, privacy, HCI, sociology and law researchers working together on both avenues over the next decades.

Acknowledgments

We would like to thank the EPSRC for support under grant EP/M027805/2. We would also like to thank Natalia Criado for her comments on draft versions of this paper.

[10] https://julianoliver.com/output/log_2015-12-18_14-39

References

[Acquisti *et al.*, 2015] A. Acquisti, L. Brandimarte, and G. Loewenstein. Privacy and human behavior in the age of information. *Science*, 347(6221):509–514, 2015.

[Altman, 1975] I. Altman. *The Environment and Social Behavior*. 1975.

[Baarslag *et al.*, 2016] T. Baarslag, A. T. Alan, R. C. Gomer, I. Liccardi, H. Marreiros, E. Gerding, et al. Negotiation as an interaction mechanism for deciding app permissions. In *Proc. of CHI Extended Abstracts*, pages 2012–2019, 2016.

[Chang *et al.*, 2016] D. Chang, E. Krupka, E. Adar, and A. Acquisti. Engineering information disclosure: Norm shaping designs. In *Procs. of CHI*, pages 587–597, 2016.

[Chopra and White, 2007] Samir Chopra and Laurence White. Privacy and artificial agents, or, is google reading my email? In *Proc. of IJCAI*, pages 1245–1250, 2007.

[Criado *et al.*, 2011] N. Criado, E. Argente, and V. Botti. Open issues for normative multi-agent systems. *AI Communications*, 24(3):233–264, 2011.

[Danezis *et al.*, 2014] G. Danezis, J. Domingo-Ferrer, M. Hansen, J. Hoepman, D. Metayer, R. Tirtea, and S. Schiffner. Privacy and data protection by design-from policy to engineering. *ENISA*, 2014.

[Dwork *et al.*, 2012] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. Fairness through awareness. In *Proc. of ICTS*, pages 214–226, 2012.

[Fogues *et al.*, 2017] R. Fogues, P. K. Murukannaiah, J. M. Such, and M. P. Singh. Sharing policies in multiuser privacy scenarios: Incorporating context, preferences, and arguments in decision making. *ACM TOCHI*, 24(1):5, 2017.

[Guerini *et al.*, 2014] M. Guerini, F. Pianesi, and O. Stock. Is it morally acceptable for a system to lie to persuade me? *arXiv:1404.3959*, 2014.

[Gurevich and Wing, 2016] Y. Gurevich and J. Wing. Inverse privacy. *Commun. ACM*, 59(7):38–42, 2016.

[Gürses and del Alamo, 2016] S. Gürses and J. del Alamo. Privacy engineering: Shaping an emerging field of research and practice. *IEEE Secur Priv*, 14(2):40–46, 2016.

[Jennings *et al.*, 2014] N. Jennings, L. Moreau, D. Nicholson, S. Ramchurn, S. Roberts, T. Rodden, and A. Rogers. Human-agent collectives. *Commun. ACM*, 57(12), 2014.

[Kafalỳ *et al.*, 2016] Ö. Kafalỳ, N. Ajmeri, and M. P Singh. Revani: Revising and verifying normative specifications for privacy. *IEEE Intelligent Systems*, 31(5):8–15, 2016.

[Kökciyan and Yolum, 2016] N. Kökciyan and P. Yolum. PriGuard: A semantic approach to detect privacy violations in online social networks. *IEEE TKDE*, 2016.

[Lima *et al.*, 2016] A. Lima, F. Rocha, M. Völp, and P. Esteves-Verissimo. Towards safe and secure autonomous and cooperative vehicle ecosystems. In *CyberPhysical Systems Security & Privacy*, pages 59–70, 2016.

[McSherry and Talwar, 2007] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *FOCS*, pages 94–103, 2007.

[Mulligan and Koopman, 2015] D. K Mulligan and C. Koopman. Theorizing privacys contestability. In *Privacy by Design*, 2015.

[Mulligan, 2014] D. K Mulligan. The enduring importance of transparency. *IEEE Secur Priv*, 12(3):61–65, 2014.

[Murukannaiah *et al.*, 2016] P. K. Murukannaiah, N. Ajmeri, and Munindar P Singh. Engineering privacy in social applications. *IEEE Internet Computing*, 20(2):72–76, 2016.

[Nissenbaum, 2009] H. Nissenbaum. *Privacy in context: Technology, policy, and the integrity of social life*. 2009.

[Ohrimenko *et al.*, 2016] O. Ohrimenko, F. Schuster, C. Fournet, A. Mehta, S. Nowozin, K. Vaswani, and M. Costa. Oblivious multi-party machine learning on trusted processors. In *USENIX Security*, 2016.

[O'Neil, 2016] C. O'Neil. *Weapons of math destruction*. 2016.

[Pagallo, 2012] U. Pagallo. On the principle of privacy by design and its limits: Technology, ethics and the rule of law. In *European Data Protection*, pages 331–346. 2012.

[Palen and Dourish, 2003] L. Palen and P. Dourish. Unpacking privacy for a networked world. In *Procs. of ACM CHI*, pages 129–136, 2003.

[Pinyol and Sabater-Mir, 2013] I. Pinyol and J. Sabater-Mir. Computational trust and reputation models for open multiagent systems: a review. *Artif Intell Rev*, 40(1):1–25, 2013.

[Rannenberg *et al.*, 2009] K. Rannenberg, D. Royer, and A. Deuker, editors. *The Future of Identity in the Information Society: Challenges and Opportunities*. 2009.

[Rho, 2017] E. Rho. Privacy norms in the context of connected & self-driving cars. In *Networked Privacy*, 2017.

[Shvartzshnaider *et al.*, 2016] Y. Shvartzshnaider, S. Tong, T. Wies, P. Kift, H. Nissenbaum, L. Subramanian, and P. Mittal. Learning privacy expectations by crowdsourcing contextual informational norms. In *AAAI HCOMP*, 2016.

[Singh, 2013] M. P Singh. Norms as a basis for governing sociotechnical systems. *ACM TIST*, 5(1):21, 2013.

[Solove, 2006] D.J. Solove. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(3):477–560, 2006.

[Spiekermann and Cranor, 2009] S. Spiekermann and L. Cranor. Engineering privacy. *IEEE TSE*, 2009.

[Stamp, 2006] Mark Stamp. *Information Security: Principles and Practice*. Wiley-Interscience, 2006.

[Such and Criado, 2016] J. M. Such and N. Criado. Resolving multi-party privacy conflicts in social media. *IEEE TKDE*, 28(7):1851–1863, 2016.

[Such and Rovatsos, 2016] J. M. Such and M. Rovatsos. Privacy policy negotiation in social media. *ACM Trans. on Autonomous and Adaptive Systems*, 11(1):4, 2016.

[Such *et al.*, 2014] J. M. Such, A. Espinosa, and A. Garcia-Fornes. A survey of privacy in multi-agent systems. *The Knowledge Engineering Review*, 29(03):314–344, 2014.

[van Blarkom *et al.*, 2003] G van Blarkom, J. Borking, and J Olk. *Handbook of Privacy and Privacy-Enhancing Technologies: The Case of Intelligent Software Agents*. 2003.

[Westin, 1967] A. Westin. *Privacy and Freedom*. 1967.

[Wortham *et al.*, 2016] R. H. Wortham, A. Theodorou, and J. J. Bryson. What does the robot think? transparency as a fundamental design requirement for intelligent systems. In *IJCAI Ethics for Artificial Intelligence Workshop*, 2016.