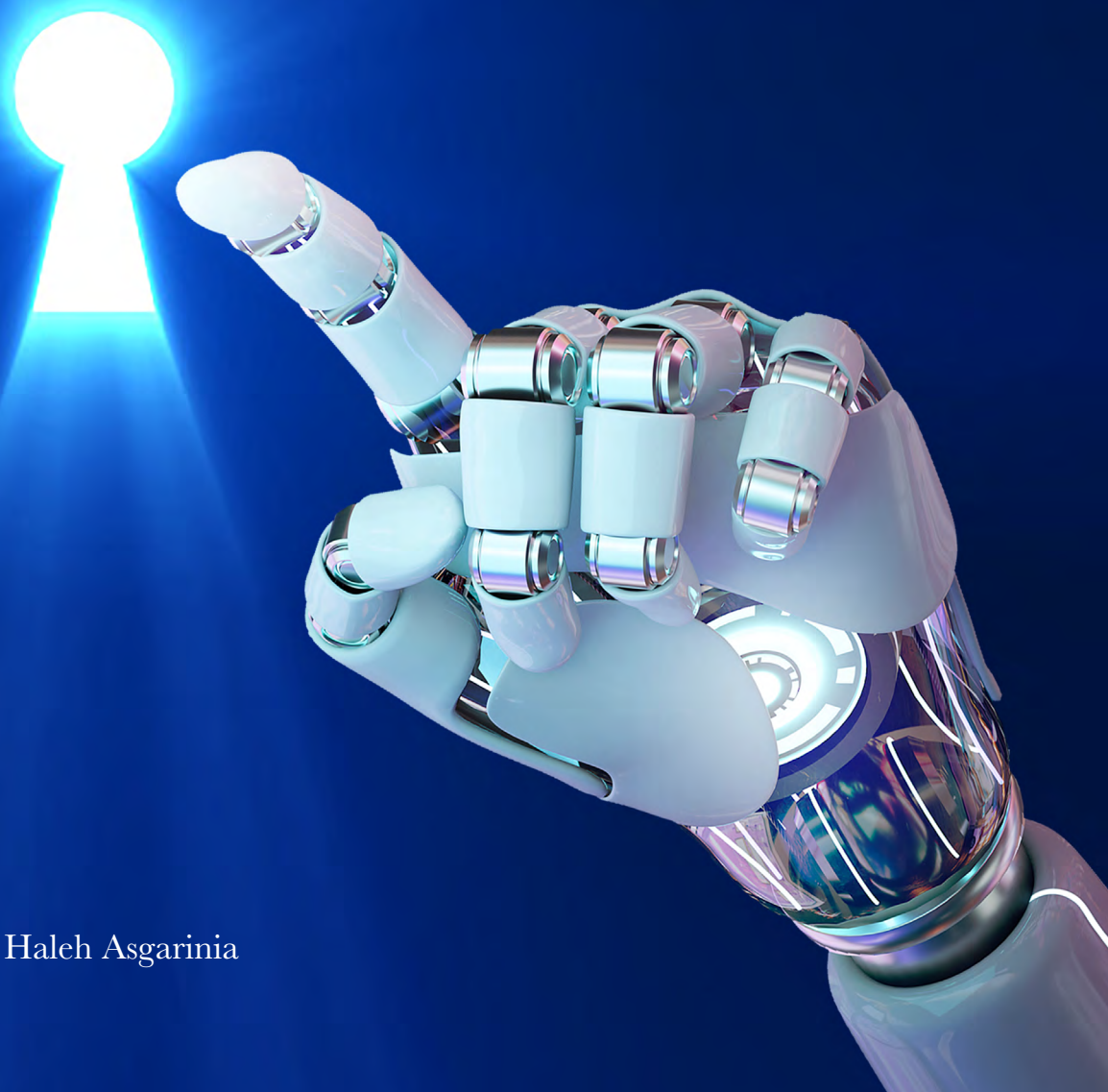


Privacy and Machine Learning- Based Artificial Intelligence: Philosophical, Legal, and Technical Investigations



Haleh Asgarinia

Simon Stevin Series in the Ethics of Technology

Privacy and Machine Learning-
Based Artificial Intelligence:
Philosophical, Legal, and
Technical Investigations

Haleh Asgarinia

Privacy and Machine Learning-Based Artificial Intelligence:
Philosophical, Legal, and Technical Investigations

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the Doctorate Board
to be publicly defended
on Thursday 16 May 2024 at 16.45 hours

by

Haleh Asgarinia

This dissertation has been approved by:

Promotor

prof. dr. P.A.E. Brey

Co-promotor

dr. A. Henschke

Research for this dissertation was made possible by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 813497.

© 2024 Haleh Asgarinia, The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author. Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de auteur.

The Simon Stevin Series in the Ethics of Technology is an initiative of the 4TU Centre for Ethics and Technology. Contact: info@ethicsandtechnology.eu

Cover design, print and lay-out: Grefo Prepress, Eindhoven

ISBN (print): 978-90-365-6011-5

ISBN (digital): 978-90-365-6012-2

URL: <https://doi.org/10.3990/1.9789036560122>

Copies of this publication may be ordered from the 4TU.Centre for Ethics and Technology, info@ethicsandtechnology.eu

For more information, see <http://www.ethicsandtechnology.eu>

Graduation Committee:

Chair / secretary: prof. dr. T. Bondarouk

Promotor: prof. dr. P.A.E. Brey
Universiteit Twente, BMS, Philosophy

Co-promotor: dr. A. Henschke
Universiteit Twente, BMS, Philosophy

Committee Members: prof. dr. M.A. Heldeweg LL.M.
Universiteit Twente, BMS, Department of Governance
and Technology for Sustainability

dr. J. van der Ham - De Vos
Universiteit Twente, EEMCS, Design and Analysis of
Communication Systems

prof. dr. E.L.O. Keymolen
Tilburg University

prof. dr. P. De Hert
Vrije Universiteit Brussel

To those who are invoking a hope that has been forgotten
and in memory of those who bravely resisted to usher in freedom

Table of Contents

Acknowledgements	xiii
Chapter 1: Introduction	1
1.1. Information Privacy and Artificial Intelligence	4
1.2. Inference	8
1.3. Research Aims and Questions	9
1.4. Methodologies and Approaches	11
1.5. Dissertation Structure and Overview of the Chapters	14
Part I	25
Chapter 2: Convergence of the Source Control and Actual Access Accounts of Privacy	27
Abstract	27
2.1. Introduction	28
2.2. Menges's Account of Privacy: Privacy as Source Control	29
2.3. Revising the Source Control Account of Privacy	33
2.4. Macnish's Account of Privacy: Privacy as an Actual Access	38
2.5. Revising the Actual Access Account of Privacy	39
2.6. Paradigmatic Cases	43
2.7. Theoretical Argument	46
2.8. Conclusion	47
Chapter 3: How Does an Artificial Intelligence System Affect Privacy? Adopting Trust as an Ex Post Approach to Privacy	49
Abstract	49
3.1. Introduction	51
3.2. Trust	54
3.3. Artificial Intelligence and Trust	68
3.4. Conclusion: Trust, Privacy, and Artificial Intelligence	73
Chapter 4: Limiting Access to Certain Anonymous Information: From the Group Right to Privacy to the Principle of Protecting the Vulnerable	77
Abstract	77
4.1. Introduction	79
4.2. Group Privacy	85
4.3. Approaches to Group Rights	89
4.4. Exploring the Nature of Goods That Qualify as Objects of Group Rights	92
4.5. Exploring the Types of Group Interests That Ground Group Rights	96
4.6. The Group Right to Privacy: A Collective or Corporate Right?	100

4.7. A Moral Principle to Protect the Privacy of Clustered Groups	108
4.8. Invasion of Group Privacy vs. Promoting Public Health	110
4.9. Conclusion	112
Part II	113
Chapter 5: Big Data as Tracking Technology and Problems of the Group and its Members	115
Abstract	115
5.1. Introduction	116
5.2. Key Ethical Issues	118
5.3. Current Measures to Address the Identified Issues	124
5.4. Conclusion and Recommendations to Improve Current Measures	131
Part III	133
Chapter 6: Design for Embedding the Value of Privacy in Personal Information Management Systems	135
Abstract	135
6.1. Introduction	136
6.2. The Layer of Values: Privacy and Autonomy	142
6.3. The Layer of Norms: Norms for Promoting Personal Autonomy	153
6.4. The Layer of Design Requirements: Design for the Value of Privacy	158
6.5. Conclusion	164
Chapter 7: Conclusion	167
7.1. Overview of Research Findings	167
7.2. Limitations of the Research	175
7.3. Recommendations for Future Research	177
References	183
Summary	203
Samenvatting	205
List of Publications	207
Simon Stevin (1548-1620)	213

Table of Figures

Table 1.1. Overview of Chapter Goals and Methodologies	13
Table 2.1. Compare and Contrast Different Versions of the Source Control and Actual Access Accounts of Privacy	45
Table 4.1. Summary of the Approached to Group Rights	99
Figure 6.1. Possible Values Hierarchy for Privacy	164
Table 7.1. Specific Norms and Design Requirements	174

Acknowledgements

‘You had to live—did live, from habit that became instinct—in the assumption that every sound you made was overheard, and, except in darkness, every movement scrutinised’ (Orwell, 1949, p. 5). How challenging it would be for one whose every aspect of one’s life had been intruded upon by the Other to think about and comprehend privacy. The writing of this dissertation would have been significantly more challenging without the support, influence, and inspiration of many people, which I acknowledge with appreciation.

I am deeply indebted to my supervisors for their invaluable guidance and support. Foremost, I am extremely grateful to Philip Brey, whose support and guidance not only helped me undertake this journey but also profoundly influenced my personal and professional development. I am particularly thankful for the insightful discussions we had, which enhanced my ability to express my ideas clearly; for his patience and direction, which helped me navigate and shape my research path; and for his trust in my ability to complete the project. Being part of the project on privacy not only extended my horizons on the topic but also impacted my home country through the various workshops and presentations I conducted there in an attempt to improve awareness of privacy and emerging technologies. Thank you for supporting me in reaching these milestones.

I would like to express my deepest gratitude to Adam Henschke, who was my daily supervisor during the latter half of my Ph.D. His valuable comments and feedback greatly improved the quality of my research and encouraged me to think more broadly. I am grateful for his guidance and advice, for sharing his knowledge and expertise with me, for his assistance in facilitating my visiting research at Birmingham University, and for his help during my time there. I also appreciate his support during challenging times, which reminded me I was not the only person in the world who had faced difficulties. Thank you for the non-philosophical conversations we had, especially those about the nature and birds of Australia, which inspired me. I eagerly look forward to the opportunity to see them someday.

This endeavour would not have been possible without the support of Kevin Macnish, my first daily supervisor. Thank you, Kevin, for helping me find myself and regain my confidence. I recall your words about standing on top of a mountain after completing my Ph.D., where everything would appear beautiful. Now, after navigating a path full of twists and turns, I can see beauty across a vast plain. Words

cannot express my gratitude for your encouragement to continue along my path. I am also grateful for the opportunity to spend my internship at Sopra-Steria Company, made possible by you. During my internship, I learned valuable lessons about the ethics of emerging technologies, and it was a pleasure to collaborate with you on writing a report. Additionally, co-authoring a book chapter on privacy and the media was a delightful experience for me.

I had the opportunity to visit Birmingham University during my Ph.D., and I would like to extend my thanks to Marten Reglitz, my supervisor there. Thank you for the time you dedicated to each piece of manuscript I sent, carefully reading them and providing insightful comments, and for the enriching discussions we had. My visit to Birmingham was one of the best chapters of my Ph.D. journey.

My Ph.D. research was part of the ‘PROTECT – Protecting Personal Data Amidst Big Data Innovation’ project. I am grateful to everyone who made it possible for me to be a part of this project. In particular, I extend my thanks to Dave Lewis, my supervisor in PROTECT, for the valuable discussions we had and for his support and guidance in conducting various workshops, which were immensely helpful. My appreciation also goes to Jessica Grene for her management of the project and to Valerie De Moor for subsequently taking over the management responsibilities. I would like to express specific thanks to my colleagues and friends in PROTECT: to Andrés Chomczyk Penedo for helping me become familiar with the GDPR, to Beatriz Esteves for introducing me to knowledge engineering, and to Blessing Mutiro for the valuable discussion on the group right to privacy, even though I remain unconvinced whether there is such a right. Collaborating with you on conducting workshops, writing deliverables and papers, and presenting at conferences was immensely beneficial for me. I look forward to future opportunities for collaboration.

Many thanks to Yashar Saghai for his diligence in managing matters relating to the PROTECT at Twente. I also extend special thanks to him for fostering a sense of belonging within the department in me, particularly by providing opportunities for sharing my experiences with him and others who felt similar pain during the tough times and horrible events in my home country.

Many thanks go to the 4TU.Ethics community. The conferences and graduate courses offered by the community helped me understand the ethics of technology. I am specifically grateful for their support during COVID, which was marked by various workshops. I would like to thank Karen Buchanan for the wonderful plant gift card. I nurtured a baby plant throughout my Ph.D. journey, and as I have flourished in my studies, so too has the plant. My appreciation also goes to Anna Melnyk and

Leon Rosσμαier for the wellbeing workshop, Sage Cammers-Goodwin for the cooking workshop, and Shaked Spier for the yoga workshop.

I extend my gratitude to Björn Lundgren for his insightful comments and feedback on the draft of Chapters 2 and 3, which were incredibly helpful and valuable. Moreover, I am grateful to him for initiating and organising group meetings to discuss privacy with Alexandra Prégent and Bart Kamphorst. I really enjoyed our discussions, which were valuable to me, and I look forward to having more meetings and group discussions in the future.

Thank you to my friends who were or are in the philosophy department at Twente. I appreciate the good times shared with Elisa Pausco, Isaac Oluoch, Kristy Claassen, Margoth González Woge, Mayli Mertens, Patricia Reyes Benavides, and Sage Cammers-Goodwin during our lunch and coffee breaks. Special thanks to Jana Mišić, Leon Rosσμαier, and Rosalie Waelen for the enjoyable moments we spent together in workshops, conferences, and trips. I hope to continue these wonderful experiences in the future.

I would also like to express my heartfelt gratitude to my parents, who have been a constant source of love and encouragement. Maman and Baba, thank you for always being there for me. When I decided to pursue my Ph.D. in the Netherlands, I was apprehensive about the sense of longing that being far from you would bring, both for you and for me. Now, I believe that enduring the distance and the brief visits was worthwhile, considering the accomplishments I have achieved. I hope you share the same feelings of pride and fulfilment, and I wish you could have been here to celebrate my achievements together. Many thanks also go to my siblings, Hadiyah, Haniyeh, and Amir Hosein, for their unwavering emotional support. Though miles apart from you, your presence has always been felt and deeply valued.

Finally, but most important, I wish to extend my deepest gratitude to my husband, Ali Reza Ghavamipour, for his endless support and love throughout this journey. Ali Reza, thank you for encouraging me to persevere when I felt disappointed, for giving me the strength to continue, for always believing in me even when I doubted myself, for engaging in discussions with patience in understanding philosophical concepts, and for celebrating each achievement with me. A heartfelt thank you for understanding me, especially during the intense final month of my Ph.D. This accomplishment is as much yours as it is mine.

Chapter 1:

Introduction

Artificial intelligence (AI) systems based on machine learning (ML) have the ability to learn without being explicitly programmed. ML algorithms are designed to learn how to perform specific tasks by generalising from data, such as making accurate predictions or identifying structures within the data (Samuel, 1959). To execute these tasks, ML algorithms require training on massive collections of data, known as training datasets (Crawford, 2021). These datasets serve as the input data for ML algorithms, which are then used to build models. The development of ML models is iterative, meaning various model configurations are tested and refined repeatedly using the training datasets. Once certain models display promise, they undergo further analysis, and the unseen data are fed into them to perform their intended task (Al-Rubaie & Chang, 2019; Aouad et al., 2007).

Data drive the success of ML-based AI systems. A mass collection of data (i.e., training datasets) is required to train ML models. Ongoing streams of data are required to enhance the accuracy of ML models. Moreover, for the systems to operate, they require a steady input of data (Crawford, 2021). The task of developing an ML model involves uncovering correlations within the data. Therefore, a developer must collect substantial quantities of data, building the training datasets necessary for training the model (Aouad et al., 2007). The accuracy of the developed model is enhanced by testing it with the information in the training datasets. When trained, the model requires new data to be inputted to execute its tasks, which include identifying correlations among the features of data it has never previously encountered (Crawford, 2021).

Massive quantities of data, which serve as the foundation for AI systems, are collected from diverse sources, each adapted to the specific field in which ML is applied. For instance, in the field of astronomy and astrophysics, data about various astronomical phenomena and objects are collected to train ML models (Rodríguez et al., 2022). In agriculture, data about environmental factors are collected to assess soil and water conditions (Liakos et al., 2018). The application of ML in healthcare shows promises in medical diagnosis, personalised healthcare, improved treatment, and the

reduction of healthcare costs (Chowriappa et al., 2014). These advancements are achieved by collecting and analysing health-related information about persons (Akselrod-Ballin et al., 2019).

Certain AI systems depend on information¹ about persons—personal information—for their development and operation. Personal information is defined as ‘any information relating to an identified or identifiable natural person (“data subject”); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person’ (EU Parliament, 2016, Art. 4(1)).

Personal information can be categorised to include a person’s purchasing behaviour, financial information, and health data. The collection of this information enables the development of models for specific applications (Chen et al., 2014). For example, information about a person’s purchasing behaviour can be used to build models used for business execution and commerce (Abualganam et al., 2022). Financial information can be employed to develop models in finance, such as determining a person’s eligibility for a loan (Citron & Pasquale, 2014). In healthcare, health information is collected to develop models for purposes such as the diagnosis of a particular type of asthma (Haldar et al., 2008).

This dissertation does not confine itself to specific applications of ML and, therefore, does not examine specific categories of personal information. Instead, it focuses on personal information in general. The main reason for not limiting the focus to specific applications of ML and categories of personal information is to provide a comprehensive understanding of the broader issues associated with the use of personal information in ML. By avoiding a narrow focus, this dissertation offers insights applicable to a wide range of ML applications that depend on personal information.

This study concentrates on ML-based AI systems that employ predictive analytics, based on inductive or deductive reasoning (see Section 1.2), to predict (personal) information through the analysis of training datasets. The functioning of these systems can be described as follows: Personal information is collected, formatted as features by developers or data scientists, and then input into the systems. Subsequently, ML models are developed to predict (personal) information. Personal information, including features such as x_1 (e.g., age), x_2 (e.g., gender), and x_3 (e.g., browsing history),

¹ Throughout this dissertation, I use the words ‘data’ and ‘information’ interchangeably, regardless of their differences.

is collected to predict a specific outcome, y (e.g., online behaviour). The symbolic ML model representing the correlation between the features (x_1 , x_2 , and x_3) of the personal information collected and the outcome y can be expressed as follows:

$y = a \cdot x_1 + b \cdot x_2 + c \cdot x_3$, in which a , b , and c are coefficients that quantify the influence or weight of each respective feature on the predicted outcome y . Once the model is developed, uncovering correlations between features, such as x_1 , x_2 , and x_3 , when new data are sent to the model, the model employs the learned correlations to identify similar patterns in the new data based on the features present, predicting the outcome y . For example, if a website visitor is 30 years old, male, and has a browsing history of sports websites, the model—in which each feature such as age, gender, and browsing history impacts the prediction by certain amounts—predicts that the visitor is likely interested in sports ads.

Regarding the above discussions, this dissertation focuses on ML-based AI systems that centrally involve personal information and has an overarching concern with privacy and ML-based AI systems. To address this concern effectively, it conducts philosophical, legal, and technical investigations. First, through philosophical investigations, this study identifies and analyses how inference, as a process associated with ML-based AI systems, impacts the definition, the value of, and the right to privacy. This dissertation addresses the existing gap in the literature concerning inference not only by exploring how inference may invade or violate privacy but also by fundamentally examining its impacts on various aspects of privacy. Second, through legal investigations, it assesses whether and how the General Data Protection Regulation (GDPR) addresses inference. Third, through technical investigations, it proposes design requirements to embed the social value of privacy into systems.

The findings of this dissertation contribute to a more comprehensive understanding of privacy and the issues arising from the activities of (or in) AI systems, suggest the expansion of the scope of the information protected by principles established by the GDPR, and enable individuals to live their lives autonomously—considering that the value of privacy is realised when a person's autonomy is protected or promoted.

In the following sections, Section 1.1 provides an overview of the analysis on how AI systems impact information privacy, outlines legal provisions for privacy and data protection, and discusses technical requirements for integrating privacy into systems. Furthermore, this section identifies gaps and limitations in the existing literature on the philosophical, legal, and technical investigations into privacy and AI. Accordingly, it illuminates the contributions of this dissertation to these respective investigations.

Although the information presented in this section does not extend to the rest of the dissertation, it is crucial in identifying gaps and limitations that are subsequently bridged and addressed throughout the dissertation.

Section 1.2 elaborates on the components and activities categorised under inference as a process associated with ML-based AI. It delves into the gaps and limitations highlighted in the previous section, specifically concerning the impacts of inference on privacy and the limitations of the GDPR in addressing inference. Section 1.3 outlines the research aims and questions. Section 1.4 discusses the methodologies and approaches adopted in this dissertation. Finally, Section 1.5 addresses the dissertation structure and provides an overview of the chapters.

1.1. Information Privacy and Artificial Intelligence

Data practices that involve personal information raise ethical issues, typically regarding privacy, as van den Hoven (2009) notes. More precisely, data practices in AI systems, particularly those involving personal information, raise concerns about ‘information privacy’.² Following Solove (2008), data practices and activities that affect *information* privacy are categorised into different groups, including **information collection**, **information processing**, and **information dissemination**. Each group further contains various sub-activities (Solove, 2008, pp. 104–105). Solove’s description provides a useful way to link existing scholarship on privacy with the changes ushered in by AI, and these activities offer a way to identify and analyse privacy issues. Adopting these activities as a basis, I conduct a structural analysis of how AI systems affect information privacy. A detailed examination of these issues is provided in the subsequent paragraphs.

² In the digital age, privacy is frequently concerned with information—more precisely, personal information. While some privacy scholars discuss privacy as a matter that solely concerns information (Parent, 1983), others have disputed this perspective, claiming that privacy may relate to more than just information (Finn et al., 2013; Rössler, 2005). According to these scholars, privacy pertains to physical space and location, as well (Rössler, 2005). Non-informational privacy has further expanded to encompass one’s decisions, modes of behaviour, ways of acting, and the life projects they pursue (Rössler, 2005). Privacy also extends to aspects related to appearance, scent, taste, touch, and sound (Macnish & Asgarinia, 2023).

I acknowledge that privacy relates to more than just information, as non-informational privacy becomes pertinent when discussing privacy issues related to new developments in the virtual world, such as the metaverse, as highlighted by Brey (2023). However, due to this dissertation’s emphasis on AI systems, particularly ML-based AI systems that process large quantities of personal information to develop models, I focus specifically on information privacy.

Information collection is a form of surveillance that threatens privacy (Véliz, 2022), as it may reveal private information that would typically remain undisclosed (Solove, 2008). Thus, collecting personal information to develop AI models constitutes a form of surveillance, threatening the privacy of those from whom information is collected.

Information processing is divided into five sub-groups of activities, each representing a threat to privacy: *aggregation*, *(re-)identification*, *insecurity*, *secondary use*, and *exclusion* (Solove, 2008). *Aggregation*, which provides extensive knowledge about individuals that they have not knowingly and willingly shared, threatens the privacy of individuals (Henschke, 2017; Solove, 2008). AI systems exacerbate privacy concerns by making it possible for the aggregation of innocuous information that leads to the revelation of sensitive information about a person (Henschke, 2021).

Re-identification raises concerns about a person's privacy, as it can inhibit their ability to be anonymous (or de-identified; Solove, 2008). Individuals in de-identified datasets can be re-identified when multiple datasets are combined or cross-referenced with one another. When different datasets are merged, identifiers that were previously removed, such as names, might be available in one of the combined datasets, leading to the possibility of individuals being re-identified (Kammourieh et al., 2017), as Ohm (2009) also highlights. The issue of re-identification has intensified with the advancement of AI systems. Previously, techniques that added noise or random elements to datasets helped protect identities. However, with advancements in ML algorithms, these algorithms can detect and remove noise and random elements, increasing the risk of re-identifying individuals in datasets (Kammourieh et al., 2017).

Insecurity involves carelessness in protecting stored information from leaks and unauthorised access (Solove, 2008). A security breach occurs when ML models leak information about the individual data records on which they were trained (Shokri et al., 2017). For example, in cases in which a person knows that data about individuals with a specific disease have been collected for research on which algorithms were trained, if individual data records are leaked, then that person might conclude that the individual whose data were leaked has the disease, resulting in an invasion of that person's privacy.

Secondary uses of data violate people's expectations about how their information will be used, and they may be hesitant to provide their data if they know about the possibility of secondary use (Solove, 2008). Secondary uses of information may occur in AI systems. In these systems, the same dataset might be repurposed to train models in different contexts or to develop AI systems other than those intended when the data

were originally collected. As Crawford notes, once information is collected to train ML models, the context in which it was gathered is considered irrelevant (Crawford, 2021). The irrelevance of context makes it possible for the information to be used for secondary purposes, which, if done without individuals' consent or awareness, poses risks to their privacy.

Exclusion refers to the failure to allow individuals to know about, manage, correct, or amend a record of identifiable information about them that others have. This lack of inclusion deprives individuals of control over their personal information (Solove, 2008), which might impact their privacy. When ML models operate as black boxes, they make end-decisions with processes that are not transparent, explainable, or answerable. The obscurity regarding the consequences of AI's actions or decisions, or the lack of awareness of alternatives, leads to a loss of control over the use of AI (Coeckelbergh, 2020). Concurrently, individuals whose data are used by AI face challenges in comprehending how their information is being used and the decisions AI makes, which prevents them from exercising control over their personal information.

Information dissemination, including *disclosure*, affects privacy. Disclosure involves the revelation of information about a person that affects the way others judge them (Solove, 2008). An ML model might establish a correlation between certain features of X , which may often be innocuous information, and information with social or group characteristics, such as religious beliefs or ethnic origin, 'which most individuals in a given society at a given time do not want widely known about themselves' (Parent, 1983, pp. 269–270). Disclosing an ML model that communicates this kind of information allows others to 'target a person', in Henschke's (2017) words, by knowing features of X about them, which leads to inferring sensitive information, such as their religious beliefs. Thus, the release of an ML model compromises the privacy of individuals, who can easily be targeted based on the information that the model reveals.

To protect natural persons regarding the processing of personal information, the GDPR was designed. The GDPR, which is the European Union's landmark data protection legislation, aims to address data protection challenges in a global and increasingly interconnected era characterised by rapidly evolving technology (EU Parliament, 2016). The GDPR established seven privacy and data protection principles related to the processing of personal data. These principles specify that personal data shall be: (1) processed lawfully, fairly, and in a transparent manner in relation to the data subject; (2) collected for specified and legitimate purposes; (3)

minimised to what is necessary in relation to the purposes for which they are processed; (4) accurate and up-to-date; (5) stored only as long as necessary for the purpose for which they are processed; (6) secured to ensure integrity and confidentiality; and (7) managed in a manner that holds the data controller accountable.

Furthermore, it is increasingly recognised that technological design expresses certain values, making it desirable to explicitly address values during the technological design process. An approach specifically focused on privacy is privacy-by-design (van de Poel, 2021b). Privacy-by-design aims to reduce privacy risks by incorporating measures that ensure that the development and design of the technology protect privacy (Strauß, 2017). Such measures could include the use of various privacy-preserving techniques. Among the advanced privacy-preserving techniques, two primary ones are those aimed at ensuring the confidentiality of data and protecting the identity of data subjects. Hence, privacy-preserving techniques integrate the value of privacy into systems focusing on the relationships between privacy, confidentiality, and identity. Techniques such as obfuscation (Brunton & Nissenbaum, 2015), encryption (Miller & Bossomaier, 2021), such as homomorphic encryption (Naehrig et al., 2011), secure multi-party computation (Zhao et al., 2018), differential privacy (Dwork, 2008), and access control (Tourani et al., 2018) can be employed to ensure the confidentiality of data. Meanwhile, anonymisation (Irti, 2022) and pseudonymisation (EU Parliament, 2016) can be employed to protect the identity of data subjects.

Nonetheless, the existing literature exhibits shortcomings in addressing a challenging group of activities (i.e., inference) associated with AI systems based on ML that affect privacy. Although inference is sporadically mentioned in the literature (see Henschke, 2021), it has not been studied as a group of activities affecting privacy. In this regard, the literature reveals a gap in identifying and comprehensively analysing how inference impacts privacy. Furthermore, the GDPR has certain limitations regarding the inclusion of inferred data within its scope, leading to ambiguity in applying its principles related to data processing. Moreover, although the aforementioned techniques help protect privacy—once data are collected, stored, aggregated, and shared—by concentrating on the value of privacy and its relation to confidentiality and a person’s identity, there is relatively limited philosophical discourse on the incorporation of the social value of privacy into systems.

1.2. Inference

Inference in relation to ML includes inductive and deductive inferences (Crawford, 2021). First, there is inductive inference: information acquired from processing and analysing training datasets is generalised for all those whose information was not in the training datasets. These inductive inferences are supported by the data available in the training datasets. For example, from ‘all apples in training datasets are red, not green’, ML induces that ‘all apples are red, not green’ (Crawford, 2021, p. 97). Second, there is deductive inference, which, according to the Organisation for Economic Co-operation and Development’s (OECD) document (2022), is referred to as model inference or using a model. After the model is developed, new data on which it was not trained are fed into it, and information about the new data is deductively inferred to derive a prediction, recommendation, or other outcome. For example, the information conveyed by an ML model is that ‘those who have the feature X are more likely to have the feature Y’. If a person A, whose data were not used to train the model, has feature X, then it logically entails that A is more likely to have feature Y.

Accessing certain pieces of information about groups of individuals derived from inductive inferences—which could enable those groups to be easily identified and targeted and would likely be used to harm them in morally objectionable ways—raises concerns about the privacy of those groups.³ Furthermore, targeting a person with the information acquired from ML models entails revealing sensitive information about that person following a deductive inference (see Henschke, 2021). Making inferences, even from innocuous information, might raise concerns about privacy, as the inferred data might reveal sensitive or intimate information about a person (for more information, see Henschke, 2017, 2021).

Although the information obtained about a group is not linked to a specific person in the group, preventing it from being categorised as personal data, it is linked to a group and thereby enables the identification of the group. Therefore, the scope of privacy must be expanded to include group privacy, and the data considered within the realm of privacy should extend beyond mere information about natural persons to include information about the group as a whole, as well (Floridi, 2014, 2017). Moreover, as Wachter and Mittelstadt (2018) emphasise, information that is (deductively) inferred from the personal data fed into ML models must also be

³ Chapter 4 argues that generalising information gives rise to concerns about epistemic injustice, while accessing specific generalised information raises concerns about group privacy.

considered personal information, and thus it must be considered within the scope of privacy and data protection.

In light of the above discussion, two activities related to inference impact privacy. First, activities that involve accessing specific pieces of information which were uncovered by a model and which would likely be used in morally objectionable ways, raise concerns about privacy at the group level. Second, activities that pertain to using a model, in the OECD's (2022) terms, with a particular emphasis on the model's output (i.e., inferred information), impact privacy at the individual level. In addition to these two activities, I also consider AI's performance to be a component related to using a model, as it impacts the quality of inferred information (OECD, 2021) and, consequently, affects privacy. The impact of this component on the (social) value of privacy is discussed in detail in Chapter 3. Thus, the activities and a component that I categorise under inference include accessing information uncovered by a model, using a model with a particular emphasis on inferred information, and AI models' performance. Accordingly, in this discussion, I regard inference as a process that includes activities and components that are either associated with it or that result from it.

According to the preceding discussions, the literature reveals a gap and a corresponding limitation. First, the gap pertains to the exploration of the impacts of inference on privacy. Through detailed analysis, I consider the inference into inferred information, AI models' performance, and accessing information uncovered by an AI model. More precisely, I claim that the gap in the literature relates to the analysis of the impacts of these aspects on privacy. Second, as indicated in this section, this analysis exposes limitations in current privacy schemes such as the GDPR, as they do not include information about groups designed by ML algorithms (henceforth referred to as clustered groups) and inferred data within the scope of privacy and data protection.

1.3. Research Aims and Questions

This dissertation has an overarching focus on privacy and ML-based AI systems. Drawing from the privacy impact assessment (see Section 1.4), this dissertation has a three-fold aim. The first aim is to investigate the impacts of ML-based AI on privacy, with a particular focus on inference. To achieve this, activities and a component related to inference are distinguished and discussed separately: activities concerning the use of a model, with a particular focus on inferred information; the component of

the model's performance; and activities related to accessing information uncovered by an ML model. Concurrently, two aspects of privacy—the descriptive and the normative—are identified and explored independently. Within the normative aspect, the value of privacy—particularly its social value—and the right to privacy—particularly that of groups—are analysed separately.

To be more precise, the first aim of this dissertation is to analyse the impact of inferred information on the descriptive aspect of privacy (the definition of privacy); the impact of AI models' performance on the social value of privacy; and the impact of accessing information uncovered by a model on the privacy of a group. The discussion of group privacy necessitates an analysis of recent discussions regarding the recognition of the group right to privacy. Thus, this dissertation extends beyond merely discussing potential harms to privacy and delves into how inference impacts the definition of privacy, its value, and the right to privacy.

The second aim includes highlighting the limitations of the GDPR in addressing privacy issues concerning activities related to inference and providing suggestions to mitigate those limitations. The third aim is to embed the value of privacy into systems by proposing design requirements. These requirements are translated from norms that aim to promote autonomy, an end for which the instrumental value of privacy is defined.

Accordingly, the main research questions (RQs) are as follows:

RQ1: How does an ML-based AI system affect privacy?

RQ2: How effectively does the GDPR assess and address privacy issues concerning both individuals and groups?

RQ3: How can the value of privacy be embedded into systems?

To respond to RQ1, the following three sub-questions (SQ) are formulated:

SQ1: How does inferred information affect the definition of privacy?

SQ2: How does the performance of an AI model affect the social value of privacy?

SQ3: What impact does accessing information uncovered by AI models have on the privacy of groups, and how can group privacy be respected?

The responses to SQ1 to SQ3, which together serve to answer RQ1, are detailed in Chapters 2 to 4 of this dissertation. RQ1 explores how inference impacts privacy and is divided into three SQs. Each SQ examines the impacts of a specific component of inference on a particular aspect of privacy. Addressing SQ1 to SQ3 jointly provides answers to RQ1. The response to RQ2 is thoroughly explored in Chapter 5, while the responses to RQ3 are outlined in Chapter 6.

1.4. Methodologies and Approaches

AI systems, which depend on personal information, have significant impacts on privacy. To comprehend the effects of AI on privacy and to ensure its adequate protection, a privacy impact assessment (PIA) can be used. A PIA ‘is a methodology for assessing the impacts on privacy of a project, policy, programme, service, product or other initiative which involves the processing of personal information and, in consultation with stakeholders, for taking remedial actions as necessary in order to avoid or minimise negative impacts’ (Wright & de Hert, 2012, p. 5). A PIA is a systematic process that should begin at the earliest possible stages, when there are still opportunities to influence the outcome of a project. This process should continue until and even after the project has been deployed (Wright & de Hert, 2012). By identifying privacy issues at an early stage and providing system designers with relevant knowledge, a PIA can help raise the system’s privacy baseline, thereby enhancing the privacy of data subjects (Tancock et al., 2010).

There are several approaches to PIAs that differ in detail and the particular contexts of application. However, regardless of the differences, conducting a PIA generally includes the following fundamental stages: first, identifying and evaluating the potential privacy effects of the processing of personal information; second, checking for compliance with privacy legislation to determine whether a technology involving personal information processing complies with relevant legislative or regulatory requirements; and third, considering how to avoid or mitigate privacy issues by taking into account some measures to ensure that the development and design of the technology protect the privacy of data subjects (Wright & de Hert, 2012). An umbrella term for such measures is privacy-by-design, which aims at reducing privacy risks by avoiding or limiting the disclosure or processing of personal information (Strauß, 2017).

In this dissertation, I adopt a PIA as the overall methodology. A PIA is a risk-based approach used to identify and mitigate risks to privacy. However, this dissertation goes beyond merely exploring privacy risks; it examines the impacts of AI on the definition of privacy, the value of privacy, and the right to privacy, and proposes design requirements to protect privacy by integrating privacy into systems. Therefore, I do not conduct a PIA per se but rather adopt it as a structural element in my analysis. A PIA is intended to answer analytical, legal, and technical questions in relation to a single system, whereas in this dissertation, I adopt it to guide me in navigating issues concerning a single process common across a class of information systems. This methodological choice underscores the importance of focusing on the process during

investigations into privacy and AI, and suggests integrating this process into the investigations of privacy issues within a specific AI application or system. I articulate the three stages of the PIA as follows:

First, the analytical stage concerns the analysis of the impacts of ML-based AI on privacy, particularly with respect to inferred information, AI models' performance, and accessing information uncovered by an AI model. It is important to note that although the analytical stage of a PIA is commonly taken to evaluate the effects of specific software or systems on privacy, in this stage, I analyse how a specific process associated with ML-based AI (i.e., inference) impacts privacy. The analytical stage is covered in Chapters 2 through 4.

Second, the legal assessment stage involves assessing whether AI that involves the processing of personal information and developing AI models complies with the GDPR and, more importantly, evaluating whether the GDPR can address privacy issues concerning inferred information. This is done in Chapter 5.

Third, the design requirement stage pertains to proposing design requirements for systems aimed at protecting privacy by embedding privacy into systems. This is accomplished in Chapter 6.

Having outlined the overall methodology, I now elaborate on the specific approaches and methodologies I employ throughout the dissertation.

Chapter 2 contributes to answering the conceptual question of what privacy entails by analysing the recent descriptive accounts of privacy. The tradition of discussing conceptual questions in analytical philosophy involves the use of various thought experiments and counter-examples, which provide tools that may help to better understand the concept under debate. This approach is adopted in Chapter 2.

Chapter 3—which concerns the normative aspects of privacy, particularly its social value—analyses various discourses on the value of privacy to understand which essential components constitute privacy as a social value. In the discussion, I consider perspectives according to which the social value of privacy is discussed in relation to trust and adopt the doxastic account of trust. Accordingly, through an epistemological analysis of trust, I explore the relationship between trust and privacy.

Chapter 4—which is dedicated to the normative aspect of privacy, specifically the right to privacy—focuses on the group right to privacy. To study whether an algorithmically designed group has the right to privacy, I analyse different approaches in the literature on group rights and outline the conditions and criteria that are satisfied in cases in which a group has a right. According to these approaches, if a clustered group can have a right to privacy, then the right must be a collective or a

corporate right. I critically evaluate the consequent and demonstrate that it is false, which then logically implies the falsehood of the antecedent.

Chapter 5 checks for compliance with the GDPR and extends the analysis beyond merely determining whether an AI system complies with the GDPR in its collection and processing of personal information. It also involves a critical evaluation of the relevant regulations themselves, aiming to identify their limitations, particularly concerning the privacy of clustered groups.

Chapter 6 proposes design requirements for privacy. Among various embedded approaches, including value sensitive design (VSD; Friedman et al., 2008), and disclosive computer ethics (Brey, 2000), I focus on one specific aspect of VSD, namely the translation of values into more tangible design requirements (van de Poel, 2013), and construct a possible value hierarchy for privacy, focusing on its instrumental value in protecting and promoting autonomy (Mackenzie, 2008; Rössler, 2005).

I now summarise in Table 1.1 the specific methodologies employed to achieve the goal of each individual chapter.

Table 1.1. Overview of Chapter Goals and Methodologies

Chapter Number	Goal	Methodology/Approach
Two	Analysing how AI impacts the descriptive aspect of privacy	<ul style="list-style-type: none"> Analysing recent descriptive accounts of privacy (Macnish, 2018, 2020; Menges, 2020b, 2020a) Utilising thought experiments and counter-examples involving inferred information Revising the proposed accounts of privacy
Three	Analysing how AI impacts the normative aspect of privacy, particularly its social value	<ul style="list-style-type: none"> Analysing various discourses on the value of privacy (Altman, 1976; Nissenbaum, 2010; Steeves, 2009; Waldman, 2015) Epistemological analysis of trust and its relationship with privacy (Goldman, 2015; Hawley, 2019)

Four	Analysing how AI impacts the normative aspect of privacy, particularly the right to privacy	<ul style="list-style-type: none"> • Analysing different approaches to group rights in literature (Newman, 2004; Raz, 1988; Réaume, 1988) • Critical evaluation of the assumption that algorithmically designed groups have a right to privacy (Floridi, 2014, 2017; Taylor et al., 2017)
Five	Identifying the limitations of the GDPR, particularly concerning the privacy of algorithmically designed groups	<ul style="list-style-type: none"> • Analysing issues that arise from the collection and processing of personal information • Critical evaluation of the relevant articles of the GDPR • Identifying the limitations of the GDPR, specifically regarding the privacy of algorithmically designed groups
Six	Proposing design requirements for embedding the value of privacy into systems	<ul style="list-style-type: none"> • Considering various embedded approaches: Value Sensitive Design (VSD; Friedman et al., 2008), and disclosive computer ethics (Brey, 2000) • Focused examination of a specific aspect of VSD: translating values into tangible design requirements (van de Poel, 2013) • Constructing a possible value hierarchy for privacy, with emphasis on its instrumental value in promoting autonomy (Mackenzie, 2008; Rössler, 2005)

1.5. Dissertation Structure and Overview of the Chapters

This dissertation contains an introduction, five main chapters, and a conclusion. Adopting the PIA as its overall methodology, which has three stages, the main part of

the dissertation is further divided into three parts. Part I is devoted to the first stage of PIA, the analytical assessment, and consists of three chapters. Each chapter responds to one of the three SQs. Part II focuses on the second stage, the legal assessment, and consists of one chapter that responds to RQ2. Part III pertains to the third stage, which proposes design requirements, and consists of one chapter that responds to RQ3. Each chapter is written as an independent research paper.

Chapters 3, 4, 5, and 6 are modified versions of the corresponding publications. The idea published in the conference paper is expanded upon in Chapter 3. The introduction section is elaborated upon further in Chapter 4. The publication corresponding to Chapter 5 proposes adopting methodological approaches instead of theoretical ones in relation to the group right to privacy. This chapter elaborates on this suggestion by exploring methodological approaches in a specific case involving clustered groups. The introduction section and section on authenticity and identification, Section 6.2.2.1, are elaborated upon further in Chapter 6.

Part I

Part I analyses the impacts of inference on both the descriptive and normative aspects of privacy. The structure for this part is outlined in three chapters. They discuss the definition of privacy, the value of privacy, and the right to privacy. The impacts of inference on privacy, including in relation to inferred information, AI models' performance, and accessing information uncovered by a model, are examined.

The descriptive aspect of privacy incorporates the presumption that privacy is a neutral concept, defined by elucidating what privacy entails (Gavison, 1980; Macnish, 2020) without endorsing whether privacy is good or worth having. Neutral terms, such as 'decrease', 'diminishment', and 'reduction', describe the states, conditions, or measures of privacy. However, the normative aspect of privacy incorporates the presumption that privacy is something worthwhile, valuable, and deserving of protection (Nissenbaum, 2010), defined by elucidating what privacy should entail (Macnish, 2020). Value-laden terms such as 'violation' and 'intrusion' are used to define what counts as a violation of privacy (Post, 1989).

Distinguishing between the descriptive and normative aspects of privacy allows one to talk about states of privacy without begging the normative question of whether these states are bad (Nissenbaum, 2010). It opens up the possibility that, in certain circumstances, a reduction in privacy is not morally wrong, as it does not entail harm (see Chapter 2). As Nissenbaum posits, such a reduction need not constitute a violation, intrusion, or incursion. Therefore, in assessing an action or practice, a

reduction in privacy is not equivalent to being morally problematic (Nissenbaum, 2010, pp. 68–69).

The normative aspect of privacy involves discussions about the value of privacy and the right to privacy. In this dissertation, I distinguish between discussions concerning the value of privacy (see Chapter 3) and those pertaining to the right to privacy (see Chapter 4). This distinction suggests that, in certain cases, particularly in certain kinds of clustered groups, although privacy is valuable and deserving of protection to prevent potential harms, the right to privacy cannot be recognised to respect the privacy of these groups.⁴ Instead, other approaches or moral principles should be considered to the privacy of these groups. Distinguishing between privacy as a value and the right to privacy prevents confusion regarding how to protect the privacy of entities that are incapable of holding rights.

Chapter 2: Convergence of the Source Control and Actual Access Accounts of Privacy

Chapter 2 aims to respond to SQ1: ‘How does inferred information affect the definition of privacy?’ In addressing this question, it delves into the definition of privacy, with particular emphasis on information inferred from shared personal data. It analyses two recent accounts of the definition of privacy, namely the ‘source control view’ proposed by Menges (2020a, 2020b) and the ‘actual access view’ proposed by Macnish (2018, 2020). It discusses objections to both accounts by examining cases involving inferred information. Given the counter-examples presented, the accounts are revised, resulting in the proposed definitions of privacy.

There is confusion and disagreement about the definition of privacy, and several theories have been developed in the philosophical literature to define it. Inness describes such confusion well in her book *Privacy, Intimacy, and Isolation*:

‘Exploring the concept of privacy resembles exploring an unknown swamp. We start on firm ground, noting the common usage of “privacy” in everyday conversation and legal argument. We find intense disagreement about both trivial and crucial issues. ... [W]e find chaos. ... [T]he ground starts to soften as we discover the confusion underlying our privacy intuitions’. (Inness, 1992, p. 3)

Modern discussion regarding the definition of privacy typically began with Warren and Brandeis, who, in 1890, define privacy as ‘the right to be let alone’. The concept

⁴ This argument is presented in Chapter 4.

then evolved through two prominent approaches: the control definition and the access definition of privacy. The account of privacy as control holds that privacy is about the control that one has over oneself, as, for example, Inness (1992) and Westin (1967) advocate. By contrast, the account of privacy as access holds that privacy is a function of the extent to which people can access a person either physically or access information about that person, as, for example, Gavison (1984) and Reiman (1995) posit.

In recent years, discussions on the control definition and access definition of privacy have become more detailed, with the nuances of control and access being more explicitly defined in scholarly theories. Menges (2020a, 2020b), who advocates the control account of privacy, refines this approach by arguing that a loss of privacy occurs when a person loses what he calls source control over their personal information. Menges emphasises the importance of one's uncoerced desires and intentions in determining whether a privacy loss has occurred. Parallel to Menges's refinement in the realm of control, Macnish (2018, 2020) introduces nuances to the access account. He argues that, for a reduction of privacy to occur, the information must be accessed. Further, though, the information accessed must also be understood by the person accessing that information. The traditional access account is therefore supplemented by a semantic account that describes the capacity to understand the information by an agent.

Apart from the aforementioned views, Henschke (2020) develops the institutional model of privacy. Instead of discussing privacy in terms of information, control, or access, Henschke argues that privacy can also be understood in relation to inequalities in power, positing that individuals need privacy to protect them from the inequalities in power exerted by institutions, which stem from the knowledge they obtain from individuals. This view aligns with the political account of privacy that developed by Véliz (2021).

In Chapter 2, I do not examine privacy from either an institutional or a political perspective, topics that deserve separate studies. Instead, I limit my focus to the competing camps of control and access. I examine different cases discussed by the defenders of the respective accounts of privacy. I then revise the proposed accounts of privacy in the face of counter-examples involving inferred information, thereby elucidating what privacy entails within the seemingly endless debates on the topic.

Chapter 3: How Does an Artificial Intelligence System Affect Privacy? Adopting Trust as an Ex Post Approach to Privacy

Chapter 3 aims to respond to SQ2: ‘How does the performance of an AI model affect the social value of privacy?’ by examining the impacts of AI on trust. Privacy is defined as a social value constituted by trust-based relationships. According to this definition, the social value of privacy depends on trust. Hence, trust is a way of realising privacy in the context in which information has been shared; trust is an ex post approach to privacy. In this regard, the authority of trust norms is established because they are constitutive of privacy. According to the commitment account of trust (Hawley, 2019), one such trust norm is competency. By investigating how a person’s competency is affected by AI, with a focus on specific features of AI’s performance, I specify the impacts of AI on trust and, ultimately, on privacy.

Different theories explain the value of privacy for individuals depending on the particular value that is associated with it. Privacy is valuable in that it is connected to human dignity (Bloustein, 1964; Warren & Brandeis, 1890). Other theories articulate the value of privacy in terms of its relationship to freedom and autonomy (Goffman, 1959; Reiman, 1995; Rössler, 2005). The importance of privacy is also underscored in terms of its connections with human flourishing and well-being (Gavison, 1980; Moore, 2010).

The importance of privacy is described not only in terms of the moral values with which it is associated, but also regarding its role in preventing specific harms or problems that arise from its absence for individuals. Hence, privacy is valuable because it protects a person against harms or problems, as outlined by van den Hoven (1997), including information-based harm, informational inequality, informational injustice, and encroachment on moral autonomy.

Analysing the value of privacy is not limited to theories that describe its importance from an individual perspective or the value that it has for individuals. Rather, different theories have been developed to explain the value of privacy from a social perspective. The social value of privacy, or the value of privacy to society, is described in several connected and overlapping senses. Privacy is valuable in that it is a fundamental condition for progress and intellectual development (Richards, 2008); it enables a person to immerse and embed themselves in public or social relationships and interactions (Fried, 1968, 1984; Hughes, 2015; Nissenbaum, 2010; Rachels, 1975; Rössler & Mokrosinska, 2013); it is a common good (Nehf, 2003; Regan, 1995); it is understood as a social value (Steeves, 2009); and it has social implications, such as impacting democracy (Henschke, 2020; Lever, 2012; Merton, 1968).

From a social perspective, privacy is also valuable because it prevents the harms or problems that a lack of privacy can cause for society. The contribution of privacy to society is conceived in accordance with the kinds of problems or harms that it safeguarded against. There is social value in protecting against each problem or harm, and the value of privacy differs substantially depending on the nature of each problem (Solove, 2008, 2015).

In Chapter 3, I focus on the social value of privacy. While admitting that privacy is necessary to develop and maintain relationships of trust with different people by giving one the ability to mediate various social relationships (Fried, 1984; Rachels, 1975), I emphasise that, as a social value, privacy is constituted in trust-based relationships. To reconcile the tension between these two perspectives concerning which notion is theoretically more basic, I argue that privacy and trust play a major role in forming each other. Fazlioglu (2023) also confirms this view, arguing that privacy and trust are intertwined with each other. In Chapter 3, I concentrate on the perspective according to which trust is a constituent of privacy, indicating that privacy depends on trust. Considering privacy as a social value constituted by trust forms the basis of the argument in Chapter 3. Through an epistemological analysis of trust and its relationship with privacy, I reach a conclusion that demonstrates how AI's performance impacts the social value of privacy.

Chapter 4: Limiting Access to Certain Anonymous Information: From the Group Right to Privacy to the Principle of Protecting the Vulnerable

Chapter 4 aims to respond to SQ3: 'What impact does accessing information uncovered by AI models have on the privacy of groups, and how can group privacy be respected?' It examines the effects of AI on privacy, focusing specifically on the right to privacy, in particular the group right to privacy. The chapter elucidates the issue of group privacy, contending that accessing certain pieces of information about a clustered group, which could be used in morally objectionable ways to harm that group, raises concerns about the privacy of the group as such. The chapter then analyses the predominant approach taken to protect the privacy of groups—the group right to privacy—which is explored, for example, by Floridi (2014, 2017) and Mantelero (2017). Subsequently, it investigates the plausibility of recognising such a right for these groups to limit access to certain information about them.

Privacy is recognised as a human right. Article 12 of the Universal Declaration of Human Rights states, 'No one shall be subjected to arbitrary interference with his privacy, family, home, or correspondence, nor to attacks upon his honour and

reputation. Everyone has the right to the protection of the law against such interference or attacks' (United Nations, 1948). Moreover, Article 8 of the European Convention on Human Rights (ECHR), which entered into force on 3 September 1953, grants individuals a fundamental right to respect for private and family life (see Grabenwarter, 2014, for a discussion on the rights granted by Article 8 of the ECHR).

Although the modern discussion on the right to privacy began with Warren and Brandeis's (1890) idea of the right to be let alone, in philosophical debate, the discussion of the right to privacy is generally considered to have started with Thomson's (1975) paper 'The Right to Privacy'. Thomson argues that 'the right to privacy is itself a cluster of rights, and that it is not a distinct cluster of rights but itself intersects with the cluster of rights which the right over the person consists in and also with the cluster of rights which owning property consists in' (Thomson, 1975, p. 306). In this regard, Thomson challenges the suggestion that there is such a thing as a right to privacy, arguing that the right to privacy can be understood as a cluster of derivative rights. Similarly, Henschke (2017) challenges singular conceptions of privacy and makes an analogy between ownership as a bundle of rights and privacy as a cluster of related conceptions, which gives rise to offering a pluralistic approach to describing privacy.

In 'What Is the Right to Privacy?', Marmor (2015) disagrees with Thomson, arguing instead that there is a general right to privacy. He provides a noteworthy philosophical account of the individual right to privacy by identifying the interests that deserve protection by imposing obligations on others. Marmor argues that the individual right to privacy is grounded in an individual's interest in having 'a reasonable measure of control over the ways in which they can present themselves (and what is theirs) to others' (Marmor, 2015, p. 4).

So far, the discussion has centred on the right to privacy as a right held by a natural person. However, Floridi (2014), who first raised the concept of group privacy in relation to big data analytics technologies, discusses the recognition of the right to privacy for groups designed by algorithms—a right ascribed to groups as a whole—which has received attention from scholars such as Mantelero (2017) and van der Sloot (2017). Floridi (2017) argues that making anonymised data available in cases in which groups can be easily identified and potentially discriminated against increases the risk of violating the right to privacy of these groups as a whole. In the discourse on group privacy, a consensus has been adopted in the work of scholars (e.g., Floridi, 2017; Mantelero, 2017; and van der Sloot, 2017) that the right to privacy of a (clustered)

group can achieve the goal of protecting these groups against discrimination. Hence, they argue that the rights to privacy of these groups must be recognised.

In Chapter 4, I discuss how limitations on accessing the information must be imposed. As the recognition of the right to privacy of the clustered group is considered to impose such a limitation, I focus on the group right to privacy. I explore the plausibility of recognising the right to privacy for clustered groups to ascertain whether this right can be defended or if moral principles must be developed and considered in the discourse on group privacy to protect clustered groups against discrimination by limiting access to certain information about them.

Part II

Part II conducts legal investigations into privacy and AI as part of the PIA legal assessment stage and consists of one chapter. Although, in April 2021, the European Commission presented a proposal for a regulatory framework for AI within the EU, this part of the dissertation concentrates on the GDPR, as this legal framework is aligned with the dissertation's objective of examining the effects of processing personal information on privacy.

The AI Act is the first endeavour to implement comprehensive, horizontal regulation for AI. The proposed legal framework centres on the particular use of AI systems and the risks associated with them. The proposed AI Act's risk-based approach distinguishes between different levels of risks that AI applications might have for fundamental rights and user safety (EU Parliament, 2023).

The EU AI Act provides regulations regarding the applications of AI systems in specific areas, such as employment, education, and healthcare. However, this dissertation does not limit its focus to specific applications of AI systems; rather, it concentrates on the specific processes associated with such systems. Consequently, this part of the dissertation focuses on the GDPR.

Chapter 5: Big Data as Tracking Technology and Problems of the Group and its Members

Chapter 5 aims to respond to RQ2: 'How effectively does the GDPR assess and address privacy issues concerning both individuals and groups?' This chapter, in addition to highlighting privacy issues, identifies the most significant ethical issues that arise from using AI systems as tracking technologies and their impacts on groups and their members. Furthermore, the chapter assesses whether the requirements and

obligations outlined in the GDPR can effectively address the identified issues, with a specific focus on group privacy.

The GDPR has certain limitations regarding the inclusion of inferred or ascribed data within its scope. It defines personal data as data relating to an identified or identifiable natural person, excluding data that are inferred or ascribed to an individual based on group membership. This exclusion introduces uncertainty regarding the application of the GDPR's principles to such types of information. In Chapter 5, I highlight the limitations of the GDPR in providing measures to protect this kind of information and, thereby, protect the privacy of a person.

At the group level, the GDPR has the potential to provide safeguards for specific groups. It provides enhanced protection for certain types of highly sensitive data, including data 'revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, and [...] the processing of data concerning health' (EU Parliament, 2016, Art. 9(1)). While protection is granted to individuals, its effects also serve to provide safeguards for groups, such as racial and ethnic groups (Kammourieh et al., 2017).

The question that arises is whether the protection of the privacy of members of a clustered group extends to protecting the privacy of the group as a whole, particularly in cases in which individuals are often incidental to analysis, and it is the group as a whole that is affected and harmed by the analysis (Floridi, 2017). In Chapter 5, I address this question, exploring privacy issues at the group level and assessing whether and how the GDPR can adequately address these issues.

Part III

Part III aims to propose design requirements, as part of the design requirements stage of the PIA. This part consists of one chapter. It suggests design requirements to integrate the social value of privacy into systems.

Chapter 6: Design for Embedding the Value of Privacy in Personal Information Management Systems

Chapter 6 aims to respond to RQ3: 'How can the value of privacy be embedded into systems?' It proposes design requirements aimed at embedding the social value of privacy in systems. It contends that privacy is valuable for the sake of personal autonomy (Rössler, 2005), a view that is modified to align with relational autonomy (Mackenzie, 2008). After discussing the three components of personal autonomy—

authenticity and identification, the genesis of desire, and goals and projects—general and specific norms are translated from them. Subsequently, I propose design requirements based on these norms, leading to the construction of a value hierarchy for privacy.

Focusing on the value of privacy, which is defined for the sake of autonomy (Rössler, 2005), I propose design requirements that are translated from norms aimed at promoting or protecting (relational) autonomy. Relational autonomy, as discussed by Mackenzie (2008), underscores the concept of autonomy as socially constructed, which, as highlighted by Cohen (2012), provides support for the social value of privacy. The design requirements proposed from philosophical investigations of the social value of privacy lead to distinct requirements, although they may overlap with some privacy-preserving techniques, such as encryption.

Chapter 7: Conclusion

Chapter 7 summarises the findings of the research by responding to RQs and SQs. Additionally, it critically examines the limitations of the research and proposes recommendations for future research aimed at enhancing the understanding of the topic.

Part I

Chapter 2:

Convergence of the Source Control and Actual Access Accounts of Privacy

Abstract

In this chapter, it is argued that, when properly revised in the face of counter-examples, the source control and actual access views of privacy are extensionally equivalent but different in their underlying rationales. In this sense, the source control view and the actual access view, when properly modified to meet counter-examples, can be metaphorically compared to ‘climbing the same mountain but from different sides’ (as Parfit (2011) has argued about normative theories). These two views can equally apply to the privacy debates and thus resolve a long-standing debate in the literature.

Keywords: access account of privacy; control account of privacy; descriptive aspect of privacy; the convergence of the access and control views

This chapter is published as:

Asgarina, H. (2023). Convergence of the source control and actual access accounts of privacy. *AI and Ethics*. <https://doi.org/10.1007/s43681-023-00270-z>

2.1. Introduction

Privacy has been defined through several theories in the philosophical literature; for example, it has been described as the right to be alone (Warren & Brandeis, 1890), a Wittgensteinian approach of family resemblance (Solove, 2008), control over information (Inness, 1996), and limited access to information (Gavison, 1980). Among these competing definitions, two views figure prominently: ‘access’ and ‘control’, which I find the most convincing. The access account of privacy holds that privacy is a function of the extent to which people can access a person or information about him or her (as held by, e.g., Reiman, 1995). The control account of privacy holds that privacy is about the control one has over access to oneself (as held by, e.g., Rössler, 2005; and Westin, 1967). This chapter does not aim to define privacy, whether as a redundant (or single concept) or as a pluralist concept (Henschke, 2017), but rather to contribute to one aspect of the debate, focusing on the two most popular accounts—control and access—and provide new insight into them.

Those who define privacy as a matter of control argue that a loss of control over one’s information constitutes a loss of privacy. However, those who define privacy as a matter of access argue that a loss of privacy only occurs when one’s information is accessed. These earlier, classical approaches to privacy did not clarify the meaning of control and the requirement for obtaining access in their theories. Recently, however, two privacy scholars have done so. Menges (2020a, 2020b), who defends the source control account of privacy, argues that privacy loss occurs when agent A loses the *source control* over his/her personal information flow. Concurrently, Macnish (2018, 2020), who defends the actual access account of privacy, argues that privacy loss occurs when another agent B *actually accesses* personal information about agent A. In this chapter, I focus on Menges’s and Macnish’s theories, as these accounts go beyond the existing descriptions of the control and access accounts, and argue that losing a new version of control—that is, source control—and understanding of that which is accessed—that is, actual access—are required for a loss of privacy, respectively. Moreover, although some hold that privacy includes non-informational aspects (Finn et al., 2013)—such as bodily privacy or behavioural privacy—here, I focus on information privacy because both the actual access and source control accounts of privacy are related to this aspect.

Throughout this chapter, I refer to a loss or diminution of privacy. I use this deliberately non-pejorative terminology to avoid being side-tracked into the question of when privacy may be waived, invaded or violated, or whether the loss of privacy leads to the violation or sustenance of a right to privacy. I do not discuss a right to

privacy or whether a loss of privacy is morally wrong, which would call for a different chapter. My aim is to provide an answer to the question of which accounts of privacy capture significant aspects of what the term means: source control or actual access. My answer is that neither account is preferable; both are extensionally equivalent.

It is important to note that focusing on the descriptive conception of privacy does not rule out the possibility of normative accounts; rather, searching for a philosophical definition of privacy can help make sense of normative debates that arise within moral or legal traditions. As Gavison rightly notes, the value of privacy can only be determined after a discussion of what privacy is and when—and why—losses of privacy are morally or legally wrong (Gavison, 1980, p. 452). Accordingly, the importance of concentrating on a descriptive conception of privacy can be defended by stating that it enables us to build a layer on top of it using criteria to determine how much privacy is good or required (Gavison, 1980; Powers, 1996). As such, the degree to which the descriptive conception can be articulated is critical. As a contribution to recent debates concerning the descriptive conception of privacy, this chapter specifies what a loss of privacy consists of, regardless of its legal or moral significance.

The chapter is structured as follows. In Section 2.2, I provide an initial definition of the source control account of privacy developed by Menges (2020a, 2020b). In Section 2.3, I discuss the problem with this account and present an alternative way to revise it in light of potential problems. Similarly, in Section 2.4, I provide an initial definition of the actual access account of privacy developed by Macnish (2018, 2020). In Section 2.5, I then discuss the problem of the actual account of privacy and present another way to revise it in light of potential problems. In Section 2.6, I provide *paradigmatic cases* that address whole comparable scenarios to see which revised versions explain the loss of privacy in the test cases. As I argue, both versions can explain the loss of privacy in the test cases. Hence, I show that the two alternatives actually converge on the same view—on an extensionally equivalent account. Finally, in Section 2.7, I suggest a *theoretical argument* to show that the two accounts of privacy from Section 2.6 are extensionally equivalent. I conclude that source control and actual access accounts of privacy can equally apply to the privacy debates and thus resolve a long-standing debate in the philosophy of privacy.

2.2. Menges’s Account of Privacy: Privacy as Source Control

Menges (2020a, 2020b) argues in favour of the control account of privacy by developing a new way to understand the relevant kind of control. In doing so, he relies

on (Frankfurt, 1969) the distinction between two different kinds of control. One is understood as the ability to do otherwise, which Menges calls leeway control, and the other is understood as having a desire or intention⁵ to act in a certain way, which he calls source control. I explain leeway and source control through detailed examples in the following paragraphs. Menges contends that privacy should be analysed in terms of source control, which provides a novel view for conceptualising privacy. This new account of control implies that an agent A's privacy is not diminished when exercising source control, even when A does not maintain leeway control. This differs from that of traditional accounts of control—being able to effectively choose *whether or not* something happens—emphasising the importance of leeway control for privacy. The conclusion is, then, that as long as A maintains source control over his/her personal information, s/he has privacy, despite losing leeway control (Menges, 2020b, pp. 34–35). To understand Menges's view, I now turn to the underlining principles as Frankfurt originally used them in arguing for moral responsibility.

Frankfurt cases (see Frankfurt, 1969) aim to show that agent A can be responsible for what s/he does because s/he can have the control which is necessary to be responsible for an action even if s/he cannot do otherwise. The main idea associated with Frankfurt cases is that the factors that explain why an agent A acts as s/he does differ from the factors that explain why A cannot act otherwise. By themselves, the latter factors do not undermine the agent's responsibility. For instance, other agents, devices, or any other external factors make it the case that A cannot effectively choose whether an event or action happens. In contrast, features of A themselves, namely their beliefs, desires, and intentions, explain why A is responsible for an action. The idea is that we do not need the ability to do otherwise to be responsible for our actions. Rather, what we need is to be the right kind of source of our actions (Menges, 2020b). The following case clearly shows the distinction between different kinds of control.

Jones resolves to shoot Smith. Black has learned of Jones's plan and wants Jones to shoot Smith. Black would prefer that Jones shoot Smith on his own; however, concerned that Jones might waver in his resolve to shoot Smith, Black secretly makes arrangements such that, if Jones shows any sign at all that he will not shoot Smith (something Black has the resources to detect), Black will be able to manipulate Jones so that he shoots Smith. As things transpire, Jones follows through with his plans and shoots Smith for his own reasons. No one else in any way threatened or coerced Jones,

⁵ It should be noted that I use the terms 'desire' and 'intention' in a technical sense. My view applies regardless of the specific propositional attitude or mental state that is relevant to a choice.

offered Jones a bribe, or even suggested that he shoot Smith. Jones shot Smith of his own accord, and Black never intervened (McKenna & Coates, 2004, Sect. 3.2).

Jones lacks leeway control because Black can coerce him into shooting Smith. That is, Black would make Jones shoot Smith even if he decides not to. Nonetheless, we still hold Jones responsible because he exercises source control over what he does when he shoots Smith, although he does not have an effective choice over whether he does it. He wants to do it, and it happens without any intervention, while he cannot do otherwise because of Black. Concurrently, if Jones did not have the desire to shoot Smith, Black would have made him do so regardless. In this case, Jones would lose source control if his action had not been related to his desire. Thus, we can have an important kind of control over what we do, although it is not possible for us to do other than desiring to do certain things—in this case, shooting Smith (Menges, 2020b).

Just as Frankfurt cases regarding moral responsibility distinguish between leeway and source control over actions, Menges distinguishes between leeway and source control over information. In this manner, Menges applies the distinction between leeway control and source control, which are typically discussed in non-informational contexts, to informational contexts. He contends that source control, not leeway control, is the kind that relates to privacy. The nature of privacy, according to the source control account, is being the right kind of source of information flows, if information flows at all. Being the right kind of source of information means that A has source control over information. Being the source control over information requires that, if the pieces of information flow to another person, this is the result of A's desire that it do so and A's desire that s/he desires to let it flow in this way (Menges, 2020a, 2020b). The following case clarifies this discussion.

Case 1: 'Imagine that I leave my diary on a table in a coffee shop and return to that shop 30 minutes later to retrieve it. When I enter the shop, I see a stranger with my diary on her table, a different table from the one at which I was sitting. I therefore know that she, or someone, has moved my diary; but have they read it? Imagine that the stranger has not yet read it but wants to know what my last entry says. She has firmly decided to read it before 3 pm and she would read it even in my presence (imagine that she is very strong and I would not be able to prevent her from reading it). I come back at 2.55 pm and tell her: "It's terrible, I'm forgetting everything these days! I hope I'm not getting ill. Actually, I wrote about it in my diary this morning. Please, look at the last pages". In response to this, the stranger reads my last entry in the diary'. (Menges, 2020b, pp. 35–36)

In this case, I lost leeway control because I lost an effective choice of whether the stranger learns or has access to certain information. I cannot do anything to stop her from accessing or learning the information. Nonetheless, an alternative to the leeway

control account, namely the source control account, says that I have source control because I have the desire to give the stranger some information about myself. Accordingly, I still have privacy because the flow of information is grounded on my desire; I am thus the right kind of source of information flow. The stranger would diminish my privacy if she learned about the last entry, even though my desires opposed this flow of information (Menges, 2020b).

The above discussions indicate that the key idea is that a loss of source control over personal information flow is necessary and sufficient for a loss of privacy to occur. Given that a descriptive definition of privacy aims to specify what a loss of privacy consists of (Powers, 1996), the *initial* definition of the source control account of privacy is as follows:

Definition 1: A's privacy is lost iff: A has lost source control over the personal information P about agent A, if information flows at all.

For Menges, a loss of source control over information flows is a *sufficient* condition for a loss of privacy to occur. Consider the following case:

Case 2: Imagine that 'you are walking outside in a storm with your diary in your bag. Unfortunately, you forgot to zip the bag completely, so the wind blows your diary out of the bag. It lands on the sidewalk with the pages facing up. Another pedestrian ... picks it up for you, but as he does so, ... he reads some of the content'.⁶ (Mainz & Uhrenfeldt, 2021, pp. 297–298)

In this case, as the flow of information is *not* grounded in what I desire; I am not the right kind of source for the information flow, and my privacy is thereby diminished. Menges thinks that a loss of privacy has occurred because source control over information flows has been lost. That is, if source control is lost or diminished, then

⁶ One might argue that this case shows more than merely a loss of source control, as the pedestrian has actual access to the information, as well. According to Menges's view, 'privacy essentially consists in being the right kind of source of information flow to another agent if the information flows at all. ... The information does not flow to another agent as long as nobody actually accesses the data and learns something about the relevant citizens. ... The source control is diminished as soon as an agent accesses the data before the relevant citizen tells them about it' (Menges, 2020b, pp. 45-46). In this case, if I freely and knowingly had asked the pedestrian—'who has not read my diary and does not plan to read it'—to read my diary, my privacy would not have been diminished, as no diminution of source control has occurred (Menges, 2020b, p. 39). The source control view only says that accessing information is relevant for privacy only if and because it diminishes being the right kind of source of an information flow. Thus, Menges says that accessing information is relevant for diminishing privacy and that the most important thing about privacy is having source control.

privacy will be lost or diminished. The loss of source control over the information flow is thus sufficient for the loss of privacy to occur.

For Menges, a loss of source control over information flows is also a *necessary* condition for privacy loss. Menges argues that, in Case 1, I have privacy because I am the right kind of source for the information flow. This is equivalent to saying that if privacy is diminished or lost, then the source control will be lost or diminished. The loss of source control over the information flow is thus necessary for the loss of privacy to occur.

2.3. Revising the Source Control Account of Privacy

Menges (2020a, 2020b) applies the split-level theory of control used in the discussion about moral responsibility to privacy. He then distinguishes between leeway and source control over information and emphasises A's desire in determining whether privacy loss has occurred. I posit that Menges has situations like Case 3 in mind when he theorises about the source control of privacy:

Case 3: 'Imagine Annabel. ... She suffers from a rare and very hard-to-diagnose genetic disorder, a piece of information about herself she wishes to keep private. One day, Annabel agrees to take part in a new medical initiative. The primary purpose of the initiative is to' (Rumbold & Wilson, 2019, p. 4) find various factors related to a different, more prevalent disease. As a participant in the initiative, Annabel donates her DNA intentionally to medical science. Suppose that Brian is a researcher trained in genetic medicine and works on medical research. He infers⁷ from Annabel's DNA profile that she has a specific gene on chromosome 6, which is related to Type 1 diabetes.

⁷ It is important to note that I distinguish between two types of inference. The first is mental inferences: imagine a case in which a person A has a certain disease D and does not want this to be known by a person B. Suppose that there is a set of symptoms S such that x has S iff x has D. If B learns that A has S, then the content of what she learned does not include A having D, but if B knows that x has S iff x had D, then learning that A has S is equivalent to learning that A has D. In terms of privacy, it does not seem to matter that information has flowed to B through mental inferences.

The second type is inferences conducted by artificial intelligence: imagine a case in which a person A has a desire to share a set of features {x1 (age), x2 (type of blood)} with a person B to determine whether s/he has disease C, while A does not want it to be known that s/he has a certain disease D. Person B uses an inference engine to derive a prediction based on shared data. The inference engine uncovers new correlations or patterns or confirms suspected relationships between data. Suppose that B uses two different inference engines to predict disease C and disease D based on x1 and x2. In the case in which disease C is inferred, no privacy loss occurs. In contrast, inferring disease D results in A's privacy loss. It is important to answer the question of whether A has a desire for B to use a specific type of inference engine to predict his/her disease.

In this case, Menges would argue that no privacy losses occurred because Annabel is the right source of control over the information flow. I agree with Menges that no loss of privacy occurs in Case 3 because Annabel has a desire to share her information with Brian, and Brian infers information that Annabel has no desire to keep private.

In each case Menges (see Menges, 2020a, 2020b) discusses, he only focuses on information-sharing without taking into account what will happen when the shared information is analysed or processed. Accordingly, Menges considers the *origin* of the information flow to be important in determining whether a loss of privacy has occurred. As Menges emphasises, once the flow of a piece of information is grounded on the desire of agent A, whose information is shared with another agent(s) B, no privacy loss occurs (see the voluntary divulgence cases, Menges, 2020a). The focus on the origin of the information flow, I believe, implies that, according to Menges, A can be the right kind of source for the flow of information inferred from an initial piece of information only if A is the right kind of source for the flow of that initial information. Thus, the flow of information which results from an intentional action by A does not lead to a loss of privacy, regardless of any information that may be inferred from it. I argue, however, that this feature of Menges's theory—that it is indifferent to potential inferences—gives rise to a counter-example. Consider the following:

Case 4: This case is identical to Case 3 (Annabel donates her DNA for research purposes), with the only difference being that Brian infers from Annabel's DNA profile that she suffers from her rare genetic disorder.

For Menges, if A is the right source of control over P, then their privacy is not lost. Concentrating merely on the origin of the information flow, as Menges does, implies that information P* inferred from other information P can never be privacy-diminishing if P is not. Hence, it might be argued that, in Case 4, Annabel has a desire to share her information P with Brian, so inferring P* from that information does not lead to her loss of privacy in Menges's view. However, I note that, if P* follows from P in some sense, then P* should be privacy-diminishing under some circumstances; this is a property that I think must be clarified in Menges's view. In Case 4, Annabel has a desire to share her information with Brian for the defined purpose, but she does not have a desire to share some potential information inferred from her information which does not comply with the initial purpose. Hence, Annabel's privacy is, in my view, essentially lost.

It might be argued that if Annabel does not want this information (P*) to be shared, then her privacy is diminished in Menges's view. She has lost source control. I agree

that her privacy is lost, but the reason for that cannot be grounded on Menges's view because Menges merely argues for having an initial desire to share information with others and does not discuss a person's desire to infer information from that shared information. I argue that Annabel's privacy has been lost because information that Annabel does not desire Brian to have ultimately flows to him. The desire related to the information inferred is not clarified in Menges's view.

As mentioned in previous paragraphs, the problem with Menges's view is that the propositional content of the relevant information-releasing desire does not include the content of the inference. Therefore, the initial version of the source control view of privacy is wrong about Case 4 because Annabel's privacy is diminished, although she has an intention to share her information, and the origin of the information was grounded on her desire. Hence, *Case 4* is a *counter-example* for the initial version of the source account of privacy. That gives me good reason to revise the initial version so that Annabel's privacy has, in fact, been lost in Case 4.

Comparing cases 3 and 4 illustrates that the initial account of source control of privacy can be revised by answering the question of what makes something the 'right' source information flow. Although Menges uses desire as the standard example of how to conceptualise his view, he also notes that he remains open to what exactly constitutes source control (Menges, 2020b, p. 37). If the desire or intention makes it the right flow, then is s/he the right source flow for that piece of information if A intends to keep P* private? How can one distinguish between cases in which B infers information from intentionally shared information that A intends to keep private (Case 4) and those of that A does not intend to keep private (Case 3)? In other words, it is important to determine what constitutes the relevant inferences that do not lead to a loss of privacy. A's intention determines how a piece of information flows to another agent. That is, if the flow of information changes, then A is no longer the right source of the novel flow of information. Thus, to identify whether drawing inferences (P*) from intentionally shared information (P) affects whether one is the right source of information, I focus on the flow of information, as any changes in the flow determine whether one is the right source of the information flow. I think a piece of information flows between different parties in a system to realise a specific purpose. Thus, a person whose data are processed and an agent who processes that data for a specific purpose play an important role in determining the flow of information. Thus, *who* engages in a system and their *purpose* for doing so determine the flow of information. It follows that any changes in these elements, which characterise the flow of information, will alter whether one is the right source control over the information.

As Case 3 illustrates, Annabel wants to know whether she suffers from a prevalent disease and desires to share her information with a medical research lab. Let us imagine that she should share her data with one of five institutes, some of which are public health bodies, and some of which are industry organisations. Annabel has a desire to share her data with a medical research lab. However, she has no desire to share her data with a ‘big pharma’ company. In this case, although Annabel should share her data and has no ability to do otherwise (i.e., preventing one institute from accessing her information), she still has privacy because she is the right source control over her information. It is important to note that, in order to avoid second-order conflicts, I assume that sharing data with a medical research lab does not imply sharing it with a big pharma company. Otherwise, I would have concluded that not only does she not want (first-order desire) to share her data with a big pharma company, but she might not have a second-order desire to share her data with a medical research lab because doing so implies sharing data with a big pharma company. Thus, Annabel’s desire determines *who* asks her question, and she does not allow the big pharma company to answer her question. Thus, the source control account of privacy does not have any problem with the first element that characterises the flow of information.

I now turn to the second element, namely the primary *purpose*, determining the flow of information. Annabel knowingly submits her DNA sample to the research lab to find the answer to her specific question. What matters, though, is that these data contain a significant amount of information beyond her specific question. Consequently, the researcher can infer more from that information beyond what Annabel specifically asked the lab to investigate (see Case 4). In such a case, the researcher (here, Brian) can not only look for the prevalent disease Annabel asked them to identify, but they can also study whether she suffers from a rare genetic disorder. That is the kind of excessive (unintended) information derived without any reason to do so. I consider this a loss of privacy even though Annabel initially had the desire to disclose her original data. Therefore, I argue, the initial purpose for which a researcher should carry out their task identifies whether the inferences derived from the information lead to a loss of privacy. Any information derived beyond the question diminishes Annabel’s privacy, as it diminishes being the right source control of information.

It is important to note that I do not claim that Annabel has the idea of the full knowledge that can be derived from her data. Moreover, I agree that the researcher may not necessarily know a priori what specific information the research requires.

However, I note Annabel and Brian can only agree on the very limited purposes and limited inferences. Any other (excessive) inferences that might be drawn from that information lead to a loss of privacy. I believe that the problem with the source control account of privacy is thus related to the second element, namely the primary purpose.

So far, I have discussed the important elements that determine the information flow. Any changes in the elements result in a novel flow of information. If the flow of information is not grounded on A's intention (or desire), then A's privacy is lost. As discussed above, processing data or accessing information in a manner that is incompatible with the initial purpose for which data were collected alters personal information flows. Regarding the fact that agent A's desire⁸ or intention, whether reasonable or unreasonable, prescribes the flow of information, any changes in the flow of information lead to a diminution of being the right source control over the flow of information. This suggests an adjustment to Menges's definition. The adjustment consists of adding that the loss of being the right source of information flow must be due to the action(s) of another agent who obtains or deduces information intended to be private. My revised definition is as follows:

Definition 2: A's privacy is lost iff: A has lost source control over the personal information P about agent A, if information flows at all, *due to the action(s) of agent B, who obtains or infers information contrary to A's preferences.*

As I have already discussed, Case 4 is a counter-example of the initial source control account of privacy, as Menges argues that Annabel's privacy has not diminished in this case. However, my revised version of this view is correct for Case 4 because it states that Annabel's privacy has, in fact, been lost. According to the revised account of privacy, in Case 4, Annabel is not the right source control because Brian changes the flow of information by changing the initial purpose and inferring excessive (unintended) information from Annabel's data. The initial purpose was to identify various factors related to a prevalent disease, while Brian changes this purpose and acquires excessive information related to her genetic disorder. Thus, the initial purpose, which should be realised in accordance with A's (reasonable or unreasonable

⁸ I believe that, in cases in which A has an unreasonable epistemic desire, their privacy is diminished, but it is not necessarily wrong. Consider A, who shares her blood sample with B for the purpose of identifying her blood type. If B uses A's information to make inferences that A is HIV positive, then A's privacy is diminished. In this case, A has an unreasonable desire that her disease will not be revealed through sharing her information with B.

epistemic) desire, is the key element in identifying whether the inferences scientists make change the information flow.

2.4. Macnish's Account of Privacy: Privacy as an Actual Access

Macnish (2018, 2020) defends the access account of privacy against the control account. The access account holds that, for a diminution of privacy to occur, the personal information must be actually accessed. Furthermore, the information accessed must be understood by the agent accessing it. The traditional access view is then supplemented by a semantic account that describes an agent's capacity to understand the information. Accordingly, if another agent B accesses personal information P about agent A without understanding its meaning, then A's informational privacy is not diminished. Macnish concludes that privacy diminution has occurred when the information is actually accessed by those who can understand it (Macnish, 2020, p. 17).

The access account of privacy holds that a loss of privacy occurs when a stranger reads my diary. For example, in Case 2, my privacy is diminished because another agent reads my diary and discovers information about me. Furthermore, this account of privacy holds that personal information which is intentionally shared with those who understand its meaning leads to a loss of privacy. A reduction in my privacy has occurred when I show someone a personal letter or invite them into my house (Macnish, 2018). Consider Case 1, in which I freely and knowingly ask the stranger to read the last entry of my diary. In response, the stranger reads it. According to the access account, my privacy is diminished when the stranger reads my diary in response to my valid consent for him/her to read the latest entry. Similarly, this view implies that we lose our privacy when we freely and knowingly tell our friends about our problems and secrets. According to this view, our privacy is diminished whenever someone else accesses personal information about us, regardless of whether we intend to share our personal information with another agent.

The discussions above indicate that the key idea, which is that the information in question must actually⁹ be accessed, is a necessary and sufficient condition for a loss of privacy. Given that a descriptive definition of privacy aims to specify what a loss of

⁹ The use of the word 'actual' is deliberate and clarifies that the privacy account discussed in this section is the access account developed by Macnish. The purpose of using this term is to emphasise that, to argue that actual access has occurred, it is necessary to understand what is accessed.

privacy consists of (Powers, 1996), the *initial* definition of the access account of privacy is the following:

Definition 3: A's privacy is lost iff: B *actually accesses* personal information P about A.

For Macnish, the fact that B actually accesses personal information P about A is a *sufficient* condition for a loss of A's privacy to occur. Macnish thinks that a loss of privacy has occurred because agent B had actual access to P and learned something new about A (see Cases 2 and 1). The actual access by another agent is thus sufficient for the loss of A's privacy to occur.

Moreover, for Macnish, the fact that B actually accesses personal information P about A is a *necessary* condition for a loss of A's privacy to occur. He cites the following case:

Case 5: Imagine that I have returned to the coffee shop after a 30-minute interval to find my diary on the table. It is unopened. I panic for a moment, but on seeing me, the stranger smiles and hands me the book. She explains that she has not opened it but saw me leave without it and collected it to await my return. She knows how intimate her own diary is, so she respected my privacy and kept it shut, as well as making sure that no one else would be able to read it. I feel an enormous sense of relief, thank her and leave with my dignity intact'. (Macnish, 2018, p. 420)

According to Macnish (2018), my privacy has not been lessened because the diary was not actually opened and read. The actual access by another agent to personal information is thus necessary for A to lose his/her privacy.

2.5. Revising the Actual Access Account of Privacy

Macnish (2018, 2020) contends that gaining access to A's personal information—for example, through a diary—leads to a reduction of privacy. I argue that a set of personal information consists of two different subsets of information: information about A that A intends to keep private and information about A that A intends to transmit or share with other agents. According to Macnish's view, when a stranger accesses the personal information I transmit, my privacy will be diminished (see Case 1). Moreover, as Macnish stresses, accessing personal information that A intends to keep private results in a diminution of privacy (see Case 6). Therefore, in Macnish's view, accessing both subsets of information leads to a loss of privacy.

Case 6: ‘Imagine that Eustace keeps a private diary. Eustace talks publicly about this diary, freely describing what it looks like but not about its contents, which he holds to be private. One day Frances is in Eustace’s office and sees the diary, recognising it from the description. She opens the diary and finds that she *can* read it. She reads through the diary and finds out that Eustace has been visiting George a lot recently. She does not realise it from the description, but Eustace and George are having a covert relationship. In this case, Eustace’s privacy has been diminished (Frances knows something about Eustace he would rather have been kept private). However, Eustace’s privacy has not been diminished as much as if Frances had been able to infer that he was in a relationship with George’. (Macnish, 2020, pp. 15–16)

In this case, Macnish (2020) argues that Eustace’s privacy has been lost because Frances had actual access to the information P about Eustace, which Eustace attempted to keep private. I agree with Macnish that accessing information that A intends to keep private leads to a loss of privacy.

I believe, however, that my privacy will not be diminished, as I intentionally shared my information with the stranger (Case 1); ‘I am including another within my realm or privacy, not lessening my privacy’ (Inness, 1996, p. 46). According to the actual access view, however, a diminution of privacy occurs even if an agent intentionally shares their personal information; instead, I contend that only accessing information P about A that A intends to keep private results in a loss of privacy. Nevertheless, accessing information that was once private but that A now intendedly¹⁰ shares with other agents does not lead to a loss of privacy. Thus, Case 1 is the *counter-example* for the initial version of the actual access account of privacy because this view incorrectly interprets Case 1.¹¹ This gives me good reason to revise the initial version of the actual access view.

I think the initial version can be revised by making a distinction between once-private, now intentionally shared information and information kept in private. I then

¹⁰ I assume that the person has privacy-preserving intentions to share some privacy-sensitive information with others. In cases where the person has non-privacy preserving intentions, there would be a reduction in their privacy.

¹¹ One plausible interpretation of Case 1 is that ‘my privacy is (voluntarily) diminished, but it is not important or morally wrong’. Proponents of such an interpretation might see privacy as being entirely neutral. I do not take that view, as I see privacy as *prima facie* good, although the discussion of its normative aspect goes beyond the scope of this chapter. I only claim that accessing once-private, now intentionally shared information is not a diminishment of privacy. I have a *prima facie* reason to object to any action that diminishes my privacy. However, I remain impartial on whether privacy diminishment is a necessary, sufficient, or criterial condition for a right violation. I solely emphasise that privacy depreciation is part of the analysis of whether the right to privacy is violated or infringed upon.

suggest excluding the subset of the once-private, now intentionally shared information from the set of private information. In this way, the scope of the actual access account of privacy is narrowed and only covers personal information which A intends to keep private. Accordingly, if an agent B understands the meaning of information about agent A, and A has intentionally shared it, then no privacy loss has occurred, because B has actual access to the information which was once private and is now intentionally transmitted, instead of accessing (intentionally) private information.

This suggests a new adjustment of the initial definition of the access account of privacy. The adjustment consists of adding that the actual access must occur when agent B accesses personal information P about agent A, which A attempts or intends to keep private. It is important to note that a set of personal information that A intends to keep private is a subset of personal information. This is the difference between definition 3 and definition 4 below. Hence, the problem is not related to learning something new about another person, but rather, understanding the information which A intends to keep private. My revised version is as follows:

Definition 4: A's privacy is lost iff: B actually accesses personal information P about A, and A intends that P remain private.

As I have already discussed, Case 1 is a counter-example for the initial version of the actual account of privacy, since Macnish argues that my privacy will be diminished if the stranger accesses my diary. However, my revised version of this view correctly interprets Case 1 by positing that my privacy will not be diminished in this case. According to the revised account, in Case 1, the stranger accesses once-private, now intentionally shared information, which does not lead to a loss of privacy.

So far, I have claimed that a loss of privacy occurs when agent B accesses personal information P about agent A, which A intends to keep P private, while no privacy loss occurs when B actually accesses P as long as P is intentionally shared. The question that may arise is how B realises that the piece of information accessed is private, or was once private and is now intentionally revealed. In responding to this concern, two different kinds of cases can be separated: first, cases in which B *knows* that the piece of information accessed is private and that A intends to keep it private; and second, cases in which B *does not know* either whether the piece of information accessed is private *or* whether A intended to share it or A was unaware that a piece of information could be accessed by B (Rumbold & Wilson, 2019).

In cases in which agent B knows that the piece of information accessed is private and A intends to keep it private, accessing P, and even any inferences from P, diminish

A's privacy. For example, in Case 6, Eustace intends to keep the information private that she is in a covert relationship with George, and she has never talked about her relationship with Frances. Thus, Frances actually accesses personal information about Eustace, which she intends to keep private, and, consequently, Eustace's privacy has been lost.

In cases where an agent B does not know that the piece of information accessed is private *or* whether it is a piece of private information inferred from P which A intended to share, or even that A was unaware that this piece of information could be accessed by B, two different responses can be considered. First, if B is unsure whether some is private or was once private and accesses it, this leads to a reduction in A's privacy. This response restricts any access to once-private information. In contrast, there might be P which A intended to share with B. This response prohibits all intentional analyses of once-private information. Second, B refrains from accessing information that they have *reason* to think was private and which A would have wanted to keep private (Rumbold & Wilson, 2019). In this way, A's privacy depends on what B could reasonably have expected A's concerns were with regard to the piece of information now accessed.

In the case of Annabel, Case 3, Brian might reasonably expect that Annabel wanted him to understand the fact that her DNA profile illustrates a specific gene structure related to a prevalent disease—simply because the information discovered does not deviate from the initial purpose for which the data were collected. Thus, Brian has reason to think that the piece of information accessed through analysis and inference is not information that Annabel wants to keep private. Accordingly, accessing this kind of information does not constitute Annabel's loss of privacy.

According to the above discussion, I claim that Annabel's privacy in Case 3 is not lost because Brian has reason to think that the information accessed is not the kind of information that Annabel wanted to keep private. However, the initial version of the actual access account, definition 3, argues that Annabel's privacy has been lost because Brian accessed private information about Annabel. Therefore, *Case 3* is the *counter-example* for the initial version of the actual account of privacy. I believe that this account can be revised again by adding the condition that B has reason to think that A wants to keep the information that has been accessed private. I suggest the below definition, which correctly interprets Case 3 by saying that Brian has reason to think that Annabel does not intend to keep the information that has been accessed private. Thus, Annabel's privacy has not been lost.

Definition 5: A's privacy is lost iff: B actually accesses personal information P about A, and A intends that P remain private, or, *B has reason to think that A intends to keep it private.*

The above definition indicates that privacy diminishment for A is not solely about A's personal decision, but also about the contexts in which they participate. In cases where B does not know whether the information accessed is private, they make a decision on behalf of A by giving a reason why accessing that information may or may not lead to a privacy diminishment for A. This means that the context can impact and affect A's privacy. Precisely, privacy has both personal and common characteristics.

Referring to the reasonable expectation in my revised version of the actual access account of privacy seems to link the descriptive aspect of privacy to norms and values, in that there is a clear set of normative values that explains what the reasonable expectation is in a certain situation. However, norms that are characterised as the reasonable expectation are different from the moral values to which the normative conception of privacy might refer. For example, reasonable expectations might refer to legal norms (purpose limitation, such as in Case 3) or cultural norms prevalent in society, which do not necessarily constitute a normative account of privacy, which is based on moral values and norms. Moreover, the descriptive account of privacy has other parts, namely actual access and source control, that are not solely values. Thus, the descriptive aspect of privacy considers multiple elements which are not reducible to the normative concept of privacy. That is why I believe that my analysis is still related to the descriptive aspect of privacy and is not reduced to the normative one.

2.6. Paradigmatic Cases

This section is the first piece of evidence that the revised views of control and access accounts of privacy are extensionally equivalent. To demonstrate this, I test the revised views on different sets of information to see which of the revised accounts explains the loss of privacy in the cases. I focus on the sets of information introduced by Rumbold and Wilson (2019). They provide abstract classes of information that can possibly be gained through analysis and inference regarding personal information P. In what follows, I categorise each of the cases explored in the previous section according to the classes of information provided and analyse how the revised accounts explain whether privacy is diminished. The sets of information are as follows:

- Public information which has always been public,
- Private information an agent intends to remain private (Case 5),
- Once-private information an agent has intentionally shared (Case 1),

- Once-private information an agent has no intention of sharing of that they are unaware was shared (Case 2),
- Information inferred from once-private information that an agent has intentionally shared and which itself counts as a piece of information that the agent intended to share (Case 3),
- Information inferred from once-private information that an individual has intentionally shared but which does not count as a piece of information that the agent intended to share (Case 4),
- Information inferred from shared information that an agent has only shared unintentionally (Case 6).

Both source control and actual access accounts hold that not all information is subject to privacy concerns. Privacy does not concern any information about agent A. It is not a loss of A's privacy if we discover that s/he wears glasses (public information; Macnish, 2018). As a result, losing source control over or accessing information that has always been public does not lead to privacy loss.

As discussed in the previous sections, both revised accounts argue that A's privacy has not been diminished in Cases 1 and 3. Previously, I also argued that A's privacy in Case 5 was not lost based on the revised version of the actual access account. Furthermore, the source control view argues that no privacy loss has occurred in Case 5. Thus, A still could be the right source control of information, and no loss of privacy has occurred.

By contrast, as I have already discussed, both revised versions of the source control and actual access views posit that A's privacy is lost in Case 2. Previously, I also argued that A's privacy was diminished in Case 4 based on the source control view. In addition, the revised access account states that A's privacy is lost in this case. Although B has reason to think that A intends to keep the information accessed private, B accesses that information, resulting in a loss of A's privacy. Furthermore, concerning the actual access account, I have already stated that, in Case 6, A's privacy is lost. Moreover, the source account of privacy argues that A's privacy is lost because the information inferred about A flows without A being the right kind of source of this flow, resulting in a diminution of A's privacy.

The results of the test of the revised versions of source control and access accounts of privacy on whole comparable cases are presented in Table 2.1. The first two columns highlight the differences between the initial accounts of privacy in answering the question of whether privacy is lost. The grey cells in the last two columns of the

table indicate cases in which the initial and revised versions of the accounts have different answers regarding the loss of privacy, providing a contrast between the two. Comparing the initial and revised accounts shows the changes that have been made.

Table 2.1. Compare and Contrast Different Versions of the Source Control and Actual Access Accounts of Privacy

	Initial version of source control account of privacy	Initial version of actual access account of privacy	Revised version of source control account of privacy	Revised version of actual access account of privacy
Case 1	no privacy loss	privacy loss	no privacy loss	no privacy loss
Case 2	privacy loss	privacy loss	privacy loss	privacy loss
Case 3	no privacy loss	privacy loss	no privacy loss	no privacy loss
Case 4	no privacy loss	privacy loss	privacy loss	privacy loss
Case 5	no privacy loss	no privacy loss	no privacy loss	no privacy loss
Case 6	privacy loss	privacy loss	privacy loss	privacy loss

Both revised versions of the source control and actual access accounts of privacy give the same answers to the question of whether privacy is diminished (see the last two columns above), while they provide different answers as to why it is diminished. Moreover, these revised versions are located somewhere between the initial ones. As Table 2.1 shows, according to the initial version of the Menges's account of privacy, privacy loss is rare; it is lost in two of six cases, while, according to the initial version of the Macnish's account of privacy, privacy is lost often, in five of six cases. Nevertheless, in the revised versions of both accounts, there is a loss of privacy in three of six cases. Therefore, I claim that these initial versions are two poles on a continuum, with intermediate forms in between.

2.7. Theoretical Argument

So far, I have tested the proposed views of privacy on paradigmatic cases. The test revealed that there is no case in which a person loses control over the information flow due to the actions of another if the personal information, which A intends to keep private, is not accessed. Furthermore, there is no case in which private information is accessed such that the person does not lose source control. The paradigmatic cases give us a practical reason to think that these proposed views of privacy are extensionally equivalent. This section, meanwhile, gives us the theoretical reason to think this is the case.

On the source control front, I claim that no loss of source control occurs when the information is not accessed, or that A loses source control of P when B actually accesses information P about A. The kind of control defended in the source control account of privacy, I believe, is not robust enough, which means that agent A is not in a position to decide whether B accesses his/her information or to stop B from accessing his/her personal information because A does not have the ability to do otherwise. A loss of privacy does not occur when B accesses information P about A, but rather when B actually accesses some information which A intends to keep private. Since obtaining information about A in a way which is contrary to A's preferences results in a loss of A's source control over P, actual access to information P about A results in a loss of source control of P. Thus, the distinction between the views collapses.

On the actual access front, I claim that no actual access is achieved such that A does not lose source control, or B actually accesses information P about A when A loses source control of P. A loses source control of P when P flows in a way that is not grounded on A's intention or desire. Since actual access occurs when B accesses information that A intends to keep private, I conclude that the loss of source control leads to actual access. Moreover, it is impossible for A to lose source control of P while B has not accessed P. B not accessing P means that P has not yet flowed. If no one has actual access to the information, A can still remain the right source control of information. If the information does not flow, that is, no actual access occurs, only leeway control is lost, and A remains the right source control of information. Thus, when A is not the right source of control over their information P, having actual access to P diminishes A's privacy. I conclude that actual access does not diminish privacy if the access relates to A in the appropriate way. Thus, the distinction between the views collapses.

The preceding discussions show that the source control and actual access accounts of privacy are extensionally equivalent but different in their underlying rationales. In other words, these are two formulas that lead to the same result: a loss of privacy. The implications for such a difference in underlying rationales can be discussed in relation to the normative aspect of privacy, when privacy matters morally. Given that the main goal of this chapter is to focus on the descriptive aspect of privacy, I briefly explain its implications. As both accounts of privacy are extensionally equivalent, I view them as two perspectives that reach the same peak of a mountain, and I see the value of privacy as a cluster that encompasses the values represented in both accounts. Privacy is itself a cluster of values that intersects with the cluster of values that comprise control accounts, such as autonomy and individual liberty (see, e.g., Rössler, 2005), and also with the cluster of values comprising access accounts, such as secrecy and anonymity (see, e.g., Gavison, 1980). By perceiving privacy as a cluster of values, we can take a pluralistic approach that encompasses all the values in the cluster to understand the normative aspect of privacy. This means that we take into account all different values of privacy in order to form a more comprehensive and inclusive understanding of its normative aspect.

2.8. Conclusion

This chapter offered new insight into the debate about the nature of privacy. There is persistent disagreement in the literature on privacy's proper meaning and definition. However, the two definitions that are prominently discussed in the literature are 'control' and 'access'. Control (Inness, 1996) and limited access (Gavison, 1984) accounts of privacy have recently been developed by identifying the kind of control that is relevant to determining whether a person has privacy with regard to certain information and by incorporating a semantic account into the limited access account of privacy. Source control (Menges, 2020a, 2020b) and actual access (Macnish, 2018, 2020) accounts of privacy are these most recent versions. Because they are the most in-depth version of the classic accounts of control and access, they were chosen as the focus of this chapter. However, the debate over which account provides the proper definition of privacy, which is presented in the traditional control and access views, persists in the most recent versions, as well. In this chapter, I demonstrated that the revised versions of the source control and actual access of privacy are extensionally equivalent. First, I discussed these views are extensionally equivalent when applied to various test cases. They only differ regarding the explanation of why privacy is

diminished. Second, from a theoretical perspective, the relationship between source control and actual access views is equality, meaning that the extensions of these views are equivalent, while the differences between these two can metaphorically be explained by referring to different sides of the same mountain.

Chapter 3:

How Does an Artificial Intelligence System Affect Privacy?

Adopting Trust as an Ex Post Approach to Privacy

Abstract

This research explores how a person with whom information has been shared and, importantly, an artificial intelligence (AI) system used to deduce information from the shared data contribute to making the disclosure context private. The study posits that private contexts are constituted by the interactions of individuals in the social context of intersubjectivity based on trust. Hence, to make the context private, the person who is the trustee (i.e., with whom information has been shared) must fulfil trust norms. According to the commitment account of trustworthiness, a person is trustworthy only if they satisfy the norm of competence. It is argued that a person using an AI system to answer a question is competent only if they are *ex post* justified in believing what has been delivered by the AI system. A person's belief is justified in the doxastic sense only if the AI system is accurate. This feature of AI's performance affects a person's competence and, as a result, trustworthiness. The effect of AI on trust as an essential component of making the context private, and thus on privacy, means an AI system also impacts privacy. Therefore, a private context is constituted when the individual with whom the information is shared fulfils the competence norm and the AI system used for analysing the information is sufficiently accurate to adhere to this norm. The result of this research emphasises the significance of the relationship between individuals involved in information-sharing and how an AI system used for analysing that information impacts the relationship regarding making the context private, as well as how it impacts privacy. The findings of this research have significant implications for improving or ameliorating privacy regulations in light of trust.

Keywords: AI accuracy; competence; social value of privacy; trust norms; trustworthiness

This chapter is a modified version of the following publication:

Asgarina, H. (2023). Ex post Approaches to Privacy: Trust Norms to Realise the Social Dimension of Privacy. *International Conference on Computer Ethics, I(1)*, Article 1. <https://soremo.library.iit.edu/index.php/CEPE2023/article/view/287>

This chapter is published as:

Asgarina, H. (2024). Adopting trust as an ex post approach to privacy. *AI and Ethics*, 3(4). <https://doi.org/10.1007/s43681-024-00421-w>.

3.1. Introduction

Privacy can be understood as a social construction we create as we negotiate our relationships with others on a daily basis. By placing privacy in the social context of intersubjectivity (Steeves, 2009), privacy is conceived as a dynamic process regulating interpersonal boundaries by drawing a negotiated line between openness and closedness to others (Altman, 1975). The dialectical approach to privacy, in which privacy can only be obtained through the negotiated interaction between social actors, captures its importance as a social value (Altman, 1975; Steeves, 2009).

The dialectical approach to privacy neglects that privacy is a social phenomenon not only because other people exist, but also because privacy concerns the social circumstances in which information flows from one party to another. The contextual integrity model of Nissenbaum (2010) elaborates on socially embedded privacy in the digital age. Nissenbaum argues that different social contexts are governed by different social norms that govern the flow of information within and outside of that context. Protecting privacy entails ensuring the appropriate flow of information between and among contexts. Privacy is a norm that regulates and structures social life (Nissenbaum, 2010).

According to Waldman (2015), although Nissenbaum (2010) succeeds in the socialising theory of privacy in terms of social interactions and the possibility for individuals to be properly embedded in social relationships, it begs the question of what a 'private context' is. Waldman responds to this question by arguing that 'private contexts are defined by relationships of trust among individuals' (Waldman, 2015, p. 559).

Drawing on the insights of Waldman (2015), a private context is constituted by relationships of trust among the individuals involved in the context. The interaction of different individuals in the social contexts of intersubjectivity based on trust constitutes privacy. Privacy is a social construction we cannot have unless we work together, which is what Altman (1975) and Steeves (2009) argue. Interpersonal trust depends upon the nature of relationships between individuals, social circumstances, and context. Since privacy depends on trust, such social circumstances are associated with the value of privacy as well, as Nissenbaum (2010) considers in her socialising theory of privacy.

Ex post approaches discuss privacy when information is shared and revealed between different individuals in a context. Trust as an *ex post* approach to privacy, as highlighted by Waldman (2015), emphasises the role of individuals in constituting a private context. Accordingly, privacy scholars have been working on trust norms and

have regulated trust-promoting norms that govern the relational duties of trustee parties (i.e., the person who is trusted) regarding how to build and cultivate trust-based relationships with trustors (i.e., the one who trusts), thereby making the context suitable for disclosures. Richards and Hartzog (2020), for example, have identified trust norms, such as protection, discretion, honesty, and loyalty.

This chapter adopts a philosophical perspective to identify trust norms, which differ from those in, for example, the work of Richards and Hartzog (2020). From a philosophical perspective, as this chapter argues, competence must be considered a norm to be trustworthy. Thus, the norm of competence must be included in the list of trust norms that Richards and Hartzog (2020) have proposed.¹² Moreover, this chapter emphasises the role of AI systems in establishing a person's trustworthiness and in contributing to making the context private. This chapter explores the significance of AI in contributing to B's trustworthiness and thereby constituting private contexts, a topic that has not been given adequate attention in the literature.

For a clearer understanding of cases in which both individuals and an AI system are involved and information is shared and revealed, consider the following case:

A person (B) uses data (q) about another person (A) to predict whether she has breast cancer (p). B cannot deduce if q then p ($q \rightarrow p$) because of his limited background knowledge. To deduce if q then p, B relies on a machine learning (ML) model to identify the possible presence of breast cancer for A. Such an ML model has displayed the potential to predict whether A develops breast cancer within certain timeframes by analysing her electronic health records and mammography patterns (Akselrod-Ballin et al., 2019). The deliverance of the ML model is a proposition in response to the following question: 'Is breast cancer present?'

Trust as an *ex post* approach to privacy emphasises the role of B, as a person trusted by A, in constituting a private context. In addition to B acting as the trustee, does the ML model, which is used to predict whether p, contribute to making the context private? This chapter argues that, yes, ML models that predict aspects such as the presence of breast cancer contribute to making the context private. Furthermore, the ML model impacts trust relationships between A and B. Since privacy depends on trust, the ML model contribute to making the context private, ultimately impacting privacy considerations. Therefore, adopting trust as an *ex post* approach to privacy not

¹² Unlike Richards and Hartzog (2020), who consider privacy solely as secrecy and formulate trust norms based on this definition, I conceive of privacy as a social phenomenon constituted by trust, and I formulate trust norms beyond merely secrecy.

only emphasises the role of the trustee, as Richards and Hartzog (2020) highlight, but also the role of AI systems in constituting private contexts.

The main purpose of this research is to investigate how an ML model affects privacy. Since ML models are the outcomes of employing AI systems, the main research question (RQ) is as follows: ‘How does an AI system affect privacy?’ To respond to this, I formulate two sub-questions (SQs): 1. ‘How do A and B cultivate or maintain relationships of trust?’ 2. ‘How does an AI system affect trust relationships between A and B?’ Answering these two SQs provides the foundation for answering the main RQ. The SQs are addressed in Sections 3.2–3.3, respectively. Each response forms a premise for the argument that concludes with an analysis of the impacts of AI on privacy. The assumptions and premises of the argument that I formulate in this chapter are presented below.

I assume that privacy is constituted by the interaction of different individuals in the social context of intersubjectivity based on trust. Privacy, in a disclosure context in which information is shared and revealed, can thus metaphorically be conceived as a realm constituted by trust-based relationships. Hence, cultivating trust in a context is essential to making that context private. Additionally, I focus on cases in which A shares data (q) with B to answer a specific question, and B responds to the question based on the AI-delivered proposition p. As a result, the particular task that B is relied upon to perform is to assert p.

Section 3.2 addresses SQ1, which establishes the first premise of the argument. It is argued that, to promote trust in a context, A and B must conform to trust norms. B can be trustworthy while avoiding unfulfilled commitments. Given that promise-making norms are the most explicit mechanism by which B takes a (new) commitment, norms of being trustworthy derive from norms regarding promise-making. Competence is one of the norms of promise-making (Hawley, 2019). As a result, trusting B’s words involves relying upon him to fulfil promise-making norms, including the norm of competence. Section 3.3 addresses SQ2, which forms the second premise of the argument. It is argued that an AI system affects B’s competence and, thus, the trust relationships between A and B. Finally, given that AI affects trust, which is a constituent component of privacy, Section 3.4 concludes that AI affects privacy. Since trust requires an accurate AI system, privacy does also.

3.2. Trust

How do A and B cultivate or maintain relationships of trust? To promote or preserve trust in the context, A and B must conform to trust norms. To identify trust norms, I consider interpersonal trust rather than trust in a group or institutional trust, and I adopt four assumptions.

First, trust is a three-place relationship involving two people and a task. According to the majority of the literature (Baier, 1986; Hardin, 2002; Hawley, 2014, 2019; Hieronymi, 2008; Holton, 1994), trust is generally a three-place relation: A trusts B to ϕ . A primarily trusts B to do some particular thing rather than trusting him in general and in every way. Second, I focus on the norms of trust from the trustee's side. Norms of trust arise between two parties: a norm to be trusting in response to the invitation to trust and a norm to be trustworthy in response to the other's trusting reliance (Fricker, 2018). The former norm lies on the trustor's side, and the latter on the trustee's side (Carter & Simion, 2021). In this chapter, I discuss the norms of trust on the trustee's side and the conditions that give rise to trustworthiness in three-place relations. Third, I adopt doxastic conditions on trust. According to doxastic accounts, trust involves a belief on the part of the trustor. When A trusts B to ϕ , A *believes* that B will ϕ (Hieronymi, 2008). Fourth, like most philosophers, I distinguish trust from mere reliance. Trust involves reliance 'plus some extra factor'. Controversy surrounds this factor, which generally concerns why the trustor would rely on the trustee to be willing to do what they are trusted to do (Hawley, 2014, p. 5).

Regarding the first assumption, trust can be a two-place or a three-place relationship. It is a relationship between a trustor and a trustee in the first instance, as in A trusting B. Two-place trust, as opposed to three-place trust, is fundamental, according to Faulkner (2015). Two-place trust is a rather demanding affair; when we state that A trusts B *simpliciter*, we ascribe A a rather robust attitude, one in which A trusts B in several respects. A three-place relationship, on the other hand, is a less-involved affair: when we state that A trusts B to do ϕ , or that A trusts B with a valued item C (Baier, 1986), we do not need to express much about their relationship. According to Carter and Simion's views in 'The Ethics and Epistemology of Trust' (2021), this difference is maintained when we focus on the trustee's trustworthiness. One can be trustworthy in general, but one can also be trustworthy regarding a particular matter. I think of a trust-based relationship as a three-place relation between two people and a task. Considering three-place trust to be a general relation of trust, B can be trustworthy with regard to a particular matter but not generally. For example, B can be trustworthy in keeping a meeting appointment but may not be

trustworthy overall. With respect to the case discussed in this chapter, B can be trustworthy with regard to the task of assertion.

According to the second assumption, I only consider the norms of trust on the trustee's side. Addressing the third assumption, in discussions regarding the rationality of trust, or whether the trust is appropriate or well-founded, it is crucial to explore whether trust essentially involves belief. Proponents of non-doxastic accounts, such as Holton (1994), argue that it is not essential for trust to involve a belief about the trustee, such as a belief that they are trustworthy. Jones (1996), for example, maintains that the trustor must have an affective attitude which is not described by belief. Trust involves affective attitudes that may lead to corresponding beliefs. Hence, the rationality governing trusting is distinct from rational belief. However, proponents of doxastic accounts, such as Hieronymi (2008) and Hawley (2019), argue that trust involves a belief on the part of the trustor. Hence, if trust is a belief, the rationality that governs trusting is drawn from rational belief. To the extent that the trustor is rationally entitled to believe that the trustee is trustworthy with respect to ϕ , the trustor thereby has an entitlement to trust the trustee with respect to ϕ (Carter & Simion, 2021). I defend a doxastic account of trust mainly because it requires less explanation as to why trusting someone would give us a reason to believe what they say; 'trust gives a reason for belief because belief can provide a reason for belief' (Faulkner, 2017, p. 113). Although discussions of the entitlement to trust and the rationality of trust are important, I do not address them because these subjects are more related to trust norms on the trustor's side than on the trustee's side. I simply assume that, for trust to be well-grounded, the trustee must be trustworthy.

Is it required for A to have evidence of B's trustworthiness to be entitled to trust B? According to Hinchman, A's trust in B is reasonable even if A has no evidence of B's trustworthiness on the relevant matter, but it is not reasonable if A has good evidence of B's untrustworthiness on that matter. It is in line with the externalist approach to trust that the trustor need not have access to or be aware of the evidence (Hinchman, 2005, p. 580). I agree with Hinchman's (2005) point that reasonable trust does not require evidence of B's trustworthiness to be available to the trustor. Again, while the rationality of trust is important, most discussion on it is focused on the trustor's side.

Finally, concerning the fourth assumption, Baier (1986) provides an influential account of trust. According to her, trust must be distinguished from mere reliance. Although we can rely on both people and inanimate objects, not everything can be genuinely trusted. Trust differs from mere reliance because, when an object breaks,

one may be disappointed, but one does not feel betrayed. However, when we trust and are let down, we feel betrayed. As expressed in Hieronymi's (2008) theory, trust requires something more than merely relying on someone to do something; it requires a vulnerability to betrayal if let down.

Most philosophical theories of trust (Baier, 1986; Hardin, 2002; Hawley, 2014, 2019; Holton, 1994) are explicitly designed to explain that trust is a form of reliance, but it is not mere reliance; rather, trust involves reliance 'plus some extra factor' (Hawley, 2014, p. 5). Different theories associate this extra factor with the motives of the trustee. If A trusts B to ϕ , then A relies upon B to ϕ ; moreover, A assumes B has the right motive for ϕ -ing (Baier, 1986; Hardin, 2002). Those theories that dispute what type of motive the trustee should have to make trust appropriate are classified as 'motive-based' theories (Hawley, 2014). The other category of theories associates the extra factor with the trustor's particular stance towards the trustee (Hawley, 2014, 2019; Holton, 1994). These theories are classified as 'non-motive-based' theories, according to McLeod (2021). In what follows, I explore whether motive-based or non-motive-based theories succeed in explaining the conditions that give rise to trustworthiness.

I begin my argument by considering the task that B is relied upon to perform in general as ϕ . In sub-section 3.2.2, I specify ϕ . Regarding this, those who are concerned with the task of assertion might skip the first three sub-sections and move to the last one.

3.2.1. Motive-Based Theories on Trust

According to motive-based theories, the conditions that lead to trustworthiness are based on the motivation a trustworthy person has. Goodwill or self-interest are two examples of such motivations.

A trustworthy person is motivated to act by virtue of their goodwill towards the trustor. According to Baier (1986), when we trust someone, we rely on them having goodwill towards us. However, Holton (1994) argues that Baier's goodwill account of trustworthiness is not absolutely correct. Primarily, relying on a person's goodwill towards oneself is not a sufficient condition for trust. A confidence trickster might rely on your goodwill without trusting you. Second, goodwill is not a necessary condition: I can trust a person without relying on their goodwill towards me. I can, for instance, trust someone to look after a third party without requiring them to have goodwill towards me.

Another motive-based theory describes trustworthy people's motives in terms of self-interest, such as in the encapsulated interests account of Hardin (2002). He contends people trust those they believe have strong reasons to act in our best interests. He claims the primary motivation of individuals we trust is to preserve a relationship with us. Trustworthy people are motivated by their own interest in maintaining the relationship they have with the trustor, which motivates them to encapsulate that person's interests in their own.

McLeod (2021), however, provides an example to demonstrate why Hardin's (2002) theory is flawed. Consider a sexist employer who is interested in maintaining relationships with female employees and treats them fairly but whose interest derives from a desire to keep them around to daydream about having sex with them. This interest conflicts with the women's interest not to be objectified by their employers. At the same time, if the women were unaware of his objectification of them, he could ignore this particular interest of theirs. He can maintain his relationships with them while ignoring their interest in not being objectified, and encapsulating enough of their other interests in maintaining a good relationship in his own. This situation, according to Hardin, would make him trustworthy. However, if the women knew the main reason for their employment, they would not find him trustworthy. Being motivated by an interest to maintain a relationship may not require adopting all the trustor's interests to be considered trustworthy by that person.

Although motive-based theories are not limited to goodwill and self-interest theories, these are the dominant viewpoints in the literature. However, since these theories do not provide an appropriate account of trustworthiness, we need other theories that identify conditions for being trustworthy that are not driven by goodwill or self-interest.

3.2.2. Non-Motive-Based Theories on Trust

The conditions that lead to trustworthiness reside in the stance the trustor takes towards the trustee (McLeod, 2021). One can be trustworthy while avoiding unfulfilled commitments, regardless of one's motivation for fulfilling commitments. A relies on B to ϕ because A believes B has a commitment to ϕ -ing (Hawley, 2014).

Holton (1994), like Baier (1986), distinguishes between trust and mere reliance. However, unlike Baier, he does not suggest that, when we trust someone, we rely on them to have goodwill towards us; instead, when we trust someone, we take a particular stance towards them, which is the participant stance. Holton highlights that, in addition to resentment and gratitude, the feeling of betrayal is one of what Strawson

(1974) calls the reactive attitudes. We normally take these attitudes towards people but not towards objects. Behind these classes of attitudes is a more general attitude, which Strawson calls the participant attitude and Holton calls the participant stance. The participant stance is a particular reactive attitude we take towards those we regard as responsible agents. When we interact with someone who provokes a reactive attitude, whether resentment or gratitude, we adopt a particular attitude that is bound with the ascription of responsibility towards them. According to Holton (1994), trust is a reliance on the participant stance: trust involves something like a participant stance towards the trustee. Despite Holton's (1994) correct identification of the participant stance as a required component of trust, Hawley (2014) finds Holton's theory unsatisfying because relying upon someone to whom you take a participant stance does not always entail trusting them; some interactions occur outside the realm of trust.

According to Hawley's (2014) view, which she elaborates on in her book *How to be Trustworthy* (Hawley, 2019), it is reasonable to trust someone to do something only if that person has an explicit or implicit commitment to doing it. To trust someone to do something is to believe they have a commitment to doing it, and to rely upon them to meet that commitment. To make her account plausible, Hawley employs a very broad notion of commitment. Commitments can be implicit or explicit, weighty or trivial, conferred by roles and external circumstances, default or acquired, welcome or unwelcome. Hawley's account of trustworthiness in the context of the commitment, in terms of avoiding unfulfilled commitment, has nothing to do with the trustee's motives. To be trustworthy in some specific respect, it is enough to behave in accordance with one's commitment, regardless of motive. One person may trust another to do something without believing them to be motivated by their commitment (Hawley, 2014, pp. 10-11,16). In what follows, I adopt the commitment account of trustworthiness and identify norms to be trustworthy in response to the other's trusting reliance.

3.2.2.1. The Commitment Account of Trustworthiness

According to Hawley (2014, 2019), commitment is at the centre of the notion of trust. The most explicit mechanism through which we take on (new) commitments is promise-making. When thinking about promises and trust, two questions arise: First, how do we decide whom to trust? Second, whose promises do we accept and rely upon? The first question is from the perspective of the promise-receiver, whereas the second is that of the promise-giver. The following argument focuses on the second

perspective and answers the following question: ‘What do good promisors do?’ In Hawley’s (2019) view, good promisors not only keep their promises, but they also make appropriate promises in the first place. Making a good promise requires a sincere intention, the permissibility of the action promised, and the competency to keep the promise. Hence, the norms regarding promise-making are sincerity, promising to act morally, and competence. Among these norms, I focus on competence as it is impacted by AI, a topic that is discussed in Section 3.3.

A good promise requires competence to keep the promise, which is a norm of promise-making: do not make promises you are not competent to keep. ‘Competences are dispositions of an agent to perform well’, and they have three components: constitution, condition, and situation (Sosa, 2010, p. 465). Similarly, the competence required to keep a promise includes these three components (Hawley, 2019). After explaining competence, I return to the competence norms for promise-making.

Consider colour vision competence in Sosa’s (2010) paper on ‘How Competence Matters in Epistemology’, for instance. A constitution competence includes rods and cones; a condition competence includes being awake and sober; and a situation competence includes adequate light. When a person’s visual systems are fully functional, they are awake and alert and they see the object in plain view, exercising colour vision competence. Not only does a person need competence in colour vision, they also need the competence to assess the required conditions and situation of the proposed competence—second-order assessment. According to Sosa (2010), then, an agent’s success relies not only on their constitutional competence, but also on their being in an appropriate shape while appropriately situated. Thus, an internal constitution, being in good shape to exercise that competence, and external circumstances are required if the performer is to be properly credited with complete competence.

Analogously, the competence required for good promise-making should encompass all three components. In Hawley’s (2019) view, constitutional competences include a steady, reliable capacity to achieve success. More precisely, I argue the notion of competence is close to exercising a reliable intellectual capacity to form a justified belief. That is, keeping the promise manifests competence when forming an epistemically justified belief that one will keep the promise to ϕ . The following paragraphs further clarify what the constitution competence of promise-making involves.

The second component of competence, the condition competence, requires a person to be awake, alert, and sober when making a promise. The third component

of competence, the situation competence, indicates that what we are competent to do depends on the external circumstances, including the physical environment, social environment, and material resources we find ourselves in. Therefore, to incur a certain commitment, we require insight not only into our capability or underlying skills, but also into an actionable feature of our environment. For instance, it is far more difficult for a doctor working in a field hospital than it is for someone working in a well-equipped hospital to save a child's life. In a challenging environment, the situation competence required for success differs from in an easy environment. Acting in different environments requires a doctor to use different competences, some of which are more difficult to develop and maintain than other. Therefore, a person who makes a commitment needs to be aware of the circumstances in which they will need to act (Hawley, 2019).

I have described how a person being competent to promise to ϕ depends on being in good shape while making a promise and the complex facts regarding their physical and social environment. Now, I return to the first component of competence: constitutional competence. To possess a corresponding competence to keep a promise, I argue one should be 'justified' in thinking they will keep a promise when making one. Goldman (1979) distinguishes two uses of 'justified': an *ex post* use and an *ex ante* use. The *ex post* use occurs when there exists a belief, and we say whether that belief is justified. The *ex post* or doxastic sense of justifiedness applies to beliefs that a subject actually holds, rather than beliefs they could hold. In contrast, the *ex ante* use occurs when no such belief exists, or when we wish to ignore the question of whether such a belief exists. The *ex ante* or propositional justification applies to a proposition (p), a subject, and their epistemic situation. If we say that a subject is propositionally justified regarding p , we mean that it would be appropriate for them to believe p ; it is applicable even if they have no belief in the specified proposition (Goldman, 2015b). Since I argue it is inappropriate to promise to ϕ while one does not possess a belief in ϕ , I use an *ex post* or doxastic sense of justifiedness. Therefore, I articulate a good promise regarding satisfying the competence norm is one in which the promisor *in fact* believes they will keep the promise, rather than believes it is *possible* to keep the promise, and their belief is *justified*.¹³

In the scholarly literature, there are different theories of doxastic justification, such as *mentalist evidentialism* and *process reliabilism*. Since I adopt the externalism approach in

¹³ In my view, a person is rationally permitted to perform ϕ only if the person has a justified belief in ϕ .

this chapter,¹⁴ I focus on the process reliabilism theory of justification. The justificational status (J-status) of a belief, according to this theory, depends on how it is formed, or caused. As the theory indicates, how a belief is causally produced is crucial to its J-status (Goldman, 2015b). Consequently, the competence required for making a promise includes a capacity to form a belief based on a reliable process. The reliabilist principle of justification can be explained as follows:

‘(R) A belief B (at time t) is justified if and only if B (at t) is the output of a series of belief-forming or belief-retaining processes, each of which is either unconditionally or conditionally reliable, and where the conditionally reliable processes in the series are applied to outputs of previous members of the series.’ (Goldman, 2015b, p. 35)

I argue that competence includes a reliable capacity to form a justified belief to achieve success in what is promised. When a person promises to ϕ when they lack such a capacity, they make a wrong promise. Consider the following example provided by Hawley (2019). A child, Cindy, is brought to the hospital with a sever and an unfamiliar condition. The junior doctor in charge of the case, Jack, promises the parents he will save their child’s life, and he sincerely intends to do so. Cindy’s condition can be treated with a certain type of antibiotic, which Jack happens to try first, saving Cindy’s life. In this case, the junior doctor genuinely intends to save Cindy’s life, which is morally permissible. Jack keeps his promise but only through sheer luck, rather than through his competence. He does not have a justified belief he would keep the promise. Therefore, his promise counts as over-promising. For simplicity, I presume the doctor was awake, and I do not consider whether he was at risk of lacking the situation competence. I only concentrate on the requirement not explored in depth in Hawley’s (2019) description, which is having a justifiable belief in accomplishing the promised action or activity.

Even if Jack believed he would save Cindy’s life, his belief, in the doxastic sense of justifiedness, would not be justified. Although Jack has no outstanding skills regarding diagnosing and treating such conditions, he promised he would save Cindy. This promise was merely wishful thinking, but it made him confident. According to Goldman (2015b), wishful thinking is a highly flawed thought process. Forming belief through wishful thinking is unjustified, meaning Jack’s belief was unjustified. Since competence includes being justified in believing what is promised, Jack was

¹⁴ As I mentioned in Section 3.2, I suppose it is not required for A to access the reasons contributing to B’s trustworthiness. Since I take an externalist reading of reasons to believe B’s trustworthiness on the trustor’s side, I do the same when analysing competence norms on the trustee’s side.

incompetent in this case. However, as Hawley (2019) points out, a lack of suitable competency does not imply incompetence in the normal sense. Jack was as competent a doctor as his peers, but he was not competent to save Cindy's life in this circumstance.

Consider another case identical to the previous one, except Jill, the senior doctor, is substituted for Jack. Jill is an experienced physician and promises the parents she will save Cindy's life, which is what she sincerely intends to do. She has an idea about the condition the child is suffering from, and whether it is treatable. In this case, Jill arrived at the justified belief she will save Cindy's life (B) by drawing inferences from her old belief. She acquired this belief from reading a medical journal (M) that reported a patient with Cindy's symptoms was treated in a specific way (x). Jill also believed M is very trustworthy in such matters, based on her experience. Jill's belief in curing specific diseases was stored in her memory and accessible to her. She made an inference from the belief retained in her memory and believed she would save Cindy's life.¹⁵

Following (R), Jill's belief in saving Cindy's life is justified because it is an output of a reliable process (inferential process) involving reliable inputs. Jill first used perceptual processes to form the belief that M reports the specific cure. The perceptual step is unconditionally reliable. According to (R), a belief is justified if it is produced by a belief-forming process that is unconditionally reliable. Jill then inferred from experience that M is trustworthy enough for her belief to be true regarding the specific disease cure. The inferential step is conditionally reliable. According to (R), the belief is produced by the inference process, which is a conditionally reliable process, and the input of this process, that is, her old beliefs, is justified. Next, the memory stage is a conditionally reliable belief-retaining process; its later outputs are usually true if the earlier inputs to it were true. Finally, she used the inferential step to infer that she would save Cindy's life. As I mentioned previously, the inferential step is conditionally reliable. Then, using principle (R), Jill's preserved belief in B is justified. Promising that she will save Cindy's life is a good promise as it meets the (internal) requirements of the competency to keep the promise. In contrast to Jack, Jill is competent to make the promise she will save Cindy's life.

In summary, A trusts B to ϕ because A believes B has a commitment to ϕ -ing. To be trusted when making a commitment, B must comply with the norm of promise-making. A good promise requires competence. The constitutional competence

¹⁵ One of Goldman's (2015b) examples inspired me to make such a case.

required for making a commitment is that B is *ex post* justified in believing he will successfully ϕ . Next, I specify the task (ϕ) that B is relied upon to perform as an assertion, and I analyse the norms associated with this specific task.

3.2.2.2. Assertion as Promising: The Norms of Being a Trustworthy Assertor

I have explained that, to promote trust, a person who is trusted to perform ϕ must conform to the norm of competence. Since the main case in this chapter is the one in which the task that the trustee is relied upon to perform is to assert $p \rightarrow \phi$ is specified with an assertion—this section explores the norms of being a trustworthy assertor.

What norms must a person meet to be a trusted assertor? Trusting other people's words involves relying upon them to fulfil a commitment, to satisfy both *promise-making* and *promise-keeping* norms. A commitment made by a speaker when making an assertion is that they speak justifiably. When *promise-making* norms are applied in the context of assertion, trustworthiness is required as competence in speaking justifiably. When the *promise-keeping* norm is applied to the assertion, the trustworthy assertor must in fact speak justifiably. This section clarifies the norms of being a trustworthy assertor (Hawley, 2019).

Asserting or telling¹⁶ involves a form of promise. One way to think of assertion as a special case of promising is to identify asserting that p to promising to p . In other words, since asserting involves a form of promise, it is a promise to p . Therefore, asserting that p is identical to promising to p . For example, when someone asserts there is snow outside, they promise there is snow outside. However, Hawley (2019) maintains it is unacceptable to identify asserting p to promising to p . When making an assertion, one need not be in a strong epistemic situation, such as when making a promise. By asserting that p , one does not become obliged to make it true that p . Thus, the account of assertion regarding promising does not entail identifying an assertion p with a promise to p .

Hawley (2019) proposes another way to assimilate assertion to promise by working out what a person is promising to do when making an assertion. She claims asserting whether p involves both

¹⁶ I continue to use the term 'assertion' rather than 'telling' because it is consistent with the terminology employed by Hawley (2019) and Brandom (1983), whose works serve as the foundation for this section.

- (a) promising to speak truthfully regarding whether p; and
- (b) speaking truthfully or untruthfully regarding whether p (i.e., keeping or breaking the promise).

Before proceeding, I modify Hawley’s account of assertion regarding the promise. The idea that identifies assertion to promise emphasises that assertion entails making a claim about something in fact in the world. This idea is rejected by Hawley (2019), who instead defends the idea that assertion involves a promise to speak in ways that match the world; a promise to speak truthfully requires promising there is a match between words and the world (Hawley, 2019, p. 52). Truth, in both propositions, that is, either there is something in the world or that words are matched to the world, is a purely metaphysical concept rather than an epistemological one. In both claims, what makes the proposition true or false is simply the state of the world. The claim’s truth value is not affected by the cognitive relations people have towards the relevant state of affairs. However, I state that assertion involves a promise to speak *justifiably*, which requires promising there is a cognitive relationship with the relevant state of affairs asserted. As Goldman highlights, ‘cognitive relations to a proposition are crucial for determining justification or warrant. A person’s justifiedness with respect to speaking as to whether p is never (or rarely) fixed by its actual truth value’ (Goldman, 2015a, p. 5). Given the difference between taking a claim to be justified and taking it to be true, I believe assertions are not faulty if the speaker lacks any evidence for its truth; rather, it is possible to have highly favourable evidence that justifies a proposition despite its falsity.

To address the concern related to the notion of truth in Hawley’s account, I propose the following requirements: asserting regarding whether p involves both

- (a) promising to speak justifiably regarding whether p; and
- (b) speaking justifiably or unjustifiably regarding whether p (i.e., keeping or breaking the promise).

There are three points to note about treating assertion as promising. First, Hawley’s (2019) view differs from a Brandom-style commitment to justify p. Second, as the condition of b) in Hawley’s account and the corresponding condition in my view illustrate, making a promise and keeping or breaking it happen simultaneously in the case of assertion. Third, the norms of promise, including competence is applied in the case of assertion. I now discuss each of these points in detail.

First, Hawley’s account of assertion in terms of promising to speak truthfully (or justifiably, in my view) differs from a Brandom-style commitment. According to Brandom (1983), asserting a sentence entails a commitment to present a justificatory

defence of it. Brandom suggests that ‘the commitment involved in asserting is to undertake the justificatory responsibility for what is claimed. In asserting a sentence, one commits oneself to justify the claim’ (Brandom, 1983, p. 641). Assertions are treated as warranted until challenged. One commits oneself to justify assertions once a specific question is raised regarding them. Although there is no end to the justification of the justification, and each justifying assertion may be questioned and need additional justifying assertions, the assertor must provide an appropriate set of justifying assertions if challenged (Brandom, 1983). For example, I assert there is snow outside to my neighbour. In responding to a challenge by my neighbour that the white stuff is not snow, but foam, I assert I saw no person, or film crew, put foam outside. Hence, I provide a set of justifying assertion(s) inferentially related to the original claim.

However, Hawley (2019) contends that assertion does not involve commitments that extend beyond the moment of making the assertion, either in terms of justification or retraction. In Hawley’s view, people who make a promise to do something become obliged to do it, but they do not become obliged to provide evidence of having done so if challenged. For example, if a son promises his mother he will finish his homework before dinner, he is obliged to do so. Nevertheless, he is not obliged to show his mother the completed homework. The son refuses to show his schoolwork because he wants his mother to trust him, to take him at his word. Otherwise, his mother’s inability to relax reveals a lack of trust. Trusting someone to keep their promises typically involves relying upon them to behave in the manner in which they committed to behaving and does not involve justificatory commitments. Similarly, a promise to speak truthfully (or justifiably, in my view) does not require an assertor to provide evidence they have spoken truthfully (or justifiably, in my view) even if challenged (Hawley, 2019). I agree with Hawley in that I think an account of assertion in terms of promise does not entail anything as extensive as Brandom’s commitment account of assertion.

Second, assertion involves a promise to speak justifiably¹⁷ and keeping or breaking that promise at the same time. The promise made in assertion is uncommon because it is made and kept at the same time, or else made and broken at the same time. For example, Clara asks Emma, ‘Do you promise to say your next word as loudly as you can?’ Emma shouts back, ‘YES!’ Emma promises to speak as loudly as she is able, and then simultaneously either keeps or breaks the promise. The promise to speak

¹⁷ Although the term ‘truthfully’ is used in Hawley’s argument, I use the term ‘justifiably’ in the remainder of the chapter for the reasons stated above.

justifiably is kept by speaking justifiably (Hawley, 2019). An assertor keeps the promise to speak justifiably once they are speaking.

Third, as I mentioned in Section 3.2.2.1, promise-making is governed by the norm of competence. When the norm is applied to the special case of promising to speak justifiably regarding whether p , the following result is obtained:

- (a) One must promise to speak justifiably regarding whether p :
 - only if one is competent to speak justifiably regarding whether p .

What elements are required for a competence norm in a promise to speak justifiably regarding whether p ? Remembering that proper promise-making requires competence, the response in this respect is to not promise to speak justifiably regarding whether p unless you are competent to speak justifiably regarding whether p . Competences, according to Sosa (2010), encompass three components: constitution, condition, and situation. I begin with the latter two components and return to the first afterward. In the case of assertion, conditional competence is achieved when the assertor is sober, awake, and alert. Situational competence is related to the circumstances in which an assertor must act or speak justifiably. Regarding the specific task of assertors to utter p , the external circumstances might be to ensure what audiences expect to hear from them, as indicated by Hawley (2019).

Constitutional competence, I argue, is close to the doxastic sense of justification for what is promised. More precisely, I claim that one is competent to keep a promise to ϕ when one is doxastically justified in believing ϕ .¹⁸ Similarly, one is competent to speak justifiably regarding whether p , only if one is *ex post* justified in believing whether p . Consequently, one has the appropriate competence to assert whether p only if one is *ex post* justified in believing whether p .

My view is that one is competent to assert whether p only if one justifiably believes whether p , which differs from the *reasonable to believe norms of assertion* proposed by Lackey (2008). Lackey highlights that one should assert whether p only if it is reasonable for one to believe whether p . According to Lackey, an assertor might fail to believe whether p ; nevertheless, they have substantial evidence indicating that such a proposition should be believed, rendering it reasonable for them to believe whether p (Lackey, 2008, p. 125). However, I claim, to be competent in asserting whether p , one

¹⁸ In this chapter, I am not discussing whether testimony transmits knowledge (or justification) or generates knowledge (or justification), nor am I discussing under what conditions hearers are justified in believing what a speaker testifies. I simply clarify the condition required to be met for an assertor to be competent in asserting that p .

must *in fact* believe whether p. A competent assertor can offer an assertion only if the assertion does in fact represent the beliefs of the assertor.

To clarify the differences between the strong requirement that one must in fact believe whether p, which is defended by myself, and the weaker requirement that it must be reasonable for one to believe whether p, which is defended by Lacky, consider the following modified version of the creationist teacher presented by Lacky:

‘Stella is a devoutly Christian fourth-grade teacher, and her religious beliefs are grounded in a deep faith that she has had since she was a very young child. Part of this faith includes a belief in the truth of creationism and, accordingly, a belief in the falsity of the evolutionary theory. Despite this, Stella fully recognises that there is an overwhelming amount of scientific evidence against both of these beliefs. Indeed, she readily admits that she is not basing her own commitment to creationism on evidence at all but, rather, on the personal faith that she has in an all-powerful Creator. Because of this, Stella does not think that religion is something that she should impose on those around her, and this is especially true with respect to her fourth-grade students. Instead, she regards her duty as a teacher to include presenting material that is best supported by the available evidence, which clearly includes the truth of the evolutionary theory. As a result, while presenting her biology lesson today, Stella asserts to her students, “Modern-day Homo sapiens evolved from Homo erectus”, though she herself does not believe this proposition’. (Lacky, 2008, p. 111)

Stella¹⁹ has strong evidence that Homo sapiens evolved from Homo erectus, and she asserts this proposition to her students despite not actually believing it herself. In this case, Stella does not possess a belief in the proposition; nevertheless, she has substantial evidence indicating that such a proposition should be believed, making it reasonable for her to believe the proposition. However, Stella must have believed that p to genuinely assert that p, because competence in the realm of assertion, I argue, requires that one offer an assertion in the presence of the corresponding belief. In this regard, a strong requirement for being competent for an assertion is required. Fulfilling the competence norms requires a stronger epistemic condition than it being reasonable for a person to believe a proposition; one must actually believe a given proposition, and that belief must be justified. To qualify as competent in asserting a proposition, one must have a doxastic rather than a propositional justification for the given proposition. Respectively, in my view, Stella does not qualify as a person who is

¹⁹ Those who argue for propositional justification need to demonstrate that Stella treats consideration, which plays the role of evidence, as evidence. If she did not see considerations as evidence, she would not have evidence. Accordingly, she would not be justified in the sense of proposition.

competent to assert whether *Homo sapiens* evolved from *Homo erectus* because she does not in fact believe the proposition. Hence, she violates a competence norm. Even if she had intended to speak justifiably, I think she would not have been competent to assert whether *Homo sapiens* evolved from *Homo erectus*.

The norm related to promise-*making* has been discussed: a competence norm. If an assertion is a matter of promising to speak justifiably regarding whether *p*, and simultaneously keeping or breaking that promise, we should expect it to be governed by the norm relevant to promise-making and by the norm relevant to promise-*keeping*. The norm related to promise-keeping is as follows:

- (b) asserting regarding whether *p* involves speaking justifiably or unjustifiably regarding whether *p* (i.e., keeping or breaking the promise).
 - One must assert regarding whether *p* only if one does in fact speak justifiably regarding whether *p*.

Trusting other people's words involves relying upon them to fulfil a commitment – to satisfy both promise-making and promise-keeping norms. A trustworthy assertor must conform to both promise-making and promise-keeping norms. A trustworthy assertor must:

- be competent to speak justifiably regarding whether *p*:
- only if one is *ex post* justified in believing whether *p* (constitutional competence);
- only if one is awake and alert (conditional competence);
- only if one can ensure what audiences expect to hear from them (situational competence).
- in fact speak justifiably regarding whether *p*.

3.3. Artificial Intelligence and Trust

I have addressed the question, 'How do A and B cultivate or maintain the relationship of trust?', and discussed the norms of being trustworthy regarding a general task of ϕ and a specific task of assertion, emphasising the role of B as a trusted person in maintaining or cultivating a trust relationship with A. I now take the final step towards answering the main question: 'How does an AI system affect privacy?' This step requires exploring how an AI system impacts trust relationships to answer the following question: 'How does an AI system affect trust relationships between A and B?' Answering this question is essential to accomplishing the main goal of this

research: understanding how an AI system impacts privacy, which depends upon trust.

How does an AI system affect trust relationships between A and B? To answer this, I examine how AI impacts B's competence. The main case study of this chapter is the one in which the assertor (B) employs an AI system to decide whether A has breast cancer (p). One norm an assertor must fulfil to be trustworthy is being competent to speak justifiably regarding whether p. In doing so, the assertor must be *ex post* justified in believing whether p. Therefore, the question that may arise is whether and how the assertor is justified, in a doxastic sense, in declaring whether p in cases in which p is a proposition delivered by an AI system. Part of the answer to this question emphasises the role of AI in justifying B's belief that p, and thus its contribution to B's competence.

When an AI system, as the diagnostic instrument, informs B that the scan or biopsy of the patient (A) indicates the presence of cancerous cells, B uses the instrument as an 'epistemic instrument' (Sosa, 2006, p. 118) and forms beliefs based on what the instrument delivers, and then acts accordingly. Grindrod (2019) refers to beliefs formed based on deliverance from an AI system in general, or an ML model in particular, as computational beliefs. 'Whether' and 'how' B are justified through believing the proposition delivered by the instrument. In other words, *how* is B's computational belief justified?²⁰

The question of how to justify B's computational belief hinges on whether such a belief can be regarded as a distinctive form of belief or as an epistemic source that can be reduced to other epistemic sources. According to Goldman (2015b), a distinctive source provides justification on its own, without depending on other sources for its justificatory power, whereas reductionism-based justification is derived from other, more basic sources. In addition to memory and perception, I consider testimony as a distinct epistemic source. In line with Grindrod (2019), I endorse the reductionism approach to computational belief, even though these types of belief cannot be reduced to memory, perception, and testimony. Rather, computational belief can be viewed as a form of inferential belief that acquires justificatory power from reliable inductive inference.

Computational beliefs cannot be reduced to memory, perception, or testimony. Memory can be dismissed because the process of obtaining a computational belief is

²⁰ In what follows, I do not consider the 'whether' question. Readers wanting to know whether beliefs formed via the results of ML models can be justified at all should see Grindrod's (2019) paper, which provides reasons to think that computational beliefs are justified.

not equivalent to remembering a certain proposition. Computational beliefs do not resemble perceptual beliefs either; perceptual experiences with an instrument justify B in believing merely that there is an instrument, rather than believing in that deliverance. As a result, computational beliefs are not completely captured as a form of perceptual belief. Computational beliefs cannot be described as the result of a testimonial exchange. An AI system is not an epistemic agent; it does not possess beliefs in the common sense. Therefore, we cannot rely upon an AI system via testimony (Grindrod, 2019).

I agree with Grindrod (2019) that beliefs formed based on the deliverance of an AI system can be reduced to a form of inferential beliefs. B might infer computational belief that p from premises that take the form of inductive generalisation reasoning or, alternatively, premises that describe what other people testify to (Grindrod, 2019; Sosa, 2006). Accordingly, B might apply at least two distinct arguments to explain how he reaches the conclusion that p. However, B is not obliged to offer A a justification for what is said, nor does B need to undertake justificatory responsibility for what he says (see Section 3.2.2.2). Since B's doxastic attitude towards the proposition that p is justified only if arriving at the belief that p is the output of a reliable process (see Section 3.2.2.1), the justification of B's belief in p that can be offered for each distinct argument is presented as follows.

First, B might reach his computational belief that p by appealing to premises that describe a merely observed correlation, which offers him inductive support for the target proposition p:

P1: The deliverance of the instrument is proposition p.

P2: B learns from experience and test data samples that the given instrument in this specific field usually delivers the correct proposition.

P3: The deliverance of propositions p by the given instrument in this specific field is correct.

Therefore,

C: p.

Suppose B uses the system with no particular view regarding its reliability. He uses the personal data of those whose diseases have not been diagnosed by the system as test data to assess the accuracy and model performance. He finds the system produces correct answers for the test data (P2) and eventually infers the deliverance of the system in this context (or specific field) is epistemically reliable (P3). Recall principle (R). This inductive inferential cognitive step involved a conditional reliable process. That is, the step's later output is usually true if the earlier input to it is true. Given that

B's experiences with tested data are the input of the process, the output that is P3 is reliable. His belief in p was then formed using another inferential step, which is a conditionally reliable belief-forming process. B's belief in p is the output of the inferential process with the input of inductive generalisation. Since both reasoning processes are reliable, then, according to principle (R), B's belief in p is justified.

Second, B might infer some computational belief that p is based on premises upon which he relies regarding the testimony of another person:

P1: The deliverance of the instrument is proposition p.

P2: Other person said that the given instrument in this specific field usually delivers the correct proposition.

P3: The deliverance of proposition p by the given instrument in this specific field is correct.

Therefore,

C: p.

Again, according to principle (R), B is justified in believing p because p is the output of the inferential reasoning process, which is a conditionally reliable process. The input of this process is a testimonial belief (P2), which can itself be considered a conditionally reliable process or unconditionally reliable process. In the debate about testimonial knowledge, there has been a great deal of discussion about whether testimony as an epistemic source can be reduced to basic epistemic sources (Hardin, 2002), or whether it constitutes a separate and distinct epistemic source (Coady, 1973). Regarding the former, testimonial belief can be formed based on the process that is conditionally reliable with the input of memory, perceptual, or other inferential belief. Regarding the latter, testimonial belief is formed based on the process that is unconditionally reliable. Either way, the input of the process is reliable. Thus, B's belief in p is justified.²¹

I have explained that B is justified in believing that p because of the existence of a valid inferential process that forms this belief. In the first case, B relies on the inductive generalisation that proceeds from the limited sample of B's case to infer his belief in p. In the second case, B relies on another's testimony to infer his belief that p. Therefore, justification of the computational belief involves, *first*, B's or, *second*, the other's cognitive accomplishments. Furthermore, either B himself tests and gains inductive support for the accuracy of the instrument, or the developer of the ML

²¹ In both arguments, B's belief in p is justified not by evidence—*beliefs* from which p can be inferred, or *perceptual* and *memory* experiences—but by non-evidentiary reasons concerning the reliability of the processes involved in forming p.

model testifies to some level of accuracy for the model; therefore, *third*, B's computational beliefs partly rely on the accuracy and the operation of the instrument. Hence, in addition to B's or the other's cognitive accomplishments, this feature of the ML model's performance contributes to the justification of the computational belief.

First, concerning B's cognitive accomplishment, does it require that B be aware of how the instrument operates to be justified in believing that p ? Does the accomplishment require that B understands how the instrument he relies on performs to form a justified belief based on what the instrument delivers? The answer, in my view, is negative. According to the above discussion, being justified in believing that p is independent of being aware of how the instrument operates; rather, it requires that the belief is the output of a valid reasoning process in which the input beliefs are reliable. Although it is often not possible to understand properly how the algorithm processes the data and reaches the outcome it does, such an opacity does not impact the reasoning process that justifies a computational belief.

However, such an opacity leads to a significant issue, which is 'epistemic responsibility gaps' (Grindrod, 2019, p. 3). According to Grindrod (2019), there is an important sense in which B relies on his epistemic community while employing instruments he does not understand. The epistemic community consists of individuals who comprehend how the instrument performs, and B can appeal to that community if they find that the instrumental inferences are incorrect. However, computational beliefs depend upon autonomous learning algorithms, which are opaque in nature, making it challenging for any member or group of members to understand the exact workings of these algorithms. Therefore, B cannot properly rely on his epistemic community to compensate for his not understanding how the instrument performs when he forms his computational belief (Grindrod, 2019).

Second, concerning the role of the other's cognitive accomplishment in justifying B's computational belief, does it require that the epistemic community be aware of how the instrument operates to testify to the accuracy of the instrument? Do the responsibility gaps impact B's justification for believing what the instrument delivers? Again, in my view, the answer is negative. It is not necessary for the person who developed an instrument or model to understand how it operates to testify to its accuracy. Without necessarily understanding how the instrument operates, the model developer can appropriately declare that the instrument performs accurately as they have credence in the instrument's performance, which is supported by testing sample datasets. Therefore, a lack of epistemic responsibility by the epistemic community has no effect on the justification of the computational belief. The lack of impact does not

imply the discussion of epistemic responsibility does not merit investigation. On the contrary, computational belief leads to a distinct structure of epistemic responsibility, which deserves detailed research, but not in the realm of appropriate assertion and trust.

Third, it is argued that B's computational beliefs partly rely on the accuracy and the operation of the instrument. Although a lack of understanding of how an AI system (or an ML model) performs does not affect the justification of the computational belief, its accuracy does. Since being justified in believing what the instrument delivers is required for B to be competent in what he asserts, the accuracy of an AI model affects B's competence. Given that trustworthiness requires competence, an AI system impacts trust relationships between A and B since B's competence requires the AI system to perform accurately.

3.4. Conclusion: Trust, Privacy, and Artificial Intelligence

How does an AI system affect privacy? This section summarises the previous discussions and answers this question. A person (B) who employs an AI system to respond to another person's (A) question (p) relies epistemically upon the system and asserts p based on what the system delivers. One norm that B must fulfil to be trustworthy is the competence to speak justifiably regarding whether p. Justification of B's belief that p partly relies on the accuracy of the AI system. Thus, accuracy is a feature of an AI system's performance that contributes to the justification of B's belief in p. Accordingly, B's competence relies on the accuracy of the operation of the system. Since trustworthiness requires B's competence while asserting p, the AI system affects trustworthiness and, consequently, the trust relationship between A and B.

Privacy is a social value constituted by trust-based relationships. Privacy, in a disclosure context, is constituted by interactions between different individuals based on trust. Since AI affects trust, AI impacts privacy. To achieve privacy as a social value, an AI system must perform accurately. Hence, the main RQ concerning how an AI system affects privacy, is explained by how an accurate AI system contributes to building trust relationships between A and B, which constitutes privacy. As a result, both B, as the trustee, and the AI system that makes B competent in his assertion contribute to the constituting of privacy.

To conclude, I believe, in contexts in which the relationships among individuals engaged in the practice of information-sharing are grounded in trust, that sharing

information, analysing, and inferring from the shared information, as well as preserving privacy, are not mutually exclusive.

3.4.1. Implication: Extending the Scope of Privacy

Does taking trust as an *ex post* approach impact the scope of privacy? To answer this question, it is crucial to study the type of information within the scope of privacy. According to Inness (1992), privacy might not protect *all* information about a person, but might involve only *intimate* information. The intimacy of information stems from the act of sharing information that is itself intimate. An act or activity is intimate iff its meaning and values draw from the person's intimate motivations, such as love, liking, or care. The act of sharing information is intimate iff it is understood to take its meaning and value from our love, liking, or care, not merely iff it conveys a desire on our part to inform another person. For example, we value showing our love letters to others as an intimate act iff it conveys the meaning that we care for them, not to extort money from them. Protecting privacy entails protecting actions (such as the dissemination of information about oneself) that are understood as expressions of love, liking, or care; privacy claims are claims to exercise control over intimate decisions and actions.

Inness's idea has two interrelated parts: the realm of privacy and privacy claims. Therefore, it is important to discuss how taking trust as an *ex post* approach to privacy affects these parts. I begin with the privacy realm part. By taking trust as an *ex post* approach to privacy, the scope of privacy is expanded to include information exchanged in a trust-based context. Unlike intimate relationships formed between friends, partners, and lovers, trust relationships are not always confined to those people who know them and are close to them. Although trust does not require a person to be in a close relationship, it subsumes cases in which the person is in an intimate relationship. In this regard, the scope of privacy is expanded to include information shared or revealed in a trust-based context.

Determining the scope of privacy does not require a perspectival assessment, because assessing trust does not demand a perspectival assessment. Unlike intimacy, which requires a personal viewpoint to characterise underlying motivations—a person can confirm whether their own actions embody love, liking, or care—interpersonal trust is independent of one's motivation. A person motivated to act is not trustworthy; rather, trustworthiness requires avoiding unfulfilled commitments or broken promises (see Section 3.2.2.1).

Regarding privacy claims, unlike Inness's taking control account of privacy, which merely emphasises a person who shares data with others, the trust-based approach emphasises the role of others and relations between them in constituting privacy as well. Privacy claims are claims that the information exchanged in the trust-based context is to be cared for. Such a claim can take the form of cultivating trust between those involved in a disclosure context by conforming to trust norms. Accordingly, protecting privacy entails promoting or maintaining trust. Therefore, regulations need to be established that focus on building, maintaining, and fostering trust in a disclosure context.

Chapter 4:

Limiting Access to Certain Anonymous Information:

From the Group Right to Privacy to the Principle of Protecting the Vulnerable

Abstract

An issue about the privacy of the clustered groups designed by algorithms arises when attempts are made to access certain pieces of information about those groups that would likely be used to harm them. Therefore, limitations must be imposed regarding accessing certain information about clustered groups. In the discourse on group privacy, it is argued that the right to privacy of such groups should be recognised to respect group privacy, protecting clustered groups against discrimination. According to this viewpoint, this right places a duty on others, for example, private companies, institutions, and governments, to refrain from accessing such information. To defend the idea that the right to privacy should be recognised for clustered groups, at least two requirements must be satisfied. First, clustered group privacy must be conceived of as either a collective good or a participatory good. Since these forms of good are of the type from which no member of a group can be excluded from benefiting, the right to them is defined as a group right. Second, there must be group interests on which to base a group right. Group interests can be either the interests of those members that are a result of their being in the group or the interests of the group as a whole that transcend the interests of its members. However, this chapter argues that clustered group privacy cannot be conceived of as either a collective or a participatory good because it is possible for some individuals to be excluded from benefiting from it. Furthermore, due to the lack of awareness among individuals that they are members of a clustered group and the nature of a clustered group itself, such groups cannot have the group interests necessary to establish a group right. Hence, the group right to privacy cannot be recognised for these groups, implying that the group right cannot be considered a means to protect clustered groups against discrimination. Instead, this chapter suggests that moral principles need to be articulated within an ethics of vulnerability to identify the moral obligations of protecting vulnerable clustered

groups. The duty owed to the vulnerable should involve refraining from accessing certain information about (or related to) clustered groups in specific contexts. This duty is not engendered by the right to privacy of such groups; it is the duty owed to the vulnerable. The findings highlight the need to articulate moral principles regarding privacy and data protection to protect clustered groups in contexts in which accessing information about them could constitute a reason for discriminatory targeting.

Keywords: clustered groups; collective right to privacy; corporate right to privacy; group privacy; principle of protecting the vulnerable

This chapter is a modified version of the following publication:

Asgarinia, H. (2024). Limiting Access to Certain Anonymous Information: From the Group Right to Privacy to the Principle of Protecting the Vulnerable. *The Journal of Value Inquiry*. <https://doi.org/10.1007/s10790-024-09980-x>

4.1. Introduction

By harnessing the potential of big data analytics and data-driven technologies, specifically machine-learning (ML) algorithms, these technologies stand at the forefront of computing advancements (Mühlhoff, 2021). Such technologies process and analyse large quantities of data²² based on patterns and group profiles to uncover new patterns or structures and/or to confirm suspected correlations within datasets (Taylor et al., 2017). Automated forms of data analytics, such as ML, have the potential to impact how groups are identified and perceived, enabling the design of new groups without predefined parameters or attributes (Kammourieh et al., 2017). Through data analytics, individuals are grouped based on the similarity of attributes, such as age, gender, and purchasing behaviour, and possible correlations can be explored. However, individuals grouped together based on certain similar attributes may not be aware they are being bound by these similarities.

The results of data analysis are often used to inform policies, target specific groups, and make decisions that may pose risks to a group. The types of actions and interventions analyses facilitate are not aimed at individuals. Instead, these actions and interventions focus on groups with some interesting property or ‘type’ (of customers, dog lovers, skiers, ...) to which the individual (a ‘token’, e.g., Alice, you, ...) now belongs. Therefore, data analytical technologies are designed to operate on the broadest possible scale, in which the individual is often incidental to the analysis (Taylor et al., 2017).

Big data analytics technologies—ML algorithms in particular—are directed at the group level and are used to formulate types, not tokens, challenging the foundations of most current ethical theories, particularly concerning privacy. Privacy has traditionally been regarded on an individual level; however, the increasing use of these technologies forces us to ask questions about privacy on a group level (Floridi, 2017). According to this point of view, the concept of privacy needs to be reshaped to help us think about the privacy of groups.

For a clearer understanding of the issue concerning group privacy, consider the following case:

A research proposal aims to uncover correlations between purchasing behaviour and highly sensitive attributes, particularly sexual orientation,²³ using cluster techniques as part of ML tasks. To achieve this, a dataset

²² It is important to note that ‘data’ and ‘information’ are used interchangeably in this research.

²³ This research was conducted by Kosinski et al. (2013) on social networking sites.

containing over 60,000 records that pertain to information about individuals, who are henceforth referred to as ‘data sources’, as recommended by Henschke (2017), who provide information on individuals’ purchasing behaviour, detailed demographic profiles, and self-reported sexual orientation. By applying clustering techniques to the dataset, distinct groups of individuals based on similarities in purchasing behaviour are identified, henceforth referred to as ‘clustered groups’.²⁴ Once these clustered groups are identified, the research explores any correlations or associations between purchasing behaviour and sexual orientation in the identified groups. This analysis aims to uncover patterns or trends in purchasing behaviour that may be indicative of a higher likelihood of having a particular sexual orientation. Hence, the information obtained from the analysis indicates a correlation between purchasing specific items and having a particular sexual orientation. For example, the analysis reveals that 88% of the members in a clustered group exhibiting specific purchase behaviour are thought to be homosexual and/or engage in some same-sex activity. By employing reliable generalisation techniques, the information obtained from the few data sources is generalised to a broader population; the derived result indicates that 88% of the population exhibiting specific purchase behaviour are homosexual.²⁵

Now, consider a totalitarian government characterised by intolerance towards homosexuals. This government formulates a policy that explicitly targets this particular group based on their purchasing behaviour. The policy aims to impose disadvantages or deprivations on the group. These disadvantages manifest through the denial of education opportunities or employment that others enjoy in society. Importantly, these unfair disadvantages are directed

²⁴ Data sources are members of a clustered group.

²⁵ I am not discussing a case in which data handlers use the information obtained at the group level to infer information about an identified individual (e.g., Anna is a homosexual because of the similarity of her purchase habits with the clustered group). For more information about whether and how putting a person in a group because their information provided is similar to others, and accordingly making inferences about an individual whose data were not used in the training dataset, violates the privacy of that individual, see Munch (2021) and Mühlhoff (2021).

Additionally, I am not discussing cases in which undisclosed attributes of a user are inferred based on the disclosed attributes of the user’s friends on social networking sites. In this regard, I do not discuss ‘Networked Privacy’ (boyd, 2011), which implies that the privacy of a user on a social network site is connected to others. However, I focus on cases in which inferences are not drawn based on confirmed relationships existing among members of a group but, rather, on the absence of ties among members of a particular group.

towards groups of people with (an assumed) specific sexual orientation. Thus, discrimination occurs against the group rather than targeting a specific individual within the group. Accessing information obtained at the group level enables the government to formulate a discriminatory policy.

Certain pieces of anonymous information not linked to an identifiable person but about a clustered group enable that group to be easily identified and targeted. Such information can be used by a corporation, private company, government, or institution to harm the group in morally objectionable ways. Accessing such information harms the privacy of a clustered group (see Section 4.2).²⁶ Therefore, limitations must be imposed regarding accessing certain pieces of information about clustered groups, as this could help protect the group against discrimination.

Almost all privacy scholars who address the concern of clustered group privacy (e.g., Floridi, 2014, 2017; Mantelero, 2017; and van der Sloot, 2017) claim that the right to privacy for such groups should be recognised. This right holds others (e.g., researchers, data handlers, or governments) under a duty to refrain from accessing specific pieces of information about a clustered group. Floridi (2014), who first raised the concept of group privacy in relation to big data analytics technologies, argues the (clustered) group right to privacy is irreducible to the right to privacy of the individuals who comprise that group. This idea is grounded in the practical reality that, even if the privacy of each individual in a group can be protected, the privacy of the group may be violated. For example, the privacy of individuals who comprise a clustered group might be protected using anonymisation techniques but access to certain pieces

²⁶ My claim is that accessing certain information about a clustered group poses a threat to the privacy of the group. However, it is worth noting that there are other potential harms that extend beyond the scope of this chapter. These harms may arise from the design of clustered groups through analysing aggregated anonymous information or the generalisation of information from a few individuals to an entire population. If data handlers or researchers have a duty to refrain from designing or generalising information, it must be considered part of their epistemic duties, including duties against stereotyping (Fricker, 2007).

Another concern might arise from publicising the information conveyed by ML models. Making information obtained at the group level public facilitates access to that information, which could constitute a reason for an agent to act in morally objectionable ways to harm the group. However, it is not always necessary for information about a group to be public to raise concerns about privacy. There are cases in which private companies develop models and internally use the generated information to discriminate against specific groups, as exemplified by Lippert-Rasmussen and Aastrup Munch (2021). Although their paper primarily argues for the individual right to privacy, I believe it is also important to discuss this issue from the perspective of group privacy. Therefore, I argue that the concern for group privacy arises when corporations, institutions, or private companies access information at the group level without necessarily making it public.

of anonymous information at the group level is still possible, which will be likely to be used to harm that group, raising concerns about the privacy of the group.

However, in response to the need for limiting access to certain anonymous information about clustered groups, I argue against the predominant approach in group privacy discourse. This approach emphasises recognising the right to privacy for these groups to achieve that end. Instead, I contend a clustered group cannot primarily have a right to privacy. I suggest that the moral principle of protecting the vulnerable imposes restrictions on accessing certain information about clustered groups, thereby protecting these groups against discrimination. Thus, although new and advanced technologies raise unprecedented concerns about group privacy, this does not necessarily imply that the right to privacy of such groups is at risk, since such a right is not primarily defensible for these groups. As a result, the privacy of a group and the group right to privacy should be regarded as distinct concepts and not be conflated.

4.1.1. Overview of the Chapter's Argument and the Proposed Approach

Although there is no consensus among proponents of the group right to privacy regarding defining such a right, Floridi (2017) and Mantelero (2017), for example, argue it is necessary to respect the right to privacy of a clustered group to prevent discrimination against that group. It follows that it is the violation of the clustered group right to privacy that provides a government, in the particular case explained above, or data handlers (in general) with certain pieces of anonymous information about that group. Such information would then likely be used in decision-making processes that target groups in discriminatory or harmful manners. Accordingly, I assume that proponents of the group right to privacy describe this right as a right to limit access to anonymous information acquired about a clustered group that could provide others with reasons to harm the group. If there were a right to privacy in terms of a right to limit access to certain information about a clustered group, it would place a duty on a government, data handlers, or any other agent to refrain from accessing certain information. The connection between a right and duty is further examined in Section 4.3.

Nonetheless, the major challenge that proponents of the group right to privacy face is how to conceive of the holder of a group right to privacy. According to Floridi (2017), the group right to privacy is a right held by a group as a whole rather than by

its members collectively. However, an alternative perspective conceives of the group right to privacy as a right held jointly by members of a group (Puri, 2022).²⁷

The lack of consensus regarding the holder of a group right in general underpins the disagreement specifically regarding the holder of the group right to privacy in particular. The two most common accounts of who or what is the bearer of a group right in general are the collective and the corporate accounts (Jones, 2013). The collective account identifies the bearer of rights as individual members of a group (e.g., Raz, 1988). The corporate account locates this bearer at the group level (e.g., French, 1984). Hence, the challenge that group privacy advocates face is determining whether to apply a collective or corporate approach to the right to privacy. Taking a collective approach to group rights is compatible with human rights because individuals jointly hold such rights (Raz, 1988), whereas taking a corporate approach considers the rights to be non-human and held by a corporation (Jones, 2013). Regarding the group right to privacy, the concern is determining whether it should be conceived as a human right or a non-human right.

One way to address the aforementioned issue is to focus on the nature of the good for which a group right is claimed. The nature of the good, which is the object of a right (that is, what we take a right to be a right to), determines who or which entity is a holder of the right. According to Raz (1988), who is one of the most prominent supporters of the collective approach, if there is a right to a ‘collective good’, then that right must be a collective right. Collective goods are those goods from which it is logically impossible to exclude any member of a society or a group from benefiting from them (Raz, 1988). On the other hand, Réaume (1988) aligns with the corporate approach and states that, if there is a right to a ‘participatory good’, then the right to that good must be a corporate right. Participatory goods are produced through the involvement of many; one cannot individually enjoy the benefits of them unless others with similar interests do. To conclude, if there is a right to a good, and that good is a collective good, then the members of a group collectively hold a right to that good. However, if there is a right to a good and that good is a participatory good, then the group as a whole holds the right to that good.

²⁷ Floridi (2017) and Mantelero (2017) defend the corporate approach to the group right to privacy, although Floridi does not explicitly refer to this term; instead, he uses the term ‘strong’ to describe his approach. On the other hand, Puri (2022) draws attention to the collective approach to the group right to privacy. I demonstrate that neither approach can be applied to describe a clustered group right to privacy (see Section 4.6).

Moreover, an interest that grounds a right determines who or which entity can claim that interest to be respected. According to Raz (1988), the aggregation of interests of members of a group grounds a collective right. However, according to Newman (2004),²⁸ who is a proponent of the corporate approach, the interests of a group as a whole, which are not reduced to the aggregation of the interests of its members, ground a corporate right. The non-aggregative interests of a group, independent of the interests of the members, ground a corporate right. According to the collective approach to rights, members of a group collectively claim for a right to protect interests aggregated among them. However, according to the corporate approach to rights, a group as a whole can claim a right to protect its non-aggregative interests.

Concerning the nature of goods as objects of rights and the interests that ground rights, my argument against recognising the group right to privacy for clustered groups is as follows:

First, I argue the privacy of a clustered group cannot be considered either a collective or a participatory good (see Section 4.6.1). Clustered group privacy is not a collective good because it is logically possible to exclude any member of a clustered group from benefiting from it. Moreover, the privacy of a clustered group is not a participatory good because of the lack of interaction between the members of clustered groups required to produce it.

Second, I argue that neither members of a clustered group nor the clustered group itself can have an interest that can ground a group right (see Section 4.6.2). Regarding a defining feature of a clustered group, members of a clustered group are unaware of being members of that group. This lack of awareness implies members cannot have an interest in virtue of being members of a clustered group that can ground a collective right. Moreover, due to the nature of a clustered group, it is impossible to assign an interest to such a group that transcends the interests of its members.

Finally, I conclude the group right to privacy cannot be a collective or a corporate right. Regarding the collective and corporate approaches to group rights, if a clustered group can have a right to privacy, then the right must be a collective or corporate right. This point entails that a clustered group cannot have a right to privacy.

²⁸ It should be noted that Newman advocates for the corporate approach to group rights, yet in Newman (2004), he uses the term ‘collective’ to describe group rights. To maintain consistency with the text, I use the term ‘corporate’ to describe Newman’s theory.

Instead of recognising the right to privacy for clustered groups, I suggest taking a moral principle for the moral obligation of protecting vulnerable clustered groups within an ethics of vulnerability. Accessing certain information about clustered groups that causes, threatens to cause, or is likely to cause harm to those groups makes them vulnerable. The duty owed to vulnerable clustered groups is to impose a limitation on accessing certain pieces of information about a clustered group that are likely to be used to formulate policies, make decisions, and act in morally objectionable ways that harm the group. Thus, the duty to respect clustered group privacy is engendered by the principle of protecting the vulnerable, not the group right.

The research findings have both theoretical and practical implications. From a theoretical perspective, a comprehensive argument is lacking in the literature on group privacy to demonstrate that the recognition of the right to privacy for clustered groups is implausible. From a practical perspective, the findings highlight the need to consider clustered group privacy and moral principles to protect the vulnerable in the context of privacy and data protection. Moreover, the findings are of practical interest due to the suggestion of the need to develop techniques to protect group privacy, such as encrypting the general patterns uncovered by ML algorithms, thereby restricting access to information about clustered groups.

In the following sections, I explore the argument demonstrating that a clustered group cannot have a right to privacy in greater depth. In Section 4.2, I provide an outline of what I consider group privacy, its realm, and the approach to it. In Sections 4.3, I explain both the corporate and collective approaches to group rights. In Sections 4.4 and 4.5, I outline the assumptions and requirements for qualifying as a right-holder under each approach concerning the nature of the good as the object of a right and the interest that grounds a right. In Section 4.6, I critically evaluate whether a clustered group satisfies the identified requirements for holding the right to privacy. In Section 4.7, I present my suggestion for respecting clustered group privacy. Finally, in Section 4.8, I address a potential objection that might argue that the invasion of group privacy is justified due to the beneficial results achieved in promoting public health.

4.2. Group Privacy

Floridi (2017) discusses the recognition of the right to privacy for groups designed by algorithms—a right ascribed to groups as a whole. As Floridi (2017) notes, the group right to privacy differs from the existing rights in the fields of privacy and data

protection in that it is not reducible to the privacy of the individuals who form such groups. As Floridi (2017) argues, opening anonymised data to public use in cases in which groups of people may still be easily identified and (discriminatorily) targeted increases the risk of violating the right to privacy of groups as a whole. Thus, accessing certain anonymous information related to a group that might be used in discriminatory ways to harm it raises concerns about the right to privacy of a clustered group as a whole, although there may not be a concern for the individual right to privacy of data sources.

Floridi's (2017) theory has two parts—which have received attention from privacy scholars who defend group privacy, such as Taylor (2017), Mühlhoff (2021), and Mantelero (2017)—the realm of group privacy and the approach to it. Regarding the former, Taylor (2017), like Floridi (2017), argues that the realm of privacy must be expanded to include anonymous information pertaining to a group with which a group can be identified and targeted. Accordingly, a new concept of privacy (i.e., group privacy) must be developed to protect this kind of information. Anonymising personal information involves the process of removing personal identifiers (Solove, 2008) or eliminating the link between data and a specific person (Barocas & Nissenbaum, 2014), thus turning personal into anonymised information.²⁹ Concurrently, information pertaining to a group can be translated as information that is not necessarily related to each individual member of the group but to the group as a whole. For example, a pile of books can have the property of 'being too heavy to be moved by a single person, despite the fact that each book in it is reasonably small and light' (Floridi, 2017, p. 89). Thus, the group, a pile of books, has a property (being heavy) that is not reduced to the properties of its members.³⁰ Therefore, an

²⁹ Since my focus is on whether the privacy of a clustered group is violated when anonymised information is accessed by an agent, I do not discuss the issues of re-identification or de-anonymisation, which involve linking an anonymised dataset with a separate dataset containing identifying information (for more information on this matter, see Barocas and Nissenbaum (2014); and Ohm (2009)). I argue that if the information uniquely identifies a person, then individual privacy might be at risk due to the personal nature of the information. In other words, if no effort is made to identify an individual, then accessing anonymised information would not violate the individual right to privacy. Therefore, my concern lies in accessing anonymised information by an agent, without exploring the discussion of re-identification, which can be addressed within the scope of the individual right to privacy.

³⁰ Although the relationship between group privacy and individual privacy is not explicitly clarified in the theories of those who defend group privacy, I assert that, based on their approach to the concept, such a relationship is one form of dependence. In this sense, group privacy is an emergent property

accumulation of individual rights to privacy would not protect the information pertaining to a group, as there are no concerns about individual privacy. In this sense, group privacy must transcend the collection of the right to privacy of the members who form that group (Taylor, Floridi, et al., 2017).

Recall the case in Section 4.1. Consider that, after collecting information, the researcher removes the identities associated with the information in a dataset for the analysis to proceed. The information discovered from the processing of anonymous information relates to the purchase behaviour of the clustered group as a whole; that behaviour is not necessarily related to each individual member of the group. Accessing this information, which might be used in discriminatory ways to harm the group, raises concerns about the privacy of the group. In this case, the nature of the information does not fall within the scope of the individual right to privacy; the scope of the individual right to privacy is, by nature, limited to personal information (Vedder, 1999). Rather, the information falls within the realm of group privacy.

Hence, establishing an account of group privacy is at least negatively useful in that it is not useful to reduce group privacy to individual privacy, just as it is not feasible to reduce special sciences to physics; reducibility to physics is regarded as a constraint on the acceptability of theories in special sciences (Fodor, 1974, p. 97). Analogously, reducing group privacy to individual privacy prevents acknowledging the significance of protecting certain kinds of information, namely anonymous information that pertains to groups.

Regarding the approach to group privacy, proponents of group privacy take the consequentialist approach, as echoed in Floridi's (2017), Mühlhoff's (2021), and Mantelero's (2017) works. Mühlhoff (2021) argues that the unfair and harmful use of information discovered using modern analytics, which leads to adverse decisions that affect the social situation, well-being, or welfare of groups, raises concerns about the privacy of groups. According to Mantelero (2017), group privacy is the right to limit the potential harm to the group itself that can derive from invasive and discriminatory data-processing. Since a group as a whole is targeted by discriminatory practices or policies (Taylor, van der Sloot, et al., 2017), its privacy is at risk. Thus, a right to privacy of a group should be respected to protect it against discrimination.

Drawing on the insights of Munch (2021), who develops the consequentialist approach to the right to privacy, accessing specific pieces of information about an entity (A) that could provide B with information that enables B to act *subsequently* in

without being reduced to its base property (i.e., individual privacy). Although a discussion on different forms of dependence relationships merits consideration, it goes beyond the scope of this chapter.

ways that would likely harm A in morally relevant ways violates A's privacy. As Munch emphasises (Munch, 2021), 'violating A's privacy could provide B (or others) with information that enables him (or them) to subsequently act in ways likely to render A worse off in morally relevant dimensions' (Munch, 2021, p. 3786). Accordingly, the right to privacy, as a right to limit access to certain pieces of information, functions as a means to protect A against discrimination.

Correspondingly, A's (instrumental) interest in privacy justifies holding others under the duty to protect that interest—the duty to refrain from accessing specific pieces of information that could be used to harm them (more information on the relationship between right, duty, and interests is provided in Section 4.3). As a right-holder, A might make a justifiable claim against a duty-bearer, such as a data-handler B, not to gain access to those particular pieces of A's information that are more likely to contribute to B forming a belief about A, which could motivate B to perform certain harmful acts against A.

I agree with Floridi (2017) that attempts to acquire certain pieces of information about a clustered group harm the group's privacy, as these pieces of information are likely to be used in certain objectionable ways to harm the group. Like defenders of group privacy, whose theories are explained in this section, I take the consequentialist approach to privacy.³¹ Therefore, I omit the more theoretical and abstract question of whether simply accessing information without making adverse decisions would also violate the privacy of the group. However, I disagree with Floridi (2017) and other proponents of the group right to privacy that the right to privacy of a clustered group must be recognised to protect the group's (instrumental) interest in having privacy respected by others to prevent others subsequently acting to harm them. I argue instead that a group right is not primarily defensible. An issue in defending the group right to privacy arises from the interpretation of the group interest, that is, whether it is the interest of those members that is a result of their being in the group (i.e., aggregative interest) or the interest of the group as a whole that transcends the interests

³¹ In addition to the consequentialist approach this chapter takes, I acknowledge other approaches to privacy, such as deontological and political ones. The deontological approach is concerned with autonomy (Munch, 2021), and the political approach addresses institutional or governmental power (Henschke, 2020; van der Sloot, 2017; Véliz, 2020). The exploration and discussion of whether deontological and political approaches defend the group right to privacy or view the group right to privacy as a collection of individual right to privacy, which might be violated due to the *collection* of information, are important but beyond the scope of this chapter. The main concern of this chapter is the consequentialist approach, which is compatible with the stance that proponents of group privacy take.

of its members (i.e., non-aggregative interest). In Section 4.6.2.2, I demonstrate that the group interest cannot be interpreted as aggregative or non-aggregative. As a result, a group interest on which a group right to privacy is based cannot be defended for clustered groups. In the next section, I explore the conditions that should be met for a group to hold a right, and in the section following that, I demonstrate that a clustered group is incapable of satisfying the conditions to hold a right to privacy.

4.3. Approaches to Group Rights

In the analysis of rights, one of the most significant questions that requires an answer is as follows: ‘What is it that rights do for those who hold them?’ Two major theories describe the functions of rights: the interest (or benefit) theory (e.g., Raz, 1988), and the will (or choice) theory (e.g., Hart, 1982). Interest theorists maintain that a right makes the right-holder better off (Raz, 1988), whereas will theorists argue that a right makes the right-holder ‘a small scale sovereign’ (Hart, 1982, p. 183). According to the interest theory, the function of a right is to further or protect the interests of its holder (Raz, 1988). The will theory posits that the function of a right is to give its holder power by enabling them to have autonomous choices (Hart, 1982). In philosophical literature, there has been a long-standing debate about which theory is better at explaining the functions of rights (Wenar, 2023). However, the focus of this chapter is the interest theory of rights. This focus is compatible with the perspective that proponents of the group right to privacy defend (Taylor et al., 2017). According to this view, the function of the group right to privacy is to protect the instrumental interest of its holder in privacy (see Section 4.2).

According to the interest theory of rights developed by Raz (1988), the function of a right is to further and protect certain kinds of interests of the right-holder. An interest of a person (A) in x justifies attributing to A the right to x only if A’s well-being is sufficient reason to hold other person(s) under a duty to do whatever will promote the interest on which it is based. Interests sufficient to hold another subject to duty are protected and promoted through rights. The right is the ground of a duty, the ground that justifies holding that other person(s) have the duty. Children have a right to education, which entails a duty to provide education for children based on the interests of those children (Raz, 1988).

The analysis of rights is mainly focused on rights held by individuals, also known as individual rights. However, a right can be held by a group, which is known as a group right. In the literature on group rights, two approaches are described based on

how we understand the group that holds the right. These approaches are discussed in Sections 4.3.1 and 4.3.2.

4.3.1. Collective Approach to Group Rights

The first approach to a group right, known as the collective approach (Jones, 1999), views groups as a collection of individuals, and a right is jointly held by those individuals. Jones clarifies the concept of the collective approach to group rights in his 1999 paper on ‘Group Rights and Group Oppression’. In his subsequent book on *Group Rights* (2009), he develops this idea. According to the collective approach to rights, the right-holder is a collection of individuals who hold rights together as a group and not separately as individuals. For example, the right to live in a beautiful town is a collective right held jointly by those who live in such a town, which imposes a duty on the government to work towards achieving this goal. A group right conceived of in this way does not imply recognising the group as a whole has a moral status distinct from that of its members. Rather, the moral standing that enables a group to hold a right is the moral standing of the numerous individuals who jointly hold the right. Although group rights are held by the individuals who comprise a group, they differ from individual rights because they are rights the members of the group hold jointly rather than singly or independently; group rights are not just an aggregation of rights held individually by the members of the group (Jones, 1999, 2009).

4.3.2. The Corporate Approach to Group Rights

The second approach to a group right, known as the corporate approach (Jones, 1999), views the group as a unitary entity that holds rights. A group as a right-holder is conceived of as a moral entity with a moral status equal (French, 1984) or similar (Preda, 2012) to that of an individual person, and its rights are not reducible to the individual rights of those who constitute its members (Freeman, 1995). For example, a union, as a corporate entity, has the right to pursue its own interests even if those interests do not correspond to the interests of each individual member. These interests may include rules allowing the union to strike and requiring workers in a particular workplace to be members of the union. However, some individual members would be better off negotiating their own contracts outside of union membership, and it may be in some individuals’ interests never to strike. Hence, the corporate right held by the union differs from the right of the individual member to negotiate their own contracts or to choose not to strike (Newman, 2004).

What marks out a group as the type of entity that might bear rights is that it possesses an identity that does not change regardless of whether some or all of the persons in the group change or not (French, 1984; Newman, 2004). The unity or integrity necessary for a group to be a right-holder is found not only in its institutional characteristics but also in the common bond and sense of identity that its members share. A group that meets certain criteria—shared cultural characteristics, including language or religion, for instance—should be considered a unit with intrinsic moral rights (van Dyke, 1977). McDonald (1991) argues that if members of a group have a sense of being normatively bound to one another by virtue of their intersubjective experience, then the group is a unitary body that undergoes that intersubjective experience. A shared understanding that makes diverse individuals into a group includes features such as a shared heritage, language, belief, or social condition.

In general, the distinction between the collective and corporate approaches to group rights lies in their respective perspectives on whether group rights should be considered human rights or non-human rights. As Raz (1988) highlights, understanding a group's rights according to the collective approach is consistent with human rights. As this right is held by natural persons, it is consistent with human rights. However, according to Jones (2013), the corporate approach encounters a problem regarding human rights. Since we typically consider human rights as the rights of natural persons, corporate rights are not human rights because they are held by artificial (or legal) persons. However, proponents of the corporate approach to group rights may argue it does not matter whether a group right is a human right or not. They might instead suggest we can treat group rights as being distinct from human rights (Jones, 2013). The significance of this distinction for the argument in this chapter is that it helps to determine, if there is a group right to privacy, whether that right can be conceived of in a way that is consistent with the human right or whether it is necessary to recognise it as a non-human right, thereby requiring changes to relevant ethics guidelines and regulations.

In the following sections (Sections 4.4 to 4.6), my aim is twofold: first, to explain that, if a group has a right, the right is described as either a collective or a corporate right; and second, to demonstrate that the right to privacy of a clustered group cannot be conceived of as either a collective or a corporate right; I conclude that a clustered group cannot have a right to privacy. Having discussed why the right to privacy cannot be defended, in Section 4.7, I suggest employing the principle of protecting the vulnerable to protect the privacy of a clustered group. In this regard, instead of

focusing on the right to privacy, we need to consider privacy a moral principle aimed at protecting vulnerable clustered groups.

4.4. Exploring the Nature of Goods That Qualify as Objects of Group Rights

One way to resolve the disagreement between proponents of different approaches to group rights is to examine the nature of the good that is the object of a right. The collective approach posits that collective rights are rights to ‘collective goods’, whereas the corporate approach argues that corporate rights are rights to ‘participatory goods’. In the following sections (4.4.1 and 4.4.2), I discuss criteria that need to be satisfied to consider a good as either collective or participatory. By exploring different types of goods that serve as the object of a group right, I can determine whether clustered group privacy can be considered a collective good or a participatory good. Consequently, if clustered group privacy is deemed a collective good, then the right to privacy of a clustered group is a collective right, whereas if it is considered a participatory good, the right to privacy of a clustered group is a corporate right.

4.4.1. The Collective Approach to Group Rights: Examining the Nature of Goods as Objects of Collective Rights

In this section, I focus on the theory of Raz (1988), as he is a prominent proponent of the collective approach to group rights. Through a discussion of his view, I identify the essential criterion that must be met to adopt this approach to group rights concerning the nature of goods that are objects of collective rights.

Why should we make moral space for rights that individuals hold collectively but not separately? Raz (1988) argues that certain public goods necessarily have a group character. If a right to those public goods exists, then there must be a collective right. A public good is characterised by its non-excludability, which means that if a good is provided to anyone in a society, no member of the society can be excluded from benefiting from it. Nevertheless, different individuals may benefit from the goods to different degrees, depending on their characteristics, interests, and dispositions.

Raz (1988) distinguishes between contingent and inherent public goods. Contingent public goods’ non-exclusionary nature is due to contingent constraints on the present state of technology. It is logically possible to exclude some people from a contingent public good, but due to limitations of technological abilities, this is not a practical possibility. For example, clean air is a contingent public good; everyone

benefits, but only because engineers have not yet invented a way to control the distribution of clean air to each individual.

In contrast, inherent public goods are goods from which it is logically impossible to exclude any member of a society. The diffuse nature of the benefits of such goods derives from the general character of the society to which a person belongs. For example, the existence of a cultured society is an inherent public good because some aspects of the good of such a society meet the inherent non-excludability criterion. Although it is possible to exclude some people from benefiting from certain goods of a cultured society, such as libraries and art galleries, by excluding them from the society to which they pertain, that does not affect the character of a cultured society as an inherent public good. The enjoyment of the benefits of a cultured society, including the aesthetic or richer aspects of life, cannot be denied to any member of the society. Raz (1988) refers to inherent public goods as ‘collective goods’.

Raz (1988) argues that the collective nature of inherent public goods (or collective goods) makes them unsuitable as objects of individual rights. I contend that the diffuse nature of the benefits of collective goods makes it difficult to establish individual rights to them, as it is difficult to determine who exactly is entitled to these benefits. This reason counters the viability of individual rights to collective goods. If there is a right to an inherent public good, it must be a collective right. Raz opposes individual rights to inherent public goods but not to contingent public goods. The right of access to clean air, which is a contingent public good, is an individual right rather than a collective one (Raz, 1988).

Based on the above discussion, I emphasise the importance of the following criterion, referred to as C1, when adopting a collective approach to group rights regarding the nature of the good that a group has a right to:

C1: Collective rights are rights to collective goods. Collective goods are goods from which it is logically impossible to exclude any member of a society or a group from benefiting. Therefore, if a good is not a collective good, and if there is a right to that good, that right cannot be a collective right. This definition implies that individuals do not hold that right to that good collectively but individually.

4.4.2. The Corporate Approach to Group Rights: Examining the Nature of Goods as Objects of Corporate Rights

In the preceding section, I discussed the criteria for a good to qualify as an object of a collective right based on Raz’s (1988) viewpoint. In this section, I focus on Réaume’s

(1988) theory to identify essential criteria for a good to qualify as an object of a corporate right. Réaume critiques the inherent public good in Raz's theory and develops the notion of a participatory good.

According to Réaume (1988), there is no individual right to some public goods, but not because they are inherent public goods, as argued by Raz (1988). It is the nature of some goods that makes them unsuitable as objects of individual rights. For instance, the nature of a cultured society is such that it is unsuitable to be the object of individual rights. Goods of a cultured society are not only produced through the involvement of many but are also valuable precisely because of the joint involvement of many. In Réaume's (1988) view, an individual likely cannot successfully claim a right to these types of goods, which she calls 'participatory goods'; participatory goods must be held by groups rather than individuals.³² Hence, there may be individual rights to some public goods but only those that are not participatory.

An individual right exists when an individual's interest is a sufficient reason to justify imposing an obligation on others to protect that interest (Réaume, 1988). The interest in question is an interest in a good an individual has, and the good is of value to the individual and is enjoyed by an isolated individual. Hence, individual rights are claimed to those public goods that can be enjoyed individually, whether or not others enjoy them. According to Réaume, when the interest in a good is of importance to the person considered an isolated individual, regardless of whether the provision of the good requires widespread co-operation, the right to that good is an individual right. For example, the right to clean air is an individual right because clean air is a good that an individual can enjoy even if no else does, although clean air cannot be produced individually (Réaume, 1988).³³

Réaume (1988) clarifies her argument that no individual right is possible to some public goods by focusing on a good and its enjoyment. According to her, there is no

³² In her paper on 'Individual, Groups, and Rights to Public Goods' (Réaume, 1988), Réaume does not explicitly mention the corporate approach to group rights. Rather, her main focus is on the nature of the good being claimed, not on the criteria necessary to qualify as a right-holder. However, the way she discusses participatory goods suggests her view is categorised under the corporate approach. In contrast, Miller (2001) argues that rights to participatory goods are collective rights that can only be held by collectivities, not individuals. Miller defends the collective approach without relying on the interest theory of rights; the view is based on the teleological account tied to joint action. Miller's idea is not discussed in this chapter because the scope of this chapter is limited to the interest theory of right.

³³ This means that individuals, governments, and businesses contribute to the improvement of air quality.

individual right to a cultured society because the provision of such goods and their enjoyment require the participation of many people. A cultured society is a complex cluster of goods with a core upon which all other aspects of a cultured society depend. The core aspect of culture is that each individual needs others to enjoy it, not merely to produce it. A cultured society requires the existence of individuals who create and enjoy rock videos, read and write literature, compose, perform, listen to music, paint and sculpt, and so forth. The greatest value in a cultured society inherently involves the presence of others with similar interests in the arts, who devote their energies to culture, and with whom one can interact and share that culture. The value of such goods is partly constituted by a particular type of participation. As mentioned previously, Réaume calls such goods participatory goods, which involve activities that not only require many to produce but are valuable only because of that joint involvement. The core aspect of culture, that is, sharing cultural experiences, is participatory because it cannot be enjoyed individually, although it is enjoyed by individuals. Réaume admits there are some goods in the cluster of a cultured society that can be privately or individually enjoyed to a certain extent. In this respect, a cultured society also has an indubitable aspect that is conceptually capable of grounding an individual right. As a result, there is no individual right to the core aspect of a cultured society, not because it is an inherent public good and the interest of a single individual is weak regarding grounding duty on others, but because it is a participatory good and the individual has no interest as an individual in such a good.

Based on the discussion above, I emphasise the importance of the following criterion, which I refer to as C2, when considering a corporate approach to group rights, which focuses on the object of group rights, meaning the type of good that group rights pertain to:

C2: Corporate rights are rights to participatory goods. A participatory good involves the presence of others who take an active and genuine interest in that good with whom one can interact with and share that good. One cannot individually enjoy the benefits of a participatory good unless others with similar interests do too. A right to a good that requires the joint involvement of many in its production and enjoyment is a corporate right. Therefore, if a good is not a participatory good, and if there is a right to that good, that right cannot be a corporate right.

4.5. Exploring the Types of Group Interests That Ground Group Rights

I have discussed certain conditions concerning the nature of a good as an object of a right to recognise whether a right to that good is a collective or a corporate right. In this section, I examine the type of interests that ground a collective or corporate right. To adopt the interest theory of rights, it is necessary to examine the interests of a group that could ground group rights. The collective approach posits that collective rights are based on the interests of the members of a group, whereas the corporate approach contends that a group, as a distinct entity, has interests that ground corporate rights. In the following sections (4.5.1 and 4.5.2), I explore each approach more fully and consider how these approaches align with interest theories of rights.³⁴ By exploring the various types of group interests that ground group rights, I determine whether it is the interests of the members of a clustered group in privacy that ground the right to privacy or whether it is the interests of a clustered group as a whole. If the former is the case, then the right to privacy of a clustered group needs to be regarded according to the collective approach. However, if the latter is the case, the right to privacy of a clustered group needs to be understood according to the corporate approach.

4.5.1. The Collective Approach to Group Rights: Examining the Types of Group Interests That Ground Collective Rights

A collective right to an inherent public good exists if the following three conditions are met: first, the right exists because an aspect of the interests of humans justifies holding some person(s) to be subject to a duty; second, these interests are the interests of individuals as members of a group in a (*an inherent*) public good, and the right is a right to that public good because it serves their interests as members of that group; third, the interest of no single member of that group in that public good is sufficient by itself to justify holding others subject to a duty (Raz, 1988, p. 208).

The first condition is required for collective rights to be consistent with human rights. Collective rights serve the collective interests of members of a group. The second and third conditions distinguish a collective right from a collection of individual rights based on the nature of the interests in the question and their weight (Raz, 1988).

³⁴ The nature of a good that is an object of a group right and the interest that grounds a group right are interrelated, as members of a group have interests in a (collective) good that benefits them collectively, without excluding anyone from its benefits. To avoid the complexity that may arise discussing these two components of group rights together, I discuss them separately.

Regarding the second condition, one condition of the existence of a collective right is that the interests in question, assumed to be protected by the right, are the interests of individuals as members of a group in an inherent public good that is good to themselves as members of the group. Collective interests are the interests of individuals that arise from their membership in a group. The collective right is a right to a public good because it serves the interests of the group's members, as outlined by Raz (1988). For individuals to have interests that might ground collective rights, they must be members of a group and be aware of their membership in that group. For example, individuals who belong to a linguistic group have a collective interest in using their own language, and the interest of each person arises from their membership in that group. Without awareness of their linguistic group, individuals cannot have an interest that arises from their group membership, and their interests remain merely the interests of isolated individuals that cannot ground collective rights.

Considering the above discussion, I emphasise the importance of the following criterion, called C3, when adopting a collective approach to group rights regarding group interests that ground rights:

C3: Collective rights are grounded in group interests. Conceiving of group interests as the interests in the collective good of those who constitute the group requires that individuals be *members* of a group and be *aware* that they are members of a group before they can have interests that might ground collective rights. Therefore, if an interest cannot be considered a group interest, it cannot ground a collective right.

The third condition of collective rights, as outlined by Raz (1988), states that the interests of several members in a good that is good to themselves as members of the group are sufficient to ground those rights. According to Raz (1988), the right of a community to their own self-determination is a collective right. Although many individuals in a community might have an interest in the self-determination of their community (e.g., an interest in living in a community that enables them to express themselves in public without repression), the interest of any single individual is insufficient to justify holding others subject to a duty to satisfy that interest. The right to self-determination is grounded in the cumulative interests of many individuals within the community. Therefore, the existence of interest does not depend on the size of the group; the existence of collective rights and their strength do (Raz, 1988).

Collective rights represent the cumulative or aggregated interests of many individual members of a group. Individuals have an interest in the collective good only with others, rather than on their own. What makes that interest matter is it being the

interest of numerous individuals for whom it is good, but their individual interests alone cannot impose an obligation on others. Only when the combination of interests of a certain number of members reaches the threshold³⁵ required for the creation of a duty on others to act in a way that secures the collective good for them does it become a collective right.

4.5.2. The Corporate Approach to Group Rights: Examining the Types of Group Interests That Ground Corporate Rights

In this section, I focus on Newman's (2004) theory to identify the criterion that needs to be met for a group interest to ground a corporate right. Newman advocates the corporate approach to group rights, as opposed to Raz (1988). In contrast to Raz's theory, which characterises group interests as the aggregation of the members' interests, Newman conceives of group interests as non-aggregated; thus, corporate rights are grounded in non-aggregative interests.

Unlike Raz, Newman argues that the aggregative account, which reduces a collective interest to a summation of the members' interests, does not provide an appropriate conception of group interests. The problem with aggregate interests is they do not allow for individual differences; they presuppose a certain absence of diversity, and they can only work in a nonhomogeneous world. When interests are incomparable in terms of some overarching value, we cannot aggregate them. Aggregation, thus, does not provide a satisfactory description of the interests of a group that are founded on different values that cannot be compared, such as the values of different life courses. Newman suggests that an acceptable conception of group interests should be one in which individual differences and incommensurable values are considered. The group interest is not simply reducible to, or even an aggregative of, the interests of its members. Rather, such interest is a set of factors facilitating the fulfilment of the individual interests of diverse members at the same time. Newman argues that group rights must be grounded on group interests or the interests of a group as a whole, which is non-aggregative interest. As an example of group interest, consider a church as a corporate entity with an interest that is non-reductive to the interests of the members. If a church were suddenly abandoned by all its members, then it would, as a corporate entity, retain a residual right to freedom of religion as a public good. If a church did lose all its members, we would not accept

³⁵ In Raz's theory, there is ambiguity regarding how it is decided what the threshold is and who decides it.

that the church, as a corporate entity, immediately lost any moral interests it held (Newman, 2004).

Drawing on the above discussion, I highlight the significance of the following criterion, which I refer to as C4. This criterion is particularly important when taking a corporate approach to group rights that focuses on the interests of a group.

C4: The type of interest that can ground corporate rights is group interest. A group interest in goods must be a non-aggregative interest that is not reducible to the numerous interests of the numerous individuals who comprise the group. If a group interest cannot be considered a non-aggregative interest, it cannot ground a corporate right.

Conceiving of group interests as interests possessed by groups and going beyond or apart from those individuals is similar to what Jones calls ‘mysterious interests’ (Jones, 2010, p. 45). As Jones indicates, this view assumes a group is a supra-individual entity, and that its interests precede those of its members.

To summarise Sections 4.4 and 4.5, Table 4.1 outlines the criteria that need to be fulfilled when adopting either a collective or corporate approach to group rights.

Table 4.1. Summary of the Approached to Group Rights

		Approaches to Group Rights	
		Collective Approach	Corporate Approach
Reasons for Recognising Group Rights	Object of Rights: Nature of the Goods	C1: Collective goods: Goods from which it is logically impossible to exclude any member of a society or a group (Raz, 1988).	C2: Participatory goods: Goods produced through the involvement of many, and one cannot individually enjoy the benefits of such goods unless others with similar interest do (Réaume, 1988).
	Ground of Rights: Group Interests	C3: Aggregative interests: Aggregation of the interests of the members of a group can ground a collective right. To have an interest that arises from group membership, individuals must be members of a group and be aware they are members of a group (Raz, 1988).	C4: Non-aggregative interests: Non-aggregative interests of a group can ground a corporate right. Group interests are ascribed to a group as a whole, which requires the group to be considered a supra-individual entity whose interests are distinct from those of its members (Jones, 2010; Newman, 2004).

4.6. The Group Right to Privacy: A Collective or Corporate Right?

I have explained two approaches to group rights: collective and corporate. The importance of distinguishing between these different approaches lies in how we understand the group right to privacy as a human right, held jointly by individuals, or a corporate right held by a clustered group as a whole. According to these approaches, if a clustered group can have a right to privacy, then the right must be a collective or a corporate right. This section demonstrates that the consequent of this conditional statement is false. Accordingly, it cannot be claimed that the duty to refrain from accessing certain pieces of information about a clustered group to protect that group's privacy arises from the right to privacy of such groups, as these groups cannot primarily have such a right. Such reasoning sets the stage for the suggestion in the next section (i.e., Section 4.7), which advocates adopting the moral principle of protecting the vulnerable to respect the privacy of a clustered group.

To explore whether the group right to privacy can be viewed as a collective or corporate right, I address two issues that stem from the criteria mentioned in Sections 4.4 and 4.5. The first issue is related to the nature of the good, and the second concerns the interests of a group. Regarding the first issue, I investigate whether clustered group privacy can be considered a collective good, as described by Raz (1988), to satisfy C1, or a participatory good, as described by Réaume (1988), to satisfy C2. The second issue I investigate is whether individuals are aware they are members of a clustered group to have an interest that arises from their group membership, to satisfy C3, and whether clustered group interests transcend those of its members, to satisfy C4. By addressing these issues, I determine whether the right to privacy of a clustered group can be conceived of as a collective right, a corporate right, or neither of these.

4.6.1. The Privacy of a Clustered Group: A Collective Good or a Participatory Good?

The first issue to consider is whether clustered group privacy is a collective good, which is comparable to Raz's (1988) definition. According to Raz, a public good is a good that everyone can benefit from without exclusion, and a collective good is a good that is logically impossible to exclude anyone from benefiting from. However, clustered group privacy may not meet the non-exclusion criterion because it is possible for some individuals to be excluded from benefiting from it. For example, consider the following case: Although the privacy of a clustered group might be protected through the encryption of patterns uncovered through ML algorithms (i.e., encryption

of the ML model), an attacker can send a query to the encrypted ML model to deduce sensitive information about an individual—a specific datapoint—which is known as an inference attack (Islam et al., 2014). In such cases, the privacy of a clustered group is preserved because the model is encrypted to prevent individuals from accessing information about the clustered group. However, despite these measures, knowledge about an individual can still be acquired, risking their privacy. This case illustrates that clustered group privacy fails to meet the non-exclusion requirement of being a public good, let alone its other inherent criterion of being a collective good.³⁶

Nevertheless, from another perspective, (networked) privacy is a collective value, as explored by Regan (1995). Regan suggests that privacy is a collective value because technology and market forces make it difficult for any one person to have privacy without all people having a similar minimum level of privacy. This point implies all persons in a network must have a minimum level of privacy for anyone to have privacy in the network. This implication highlights the importance of protecting the privacy of all individuals in a network to ensure anyone's privacy is protected. It could be criticised that a person or company with significant power can retain their privacy while forcing others to reveal information about themselves, which is contrary to Regan's definition of the collective value of privacy. I think that what Regan meant is emphasising the necessity of the relationship between privacy protection for each person whose data are shared in a network and the rest of the network who share their

³⁶ An objection might arise that because 'individual privacy' is undermined, it does not follow that 'the good of clustered group privacy is excludable'. In response, I refer once more to the kind of information that group privacy and individual privacy protects to clarify that the case mentioned here demonstrates the exclusionary nature of clustered group privacy. Information pertaining to a group predicts the behaviour of the group as a whole; that behaviour has been revealed based on comparing and contrasting the behaviour of all members of the group. However, information that pertains to an individual predicts their behaviour based on analysing their past behaviours (Kammourieh et al., 2017). Group privacy protects information that pertains to a group, thereby preventing the acquisition of knowledge about its members. Such knowledge is obtained by inferring group characteristics about them, although knowledge about an individual might be acquired by analysing their personal information. In the case mentioned here, the member is excluded from benefiting from clustered group privacy through the inference of group characteristics about them. Given that the model is encrypted, and generalisation from information about a single person that results in group information representing the model is impossible, the person is excluded from the clustered group privacy while the group has privacy. Likewise, just as excluding members from libraries demonstrates the contingent nature of such a good, excluding members from clustered group privacy demonstrates the contingent nature of that good. This is unlike the diffuse nature of certain goods, such as the core aspect of a cultured society, from which the exclusion of a member is impossible.

data: for one person to have privacy, all persons in a network must have privacy. However, this view does not imply a necessity relationship between privacy protection for those whose data are shared in a network and those who hold or control the shared data, such as a company with great power. In Regan's view, privacy is a collective value because it is impossible for one to have privacy in a network while the privacy of others in a network is violated.³⁷

Consider the value of privacy on a social networking site from Regan's perspective. Through a social network analysis, the undisclosed attributes of a user can be inferred based on the disclosed attributes of the user's friends on social networking sites. For instance, a user's sexual orientation can be somewhat reliably inferred by analysing the nature of the relationships they maintain and their interactions with others. These inferences are made based on confirmed relationships among users (Barocas & Nissenbaum, 2014). The privacy of a user on a social networking site is protected only when the privacy of their group of friends is protected. Thus, privacy on a social networking site can be understood as a collective value, according to Regan's perspective.

However, in the case of clustered groups, there are no meaningful or recognised relationships among the individuals grouped together. The individuals are unaware of being grouped together based on the similarity of their features (see Section 4.6.2.1). This lack of relationships means privacy cannot be considered a collective value, according to Regan, in a clustered group.

Hence, the privacy of a clustered group cannot be understood as a collective value because it does not fulfil the non-exclusion requirement for a good to be conceived of as collective. Additionally, there are no meaningful relationships among the members of a clustered group, making it difficult to conceive of clustered group privacy as a collective value according to Regan's (1995) view.

Clustered group privacy cannot be considered a collective good, but can it be conceived of as a participatory good? To determine whether clustered group privacy can be considered a participatory good, it is important to examine whether the production of privacy and its *enjoyment* require participation from many individuals.

³⁷ I aim to clarify the non-excludable nature of privacy in Regan's argument, although she does not explicitly refer to it, but rather to its collective nature. Regan's argument suggests that privacy can be conceived of as a collective or non-excludable good, while clustered group privacy cannot, due to a lack of ties among the members of a clustered group. Accordingly, the important point is to distinguish clustered group privacy from the privacy that Regan defines and the consequent ways of conceiving them.

According to Réaume (1988), although participation by many is a necessary condition for a good to be considered participatory, it is insufficient. For instance, clean air is produced through the participation of many, but it can be enjoyed by an individual alone regardless of whether others enjoy it. Thus, clean air cannot be considered a participatory good. However, a cultured society is considered a participatory good because an individual has no interest in it as an isolated entity. My claim is that clustered group privacy is similar to clean air, rather than a cultured society, since it can be enjoyed individually without the involvement of others and, thus, cannot be considered a participatory good.

Overall, clustered group privacy cannot be considered a collective or participatory good. Fulfilling the outlined criteria is important because group rights are rights to a collective or participatory good. Given that the nature of clustered group privacy cannot satisfy either C1 or C2, I conclude that the right to privacy of a clustered group cannot be considered a group right. If a clustered group can have a right, that right should be a group right, meaning a clustered group cannot have a group right to privacy. Therefore, the disagreement between proponents of collective and corporate approaches to group rights regarding whether the right to privacy of a clustered group is a human or a non-human right is unfounded.

4.6.2. The Interests of a Clustered Group: Aggregative or Non-Aggregative?

The second issue investigated is related to the interests of a clustered group, which could ground a group right. According to Raz (1988), a collective right is grounded in the aggregation of the interests of individuals arising from their membership in a group, referred to as C3. This view requires that individuals are aware of their membership, and that their interests can be aggregated to establish a collective right. However, Newman (2004) argues a corporate right is grounded in the non-aggregative interests of a group ascribed to the group as a whole, referred to as C4, implying the group has interests beyond and apart from those of its individual members. To determine whether the group right to privacy can be a collective or corporate right, I examine the interests of a clustered group in privacy. First, I consider whether individual members of a clustered group can have interests arising from their membership that can be aggregated to justify a collective right. If the interest of an individual does not arise from their membership of a group, then the aggregation of interests means the aggregation of an isolated individual's interest, which cannot ground a collective right. Thus, it is important to explore whether members are *aware*

of their membership of a clustered group, which is a relevant criterion for determining whether a clustered group can bear a collective right to privacy. Second, I examine whether a clustered group can have interests that *transcend* those of its members to adopt a corporate approach to right. Having non-aggregative interests is a relevant criterion for determining whether a clustered group can bear a corporate right to privacy. In the following sections, I answer the following questions: ‘Is a clustered group self-aware?’ and ‘Does a clustered group’s interests transcend those of its members?’

4.6.2.1. Is a Clustered Group Self-Aware?

Raz argues the interests of members of a group in a collective good provide a ground for a collective right (Raz, 1988). If an individual has an interest in a good, they might have a right to it as an individual. However, if an individual, as a member of a group, has an interest in a good, they have a right to that good in combination with others. Since a public good benefits a certain society or group, an individual must be aware they are a member of a group to have interests as a member of that group in the good that benefits their group; otherwise, their interests might ground individual rights instead.

Perceiving oneself as part of a group can come from being aware one shares common characteristics with others. For example, members of a certain religious group recognise they share common characteristics and belong to the group. Moreover, individuals recognise they are members of a group because they themselves form the group. Examples of such groups include music lovers, bikers, and sports fans.

On the one hand, as Kammourieh et al. (2017) argue, members of a clustered group may be unaware of the specific sets of characteristics or attributes used to group them with others. Machine-learning algorithms cluster individuals based on similarities, such as purchasing certain items. An individual who buys a certain item may not know this action caused them to be grouped with others who made the same purchase. Thus, members of a clustered group may be unaware they share certain characteristics with others, and that this similarity caused them to be grouped together. Since any changes in the purpose of analysing data lead to the design of a new group, individuals may be grouped with others based on each action they take or each characteristic they possess. Being aware they are grouped with others based on certain actions or characteristics would require individuals to consider each feature or attribute of themselves as a group feature.

On the other hand, a clustered group is not formed by its members; the technologies used and their constraining affordances play a role in its formation (Floridi, 2017). Thus, individuals may not recognise themselves as members of a clustered group because they do not play a role in its formation.

Since individuals are unaware they are members of a clustered group and cannot recognise themselves as such, they do not have interests as members of that group in general or in clustered group privacy in particular. Therefore, the interest in privacy is an individual interest, which grounds an individual right to privacy rather than a group right to privacy.

As a result, a clustered group is not a self-aware group because its members are unaware of the specific characteristics used to group them together, and the group is not formed by its members to enable individuals to recognise themselves as a member of the group. Since an individual's awareness of their membership of a group is required to have interests that can ground collective rights, members of a clustered group cannot have interests that ground a collective right to privacy.

4.6.2.2. Does a Clustered Group's Interests Transcend Those of its Members?

According to Newman (2004), non-aggregative interests of a group must be protected by corporate rights, because these interests are attributed to a group as a whole, and the rights that secure these interests must be held by the group as a whole. To determine whether the right to privacy of a clustered group can be conceived of according to the corporate approach, it is necessary to establish whether non-aggregative interests exist for a clustered group and, if so, to secure those interests through a corporate right to privacy.

The idea of conceiving of group interests as non-aggregative interests assumes that a group is a supra-individual entity (Jones, 2010) with unity, capable of having interests in a good that go beyond or apart from those of its individual members. To assess whether non-aggregative interests exist for a clustered group, I examine two key issues. First, I determine whether a clustered group exists as an independent entity or whether it is ultimately reducible to individuals. Second, I investigate whether a clustered group has the integrity and unity required to ascribe interests to it.

The issue that first needs to be investigated is whether a clustered group is independent of its members or whether it can be reduced to a collection of its members. If the former is true, then a clustered group can be considered an entity that can have interests apart from and beyond those of its members. However, if the latter

is true, then the interests of a clustered group are simply a collection of the interests of its members. I argue we can only conceive of a clustered group having a right to privacy according to the corporate approach if the former is true.

To determine whether a clustered group is an independent entity or can be reduced to a collection of its members, as mentioned by Floridi (2017), I examine the most discussed distinction of realism versus nominalism, or holism versus individualism. In brief, realist views posit the existence of ‘kind-’ or ‘type-’level entities, whereas nominalist views deny the existence of such entities (Floridi, 2017). In various forms of realism, types or universals, such as groups, are genuinely existing entities distinct from their instances, such as individual members of a group. According to the realist view, types are objective and observer-independent and exist before the interest in identifying them is specified (discovered). However, the nominalist view only allows for the existence of tokens or particulars, such as individuals, rejecting the existence of types. According to the nominalist view, types are subjective, observer-dependent, and invented (Floridi, 2017; Gellner, 1959).

Floridi (2017) proposes a relationalist approach to the ontology of clustered groups, which takes a middle ground between realism and nominalism. He argues clustered groups are the outcomes of selections made by a data analyst on observables (information) for specific reasons and are linked to the constraining affordances offered by the technology used to analyse the gathered information. Clustered groups are not discovered or invented; they are *designed*. Such groups are the end result of a combination of the mind (of the analyst) and the physical world (of the technology). This perspective suggests clustered groups do not have a real, independent existence. However, they are not simply subjective entities that can be reduced to individuals; they are a combination of both objective and subjective elements (Floridi, 2017).

To determine whether a clustered group has a non-aggregative interest that necessitates a corporate right to protect that interest, the first issue that needs to be addressed is whether a clustered group exists as an independent entity or is ultimately reducible to individuals. According to the relationalist approach, a clustered group is not merely a subject-dependent entity but is designed through a combination of objective elements. However, even if I accept this metaphysical approach, which defines the existence of a clustered group in terms of the relationship between both subjective and objective elements, it is important to examine how it exists—whether it exists as an integrated or unitary entity to which an interest can be ascribed.

The second issue to investigate is whether a clustered group has the integrity required to ascribe an interest to it. This criterion is necessary because I claim that,

for group interests to transcend the interests of its members at any particular time, it must be a single, continuous entity over time, such that any changes in the group's membership do not change its interests. Otherwise, the interests of the group reduce to a mere aggregation of the interests of individuals who are its members at a specific moment. Hence, the following question arises: 'Can we think of a clustered group as an integrated or unitary entity?'

The integrity and unity of a group are found either through changes in membership having no effect on the group's identity (French, 1984; Newman, 2004) or in the common bond and sense of identity shared by its members (McDonald, 1991). When considering the design of clustered groups, a data analyst at time t_1 might select specific feature values of individuals for analysis using a technology. However, in t_2 , the feature values might change, resulting in the inclusion or exclusion of different individuals in a group and the design of different groups. Clustered groups are a type of aggregate collectivity, which French (1984) describes as a 'statistical collectivity', or a 'set' in Newman's (2004) view, and 'mere collections' in List and Pettit's (2011) view. The identity of the cluster is that of an aggregate; its identity rests solely in the aggregation. Any change in the cluster membership will always result in a change in the identity of the cluster. In contrast, for example, we can think of a football club or trade union that remains the same club or union over time, even though some members leave the group and others join it (Jones, 2010). Furthermore, a lack of shared understanding stemming from a shared social condition, for example, results in a lack of a sense of shared identity among its members. Hence, a clustered group cannot be conceived of as having a continuing identity. The interests of a clustered group can be understood as merely the summation of the individual interests at a specific time.

As a result, the interests of a clustered group in general (and interest in clustered group privacy, in particular) do not transcend those of its members, because the existence of a clustered group and the way it exists do not allow it to have interests, and if it has an interest, it is merely an individual's interest in privacy. Since a corporate right can be justified if a group has non-aggregative interests that require protection (C4), and clustered groups are incapable of having such interests, it follows that clustered groups cannot have the right to privacy in the corporate sense.

To summarise, in Section 4.6, I demonstrated that a clustered group right to privacy cannot be approached as either a collective or a corporate right, which entails that a clustered group cannot have a right to privacy. The reason for this finding was presented in two parts. First, I focused on the type of good that can be the object of a

group right, which must be either a collective good or a participatory good. A collective good can be the object of a collective right, whereas a participatory good can be the object of a corporate right. I explored whether privacy, as an object of the clustered group right, can be considered a collective good or a participatory good. I concluded that privacy does not fit either category.

Second, I focused on the interests that can provide a basis for group rights. The interests of a group can be either an aggregation of the interests of its members or non-aggregative interests that are beyond and separate from the interests of its members. The right that protects the aggregation of the interests of a group is a collective right, whereas the right that protects the non-aggregative interests is a corporate right. I explored whether a clustered group can have aggregative or non-aggregative interests, in general, and an aggregative or non-aggregative interest in clustered group privacy, in particular. To have the types of interests that ground a collective right, members of a clustered group must be aware that they are members of a group. However, I argued that a clustered group is not a self-aware group, which implies its members do not have the types of interests that ground a collective right. Hence, a clustered group cannot have a collective right to privacy. In addition, I explained that a clustered group cannot have non-aggregative interests in a good due to the way it exists, implying it cannot possess the interest that can ground a corporate right. Hence, a clustered group cannot have a corporate right to privacy.

If a clustered group can have a right to privacy, then the right must be a collective or corporate right, meaning I conclude that a clustered group cannot theoretically have a right to privacy. In the case discussed in Section 4.1, the group right to privacy cannot be considered a means of protecting the group against discrimination through respecting the privacy of the group, since it is implausible to assign such a right to the group. As a result, to protect the privacy of the group, a moral principle aimed at imposing limitations on accessing certain pieces of information about the group is required.

4.7. A Moral Principle to Protect the Privacy of Clustered Groups

I argued that a clustered group cannot theoretically have the right to privacy. Accordingly, the duty of protecting the clustered group privacy by refraining from accessing certain pieces of information cannot be entailed from the group right to privacy, as such a right cannot be recognised in the first place. Instead, I suggest establishing a moral principle for the moral obligation of protecting vulnerable

clustered groups as a requirement for respecting group privacy within an ethics of vulnerability. Accessing certain pieces of information about a clustered group that are likely to be used in morally objectionable ways to harm the group makes the group vulnerable. Since the source of vulnerability is the access to certain pieces of information, the duty owed to a clustered group as the vulnerable group is to refrain from accessing such information. Therefore, there is a need to expand moral duties to encompass vulnerable clustered groups, in addition to animals and environments, for example. Instead of thinking about duties engendered by the group right to privacy to limit access to certain pieces of information, we need to think about duties owed to the vulnerable to protect and respect them.

Rogers (2013) highlights the need for an ethics of vulnerability³⁸ to analyse the concept, articulate the sources and circumstances of vulnerability, identify those who are particularly vulnerable, and describe appropriate responses and protections for them. According to Rogers, such an ethics must clearly specify grounds for duties owed to the vulnerable. More specifically, in the areas of healthcare, she provides conceptual clarity regarding vulnerability to eliminate confusion about who falls under this category and what specific duties should be fulfilled to ensure their protection (Rogers, 2013).

A moral principle within an ethics of vulnerability might be Goodin's (1986) principle of protecting the vulnerable (PPV). According to this principle, we each have special obligations to protect those who are particularly vulnerable to our actions and

³⁸ In feminist ethics, the concept of vulnerability and care holds a central position. Certain feminists ground moral duties of care in response to those who are vulnerable. Vulnerability, in feminist ethics, is understood as an ontological condition of human existence, arising from our embodiment, neediness, and social and affective natures. Our obligations towards the vulnerable encompass the responsibility to provide care for them (Dodds, 2013).

Considering that the focus of this chapter is on clustered groups, I choose not to explore feminist ethics concerning the obligations towards these vulnerable groups, such as providing care for those groups, mainly due to the nature of vulnerability, which does not arise from an ontological condition of such groups. Instead, I adopt a broader perspective to define the source of vulnerability, arising from accessing certain pieces of information about clustered groups. Therefore, I explore duties towards the vulnerable from an ethics of vulnerability regarding protecting groups. In this context, I concur with Roger's assertion regarding the importance of articulating an ethics of vulnerability to identify the vulnerable and the corresponding duties owed to them for their protection. Moreover, the ethics of care as a feminist ethics seeks to preserve and promote an 'actual' human relationship between people (Held, 2006). Since an actual relationship does not exist between an individual in a target group and a data analyst or researcher, I do not adopt ethics of care to define a moral principle to care for the vulnerable.

choices. The PPV emphasises that vulnerability generates moral responsibility, compelling us to take measures to protect the vulnerable.

Since vulnerability arises from accessing information uncovered about a clustered group that is used to harm that group, there must be a moral response to limit such access. I claim that data analysts, who design clustered groups using ML algorithms, bear responsibilities³⁹ to protect the vulnerable. Privacy-preserving techniques can be employed to prevent access to information about the group. By utilising encryption measures, for example, data analysts can prevent access to information obtained at the group level. Patterns uncovered by ML algorithms must be encrypted, specifically the ML models. In this way, data analysts protect the vulnerable clustered group (see Section 4.1) from adverse policies and decisions made by the government. The use of encryption necessitates formulating policies and regulations to determine how and under what conditions key (i.e., an encryption key) distribution or key access to decipher encrypted models is implemented.

Vulnerability is contextual and variable, making it imperative to approach it as a matter of investigation rather than assumption (Rogers, 2013). To identify the vulnerable, data analysts must carefully consider the contextual factors that may indicate potential harm to a clustered group. For example, accessing information uncovered at the group level that identifies the correlation between specific purchasing behaviour and a particular sexual orientation may not render the group vulnerable in one society. However, in another society, such information could enable the government to formulate discriminatory policies targeting the group. To fulfil their moral duties in protecting vulnerable groups, data analysts first need to be aware of the contextual factors that indicate vulnerability.

4.8. Invasion of Group Privacy vs. Promoting Public Health

Critics may point out that limiting access to information obtained at the group level should not be imposed due to the significant benefits that access to information brings to society, particularly in promoting public health. The question that may arise is whether restrictions should be placed on accessing certain pieces of information that are more likely to be used in a discriminatory way to harm a group, considering the necessity of accessing group-level information for public health purposes. This point raises the broader question of how to reconcile the potential conflict between

³⁹ Similarly, parents are responsible for protecting the children to whom they gave birth.

protecting group privacy and advancing public health. Based on a careful consideration of how public health can be effectively promoted, I argue that the concern regarding a contradiction between protecting group privacy and promoting public health is unfounded.

Although there is no prevailing consensus on public health ethics in the literature, Rogers (2013) highlights that most scholars adopt a broad social justice approach, emphasising the need for special duties towards the poor who experience ill health as a result of systemic social disadvantage. Within public health practice, socially vulnerable populations are those more likely to face a heavier burden of ill health. Research on the social determinants of health (e.g., see *Just Health* by Daniels, 2007) has revealed that various forms of deprivation and disadvantage—such as economic, educational, financial, occupational, and social factors—are closely linked to poor health status. Given the presence of health inequalities associated with social vulnerability, the practice of public health prioritises addressing the health disparities stemming from systemic social disadvantage (Rogers, 2013).

To promote public health, the social determinants of health disparities and the underlying factors that contribute to these disparities need to be identified. Upon identification of these factors, policies and intervention efforts need to be formulated to address the source of health inequalities and, ultimately, promote public health. Regarding the knowledge of social determinants of health disparities required to improve public health, such insights can be gained from employing cluster techniques. Hence, a pattern must be uncovered from clustered groups that identifies underlying factors contributing to health disparities in a society. As I mentioned previously, this pattern is required to inform policymakers about how to formulate policies to improve public health. The breach of the privacy of clustered groups happens when the uncovered information is used in certain objectionable ways to harm these groups. However, I argue that, if such harm occurs, we cannot advocate for improving public health in society. The discriminatory use of the uncovered information stands in contradiction to the promise of promoting public health, which entails providing social justice for the vulnerable. Therefore, respecting the privacy of clustered groups is not in conflict with promoting public health.

To protect the privacy of clustered groups, it is necessary to articulate the PPV, which entails protecting and respecting vulnerable clustered groups. Vulnerability arises from accessing certain pieces of information about groups that can be used in morally objectionable ways to harm those groups. Therefore, the duty owed to those

identified as vulnerable in a specific context is to impose limitations on accessing the information.

4.9. Conclusion

Accessing certain pieces of anonymous information about a clustered group, enabling the group to be easily identified and targeted, which would likely be used to render the group worse off in morally relevant ways, harms the privacy of a clustered group. Protecting the privacy of a clustered group by imposing limitations on accessing certain pieces of information about the group leads to protecting the group against discrimination. This chapter argued the duty to refrain from accessing certain information about a clustered group cannot be entailed in the right to privacy of these groups, as they cannot primarily have a right. Instead, I proposed that the moral obligation to protect vulnerable clustered groups, as a requirement for respecting clustered group privacy, should be established regarding privacy and data protection guidelines and principles. The duty to respect clustered group privacy is not entailed in the group right to privacy of clustered groups but in the moral PPV clustered groups within an ethics of vulnerability. The necessity of limiting access to certain pieces of information about groups uncovered by ML models stems not from the need to respect the right of the group to privacy but from the imperative to protect groups from vulnerability, emphasising the paramount importance of protecting group privacy in the age of artificial intelligence.

Part II

Chapter 5:

Big Data as Tracking Technology and Problems of the Group and its Members

Abstract

Digital data help data scientists and epidemiologists track and predict outbreaks of disease. Mobile phone GPS data, social media data, or other forms of information updates such as the progress of epidemics are used by epidemiologists to recognise disease spread among specific groups of people. Targeting groups as potential carriers of a disease, rather than addressing individuals as patients, risks causing harm to groups. While there are rules and obligations at the level of the individual, we have to reach a stage in the development of data analytics in which groups are protected as entities. This chapter offers a deeper examination of harms to the groups.

Keywords: discrimination; GDPR; group rights; pandemic surveillance; right to privacy

This chapter is a modified version of the following publication:

Asgarina, H. (2023). Big Data as Tracking Technology and Problems of the Group and its Members. In K. Macnish & A. Henschke (Eds.), *The Ethics of Surveillance in Times of Emergency*. Oxford University Press. <https://doi.org/10.1093/oso/9780192864918.003.0005>

5.1. Introduction

Digital data help data scientists and epidemiologists track and predict outbreaks of disease. Mobile phone GPS data, social media data, or other forms of information updates as epidemics progress are used by epidemiologists to recognise disease spread among specific groups of people. Given the gravity of the risk that certain groups are exposed to, restriction of movement or surveillance could be imposed on them, as we have seen in recent years. In order to control outbreaks of disease, quarantine decisions are taken based on tracking the transmission of the disease on the group level (Taylor, 2016). For example, new data sources have been employed in high-stakes scenarios to track a range of life-threatening diseases, including cholera in the wake of the 2010 Haiti earthquake (Bengtsson et al., 2011), malaria transmission via network analysis (Tatem et al., 2009), and COVID-19.

In the case of the 2010 cholera epidemics, anonymous cell phone data were used to track and predict cholera epidemics in Haiti after the 12 January 2010 earthquake. Researchers used call records to investigate population movements after cholera struck coastal towns and surrounding areas, demonstrating that many who left these areas moved to cities. This knowledge was crucial because people leaving cholera-affected areas carried the disease with them (Bengtsson et al., 2011). Mobile phone records have also provided a valuable data source for characterising malaria transmission, enabling policymakers to modify and implement strategies for further preventing transmission (Liu et al., 2012). Using data from mobile phone networks to track population movements has therefore helped improve responses to disasters and disease outbreaks.

More recently, in response to the COVID-19 crisis, big data analytics helped public officials in making decisions about how to reopen society safely and how much activity to allow. To accomplish this, epidemiological models that capture the effects of changes in mobility on virus spread have been developed by reflecting on patterns of human interaction at non-residential locations of interest, such as shops, restaurants, and places of worship. The results of such findings could be used to infer which activities should be continued and which should be avoided. According to the model, infections in venues such as restaurants, gyms, and religious establishments play a disproportionately large role in driving up infection rates, restricting the reopening of such establishments, and making them a key target for control (Chang et al., 2021). As a result, big data analytics as tracking technologies can help authorities control and manage the COVID-19 pandemic and bring a premature end to epidemics.

Policymakers and authorities use information derived from big data analytics to target groups or persons. When an entity is the target of information, this means that observers, policymakers, or authorities have information that they relate to an entity in the world (Henschke, 2017). The observer uses the information to target the person who is infected and the person who is at risk of infection because they have been in contact with the infected person. Moreover, the observer can target groups as potential carriers of a disease, rather than addressing persons as patients.

Though promising, pandemic surveillance brings a series of challenges for those targeted by the information derived from big data analytics. Targeting persons and groups risks causing harm to a person, as a member of a group, and to a group qua group. Three of the ethical issues raised by targeting a person with the information generated at the aggregate level are consent, social justice and fairness, and privacy. The negative consequences of data processing at group level are the risks of group discrimination or stigmatisation. In these types of cases, the problem is not that this or that specific person has been harmed, but that the group as a whole is affected and thereby undermined (van der Sloot, 2017). These ethical issues and harms are discussed in the following sections.

These risks are well known, and predate the COVID-19 pandemic. The EU General Data Protection Regulation (GDPR) is considered a key to the successful development of technologies to tackle the COVID-19 pandemic (Mikkelsen et al., 2020; Newlands et al., 2020). The consent of a person to the processing of their health data is discussed in Article 9. Articles 21 and 22 address concerns about discrimination. To protect a person's privacy, Article 4 identifies which types of information should be kept private. I show that none of these principles can protect a person from the harms that arise when they are the target of pandemic surveillance. These suggest that a specific regulatory framework be developed, focusing on safeguarding information attributed to a person because they are a member of a particular group.

The cluster-type (or statistical) groupings designed by big data analytics are sources of information for making policy decisions without focusing on individual identifiability. Regarding this, obligations or regulations developed to protect individuals from the misuse of their data are not helpful at the level of the group, as groups created by algorithms or models expose those groups to potential harms without identifying individuals. Furthermore, current rules or regulations cannot protect groups against potential harms, partly because they focus on individual data protection concerns, and partly because many of the uses of big data that involve

algorithmic groupings are so beneficial in furthering scientific research and improving public health. These suggest that while there are rules and obligations at the level of the individual, we must reach a stage in the development of data protection where groups are protected against discrimination. While Mantelero (2017) argues that the group should be granted the privacy right in order to limit the potential harms that can result from invasive and discriminatory data processing, I here investigate the feasibility of assigning group rights to the group clustered by big-data analytics as a means of protecting that group against discrimination.

In the first part, I look at the ethical issues raised by aggregate-level conclusions generated from big data that target people as members of groups, and groups qua groups. The second part offers recommendations for how to improve current safeguards for persons as members of specific groups and for groups as a whole.

5.2. Key Ethical Issues

In this section, I first look at ethical issues raised by aggregate-level conclusions derived or discovered from big data while targeting a person as a member of a group. Three of the ethical issues, consent, social justice and fairness, and privacy, are discussed in this section. Second, I look at ethical issues raised by the targeting of a group qua group. Group discrimination or stigmatisation are discussed in this section. I acknowledge that there are other ethical issues not listed here, and so this list is not intended to be exhaustive. However, it covers the major issues that arise in the literature.

5.2.1. Ethical Concerns Raised by the Targeting of a Person as a Member of a Group

This section deals with ethical issues that arise due to a person being targeted as a member of a group. To approach this, I first provide a brief overview of the various types of groups created by data technologies. The distinction between different groups enables a clearer explanation of the ethical issues. Data technologies are used to discover new patterns and relations in data. Those patterns and relations may concern numerous entities leading to profiles being formed, which in this context would be profiles of people. A profile which is a property or collection of properties of a particular group of people is known as a group profile. Group profiles are divided into two types concerning the distributivity of properties forming group profiles. First, if a property is valid for each individual member of a group, this is called distributivity or

a distributive property. Second, when a property is valid for the group and for individuals as members of that group, though not for those individuals as such, this is called non-distributivity or a non-distributive property (Vedder, 2000).

Distributive generalisations and profiles attribute properties to a person, or a group of people, in such a way that these properties are actually and unconditionally manifested by all members of that group. For example, having a bad health condition may be distributed among all members of a group (those who have that condition). Non-distributive generalisations and profiles, on the other hand, are framed in terms of probabilities, averages, and medians, or significant deviations from other groups. They are based on comparisons of group members with one another and/or comparisons of one group with other groups. As a result, non-distributive generalisations and profiles differ significantly from distributive generalisations and profiles. Non-distributive generalisations and profiles apply to people as members of the reference group, but these individuals do not have to display these properties in reality (Vedder, 1999). For example, in epidemic research, a property may be assigned to a patient because the person belongs to a reference group, such as having a specific disease, which is non-distributive profile information, even when the patient does not get sick from the disease. In such a circumstance, the person is being judged and treated on the basis of belonging to the ‘wrong’ category of persons.

In a distributive group profile, each individual member of the group is examined, the property discovered is assigned to each member, and the group inherits the property. For example, each patient in a group is diagnosed with a certain disease based on the presence of a certain symptom, and the property is then assigned to the entire group. We can conclude that the group inherits a distributive profile shared between all members of the group. However, in a non-distributive profile, the pattern or property discovered in a group is only distributed among parts of the group. In such cases, though, the property is ascribed to each member of a group because they are the members of the group (Vedder, 2000), and not because they necessarily have that property. As a result, while the probabilistic property is ascribed to the group, attribution of that property to each and every individual member is invalid because that property may or may not ascribe to a particular person in the group. For example, when a group profile states that 90% of the patients in the group have a particular symptom, no one can tell on those data alone, which patients actually do have the symptom. The link connecting the non-distributive profile to the individual to whom the group profiles may apply is opaque. Hence, this type of group profiling represents a group and reveals attributes that may (or may not) be applicable to the individuals

in the group, and is only applicable to the group as such (de Andrade, 2011). Thus, assigning a non-distributive group profile to a group does not imply assigning that property to each of the group's members, implying that a group and its members do not share the same property.

I can now turn to the ethical issues that arise when a person is targeted using information derived from big data analytics.

5.2.1.1. Consent

Consent has been a point of debate and concern since its position of dominance in the post-Second World War Nuremberg Code, a set of ethical principles for human experimentation to ensure that harms to humanity like those in Nazi 'medical' experiments would never occur again (Annas & Grodin, 1995; Macnish, 2019). The purpose and justification of consent provisions are to provide reasonable assurance that a patient or research subject has not been deceived or coerced (O'Neill, 2003). Hence, when research is aimed at impacting the conditions of its subjects, it is necessary to pay attention to research subjects' consent and awareness.

The function of consent in the big data era should be to help reduce harms associated with targeting members of a specific group. An example of potential harm perceived on group membership and not on individuals is tracking migrants fleeing a capital city in order to target cholera prevention measures (Bengtsson et al., 2011) through restriction of their travel. In this case, the question arises of how to manage big data sources in terms of consent and awareness among research subjects—as members of a specific group. To gain a better understanding of the issue, consider how group profiles are designed once more. Big data analytics are used to design group profiles to help control disease outbreaks, which are often based on fluid and contingent factors such as postal code, health status, and being in a public place at a specific time. In such cases, groups are not stable but fluid, and they are not unique or sparse but rather omnipresent and widespread. Group profiles can be designed in a fraction of a second and changed by changing the purpose and needs of grouping individuals in a specific way, so who is in and who is out of a group profile can change frequently (Floridi, 2017; van der Sloot, 2017). Thus, the issue is how to seek and obtain consent when members of a group may be unaware that they are part of a group and are included in a group because they share characteristics such as being in the same place at the same time.

In the context of big data analytics, there are two main limits to obtaining consent from those who are surveilled and grouped in a specific way. First, due to the

unforeseen inferences drawn from data analytics, the possible risks and benefits might not be anticipated or anticipable at the time of initial data collection. Second, the problem stems from an inability to provide individuals with the option to choose which types of groups they want to be a part of and then make group decisions based on that. While novel approaches to consent are being developed (e.g., dynamic consent, open consent, e-consent, Budin-Ljøsne et al., 2017; and Kaye et al., 2015), there is still a lack of giving individuals the choice to decide whether to be a member of a specific group simply because they share characteristics with other members of an algorithm-designed group.

5.2.1.2. Social Justice and Fairness

Group profiles, in this context, are designed and used only for pandemic research purposes, with guarantees that access to them is restricted to some researchers who do not share the information with others. However, things change when these guarantees are not present. The information in the profiles may then be made available to others, becoming part of the body of public knowledge in society, or the information may be used for entirely different purposes. For example, the information derived from people's health data could be used for other purposes and by third parties: for job selection procedures, insurance, loans, determining who can and cannot get back to work, or determining who can and cannot access public spaces like subways, malls, and markets (Morley et al., 2020; Sharon, 2020; Vedder, 2000). If this type of mission creep (Mariner, 2007) occurs, then values of social justice and fairness are at stake.

First, when the allocation of goods and amenities in society is based on health criteria, social justice is at issue. Generalisations and profiles can be used to help public and private entities formulate policies, or they can be absorbed into public knowledge. When the information contained in the generalisations or profiles is sensitive in nature, the situation becomes more complex because it might render members of the group vulnerable to prejudice or it may be used to make decisions regarding the allocation of scarce welfare resources. Information about people who have a high risk of developing certain diseases, especially those which may indicate a likely lifestyle, for example, can lead to stigmatisation and prejudice. This information might be used to provide or restrict access to services such as insurance, loans, or jobs for members of a specific group. As a result, social justice challenges arise from some of the policy reactions to the information discovered from group profiles (Vedder, 2000).

Second, fairness is at stake because an individual may be judged or treated based on merits or characteristics that he or she did not acquire voluntarily, such as a poor

health condition. However, because the feature is one of the group and not necessarily of the individual, a person as such may not exhibit or even experience those characteristics at all. This occurs when non-distributive generalisations and profiles are used instead of distributive generalisations and profiles.

5.2.1.3. Privacy

Data technologies are used to find patterns or relationships in a dataset through maximising dissimilarities between groups and optimising similarity within a group (Aouad et al., 2007). As mentioned above, the patterns or relationships uncovered could apply to various entities, resulting in the formation of individual or group profiles. Group profiles may be used to infer characteristics to individuals (Henschke, 2017). For example, the aggregation of data may result in the knowledge that those with low oxygen saturation may be more likely to be infected with COVID-19. Thus, algorithms design a group with low oxygen saturation, which is labeled as having a high risk of infection. Consider the case where a person's data were collected, stored, and processed, and the information 'low oxygen saturation' is attributed to him or her. This information might help clinicians make early decisions regarding the arrangement and organisation of medical resources and early interventions to improve the health outcomes of this patient (Benito-León et al., 2021).

However, inferring group characteristics to individuals threatens the privacy of the individual as a member of a group. Inferred information tells us something about individual members of those groups in a very qualified way (Vedder, 2000), assuming that the information is produced in a sound and reliable way. When an individual member intends to keep that information private, or when the information inferred is contrary to an individual member's preference, the privacy of members, rather than individual privacy, is threatened. The reason for this is that issues of individual privacy arise when the information derived is uniquely about a specific individual, meaning that the link between that individual and the information derived is strong. However, there are privacy issues when the link between the information derived and that individual is weak, especially in a non-distributive group profile, meaning that the information produced could have been formed from another source. In such cases, privacy claims are derived from group claims following the aggregation of the data (Henschke, 2017). As a result, given the lack of direct connection to the individual source, inferring group characteristics to individuals in situations in which a person is not a unique source of data threatens the privacy of members, implying that a more

in-depth examination of how the privacy of groups' members is considered in the context of data protection is required.

5.2.2. Ethical Concerns Raised by Targeting a Group qua Group

In this section, I look at group discrimination or stigmatisation when a group is targeted. Consider an epidemic that appears to target certain minorities disproportionately, resulting in additional restrictions being imposed on those minority groups, regardless of whether members of the group have the disease. In the early period of the AIDS pandemic, people who were heroin users, male homosexuals, hemophiliacs, and Haitian were seen to be most at risk of contracting AIDS, and so membership in the '4H club' led to significant risk to members of those groups (Garrett, 1996). In what follows, group discrimination or stigmatisation is discussed.

5.2.2.1. Group Discrimination or Stigmatisation

Contact tracing apps, GPS ankle monitors and other wearables, cell phone location data collection, genomic testing, and targeted quarantines, among other bio-surveillance technologies being used to respond to the COVID-19 pandemic, have the potential to exacerbate discrimination against racial minorities and immigrants. As a result of the COVID-19 pandemic, racial disparities in health outcomes have increased, while communities of colour, immigrants, and other marginalised groups have been blamed for spreading the disease. Disturbing disparities in COVID-19 surveillance of racial minorities have emerged, for example, in the United States. In New York City, Black or Latinx people made up 92% of those arrested for violating COVID-19 protocols, such as social-distancing requirements. Black people were targeted by government authorities at four and a half times the rate of White people for such violations (Sundquist, 2021). As a result of 'inappropriate surveillance' (Macnish, 2012), certain population groups, namely immigrants and certain non-White racial groups, are discriminated against and blamed for disease outbreaks, which may represent a biased evaluation and become a source of social discrimination.

Making inferences and drawing conclusions about groups based on an extensive collection of information threatens the group's privacy because revealing this information increases the risk of potential harm to the group itself. Hence, the surveillance technologies used in the fight against COVID-19 have an impact on the privacy of some groups, such as marginalised communities. That is, even if all

members of a marginalised group are individually protected from unwanted intrusion and targeting, the group as a whole is not protected against disproportionate surveillance, implying that individual privacy can be effectively protected while the group as a whole is not adequately protected.

Consider a situation in which each individual knowingly shared his or her data and agreed to the type of processing to be performed at the time. Assume that the lawfully obtained and lawfully processed set of personal data enabled an analyst to draw sophisticated inferences—say, on the likelihood of disease outbreaks among populations—predicting the behaviour of a group of individual data subjects as a group. Such inferences would be based not on analysing past individual behaviour to predict future individual behaviour, but rather on comparing and contrasting the behaviours of all members of a group defined by one or more shared characteristic (Kammourieh et al., 2017). Disclosing information discovered about a group therefore increases the risk of harm to that group’s privacy because it increases the risk of discrimination against the group.

5.3. Current Measures to Address the Identified Issues

In this section, I look at the current guiding regulation regarding data protection, the GDPR, to explore the suitability of existing legal frameworks to address and mitigate the identified issues. I demonstrate that further work is required to address the identified issues and that specific rules or regulations need to be developed that differ from those already existing regulations in the field of data protection.

5.3.1. Protecting Persons against Harms

5.3.1.1. Consent: Article 9 of the GDPR

Article 9 provides the legal ground for special categories of personal data in the context of epidemics. Processing of special categories of personal data, such as health data ‘for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health’ is allowed (EU Parliament, 2016, Art. 9). These special categories of personal data are processed for reasons of public interest without the (explicit) consent of the data subject. This Article is unable to address the identified consent issues and instead introduces a new issue in the form of the privacy–health trade-off.

Giving up privacy for public health stems from a trade-off between privacy and security, which was framed after the World Trade Center attacks in 2001. In that debate, the tangible threats of terrorism justified an expansion of governments' surveillance powers, and the creation of global surveillance that still exists today (Ross, 2020). Scholars also wonder what will happen to the epidemiological surveillance constellation once the pandemic is over and whether the erosion of privacy will become part of a permanent state of surveillance against new viruses (McGee et al., 2020). It is crucial to know how much privacy should be given up for public health, which then determines the governments' surveillance power over citizens. As they stand, the current Articles are unable to address the identified consent issues (see section 5.2.1.1) and instead introduce a new issue in the form of the privacy–health trade-off.

5.3.1.2. Justice and Fairness: Articles 21 and 22 of the GDPR

Profiling and discrimination concerns are reflected in the GDPR, especially in Article 21. This Article introduces the right of data subjects to object to personal data processing, including profiling, at any time. If the purpose of data processing is direct marketing, the data subject will have a right to object. In all other cases, data processing must stop, unless the data controller demonstrates compelling legitimate interests for the processing that overrides the interests of the data subjects (Wachter, 2018).

However, the scope of the Article is limited to individual profiles that analyse or predict specific aspects of natural persons without taking into account harms that arise when a person is considered as a part of a whole group, particularly non-distributive group profiles in which the analysis or prediction is performed by comparing and contrasting the behaviour of all members of a group, rather than predicting behaviour of a specific person based on his or her available data.

Article 22 introduces safeguards against automated decision-making, including profiling, but only when data processing is solely automated and has legal or similarly significant effects. The applicability of these safeguards seems to be very limited because 'solely automated' and 'legal or similarly significant effects' remain undefined in practice (Wachter et al., 2017). Correspondingly, the scope of data protection law needs to be defined to offer a more promising approach for data subjects to maintain control over how the data is used for services and future opportunities.

5.3.1.3. Privacy: Article 4(1) of the GDPR

Article 4(1) determines which types of information are protected by GDPR. Personal data allowing for identification of a natural person, including online identifier or factors specific to the physical physiological, genetic, mental, economic, cultural, or social identity of that natural person, are protected. What follows from the definition of personal data is that there is some direct connection between a particular person and his or her data. In other words, the right of controlling data applies to personal data in the strict sense of data relating to an identified or identifiable person.

Nevertheless, applying the narrow definition of personal data and protective measures in terms of safeguarding an individual's control over group profiles is problematic. The information derived from a group profile cannot be about an individual based on the available data about the individual (Loi & Christen, 2020). This means that what must be protected is no longer raw data such as names, as an identifier, but rather valuable information that can be inferred from datasets (Kammourieh et al., 2017). Furthermore, the information is primarily ascribed to a person on the basis of belonging to a group, meaning that the links between that individual as the source and the information generated are weak (Henschke, 2017). As soon as the data have ceased to be personal data in the strict sense, it is not clear how the principle should be applied. For example, the right of controlling data does not apply to information derived from personal data (Vedder, 2000). As a result, in the age of big data and information inferred, the interest in informational privacy no longer provides sufficient protection to the individual members of a group; it focuses solely on information collection rather than analysis of aggregation data (Kammourieh et al., 2017).

5.3.2. Protecting Groups against Harms

According to Mantelero (2017), group privacy is the right to limit the potential harms to the group itself that can derive from invasive and discriminatory data processing. At the group level, the right to privacy can be perceived as a duty of the state not to use its powers arbitrarily. A group right to privacy prevents the arbitrary use of power, such as discriminating illegitimately between different groups in society or exercising power for no reason at all (van der Sloot, 2017). Understanding group privacy in terms of protecting groups against the possible negative consequences of generalisations and profiles cannot be reduced to individual privacy, meaning that the protection of group members cannot protect the group itself.

It could be asserted that, in some cases, the protection of individuals can protect specific groups. The GDPR, for example, has the potential to provide safeguards against groups. It provides enhanced protection for certain types of highly sensitive data, including ‘revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, and [...] the processing of data concerning health’ (EU Parliament, 2016, Art. 9). While this is a protection granted to the individual, its effect is also to protect specific groups that are more vulnerable to targeting (Kammourieh et al., 2017). As a result, the GDPR has the potential to limit the discovery of information about existing groups, such as racial groups.

However, the GDPR is mainly focused on protecting individual identity and on safeguarding personal information. In an era of big data, where information about groups is extracted from data, or where more information is discovered about existing groups, the individual is often incidental to the analysis. Thus, the problem is not that this or that specific person has been affected, but that groups have been harmed. Since the group is exposed to the risks derived from the creation and use of inferred data, the infringement takes place at the group level while the rights and remedies are granted at the individual level (Kammourieh et al., 2017; Taylor et al., 2017; van der Sloot, 2017). Regarding this, defenders of the group right to privacy, for example, Floridi (2017), Mantelero (2017), and Taylor (2017), highlight the importance of assigning rights to a group to protect that group against discriminatory harm. As Floridi (2017) points out, granting this right to groups is different from the existing rights in the fields of privacy and data protection in that this right to privacy is not reducible to the privacy of the individuals forming such groups.

The preceding discussions highlight the importance of developing group rights to privacy to address issues that revolve around the risks of discrimination and the adverse outcomes of big data analysis. In what follows, I discuss the feasibility and problems of ascribing the group right to privacy to a clustered group.

5.3.2.1. Group Rights to Privacy

So far, I have discussed that, in contrast to how privacy has traditionally been conceived on an individual level, the era of big data raises new questions about privacy on a group level. In such cases, access to personally identifiable information of individuals is less likely to cause harm. Harm is more likely to occur when authorities or corporations draw inferences about people on a group level. As a result, the concept of privacy must be stretched and reshaped in order to help us think about groups. Floridi (2014) was the first to bring up the concept of group privacy in relation to big

data analytics insights. He argued that it is crucial to investigate whether groups have privacy rights that are not reducible to the privacy of the individuals who make up those groups.

According to Floridi's argument, a group right to privacy is a right held by a group *qua* group rather than its member severally; it is referred to as a group right in the *strong* sense, the corporate approach to group rights. Right-holding groups are conceived as moral entities in their own right, with a being and status similar to that of an individual person. This viewpoint holds that a group has an identity and existence distinct from its members. Accordingly, unity and identity are necessary for a group to be the type of group that can bear rights (French, 1984; Newman, 2011). For example, French (1984) contends that the Gulf Oil Corporation's rights and responsibilities in purchasing or selling property, or in being responsible for environmental pollution and cleaning it up, are not reducible to the individuals currently associated with it. Organisations of this type have identities that are not exhausted by the identities of the people who work in them; one person leaving and another joining does not form a new organisation. As a result, a group's unity or identity distinguishes it as the type of group that might have rights.

However, proponents of group rights in the *moderate* sense, the collective approach to group rights, such as Raz (1988), argue that groups are not conceived as having independent standing, but rather as having rights shared in and held jointly by the group's members, rather than being a mere aggregation of rights held by the group's members individually. The individuals who comprise the group have a right that none of them have as independent individuals. In this view, collective rights are ascribed to a specific collection of individuals because there are some sorts of public goods that can only be held by the collectivity (Raz, 1988). In respect of the publicity production of such goods, participants in a participatory activity possess collective rights. For example, the provision of a cultured society requires participation amongst members of the group; each individual needs others in order to produce it. Accordingly, there is no individual right to a cultured society, but rather participants in a joint action possess collective rights (Miller, 2001).

I argue that, in the case of a group designed by algorithms or data technologies, it is implausible to regard a group right to privacy as a group right in either the strong or moderate sense. The reason for this is that in this kind of group, the essential criteria for both strong and moderate approaches on group rights to privacy are not met. On the one hand, because of the lack of integrity or unity needed to hold a right, a group right to privacy cannot be described based on strong approaches. A group right to

privacy, on the other hand, cannot be conceived based on moderate approaches because members of the relevant group cannot perform a joint action to produce any good simply because they cannot realise the condition required to constitute a joint action, which is 'believing that their action is dependent on the action of other members' (Miller, 2001, p. 57).

A group right to privacy (for a group designed by algorithms) cannot be theoretically explained using either strong or moderate approaches to group rights. Instead, can we adopt methodological approaches to argue that the need for a moral group right to privacy is practical? This implies that a moral group right to privacy does not objectively exist in the sense of being independent of human beings but that it is rather constructed by human beings to solve practical problems. Subsequently, the question is, 'Can a moral group right to privacy be the outcome of a construction procedure?' To respond to this question, I employ society-based constructivism, as developed by Copp (1995), which argues that no moral codes are independent of a society's construction procedure.

Society-based constructivism posits that moral codes are rationally chosen by a society as a whole to contribute to its ability to meet its non-moral values and needs. The normativity of a moral code, which contributes to the realisation of a society's values, should be explained in such a way that a rational society should choose to serve its non-moral values and needs. Accordingly, 'a moral code is justified in relation to society just in case the society would be rationally required to choose the code to serve in it as the social moral code, in preference to any alternative' (Copp, 1995, p. 104).

Society as such makes a choice or has a preference when two conditions are met. First, there is a property that society could have in principle. Second, members of the society are nearly unanimous in preferring that society have the property. The first condition distinguishes between individual preferences (e.g. buying lottery tickets) and societal preferences (e.g. democracy) within a society. In the first case, most members of society would prefer the same option or property for themselves, which would be interpreted as a claim about most members' preferences rather than the preferences of society as a whole. In the second case, however, democracy is a possible societal preference and may be an option for society. The second condition requires a consensus over societal options among members of society, which does not necessarily mean that each member would prefer a societal option but rather that they are nearly unanimous in preferring an option for society. When there is a property that a society could have and its members are nearly unanimously in favour of the society having this property, then the society as a whole prefers this property (Copp, 1995).

The above discussion does not imply that no moral code is justified unless there is a society that is rationally required to choose one. As Copp points out, ‘if a group is neither a society nor a part of a society, then a code is justified as a moral code in relation to it just in case, if the group were a society, the group would be rationally required to select the code to serve in it as the social moral code, in preference to any alternative, to satisfy its non-moral values’ (Copp, 1995, p. 122). It may be rational for a group of people who have no inclination to interact with each other and thus do not qualify as a society to choose a code as a moral code.

Although there is no interaction between members of a clustered group, as this group is statistically designed using algorithms and does not form part of a society, they can choose a moral right to privacy. If members of a clustered group were to be part of a society, this right would satisfy their non-moral values, including exchanging information, sharing feelings and emotions, making plans, and taking collective action to achieve their goals. This is the right to confidentially associate with others, which Bloustein (2019) referred to as the right to family or relational privacy. Maintaining the confidentiality of information that members share among each other emphasises the importance of privacy to the success of interpersonal associations; this is a collective action that cannot be accomplished by isolating one member of a group from the rest. This implies that family or relational privacy can be interpreted as a societal property and that members of a clustered group would nearly unanimously prefer to have a right to family or relational privacy, which would enable them to privately associate with each other.

While members of a clustered group would choose the right to family or relational privacy as a moral code to satisfy their non-moral values, it is important to distinguish this right from a clustered group right to privacy. The latter aims to protect certain information about a clustered group uncovered through the analysis of similarities within the group and dissimilarities between it and other groups. Therefore, it is the right to family or relational privacy that might be constructed by a society, which is distinct from the moral group right to privacy.

In conclusion, while it is important to protect certain information about a clustered group to prevent discrimination, the group right cannot be recognised as a means of protecting the group against discrimination by protecting this information; rather, other means or moral principles must be taken into account.

5.4. Conclusion and Recommendations to Improve Current Measures

Big data analytics have the capacity to uncover new information, find patterns and predict behaviour, allowing for the algorithmic creation of totally new groups. In this regard, it is necessary to reconceptualise the risk of data harm to include the problem of the group and its members. For researchers, it is difficult to manage the source of big data regarding consent and awareness on the part of research subjects. A further problem is that the application of data technologies undermines the values of social justice and fairness, since an individual may be judged or treated on characteristics they did not acquire voluntarily (or at all). Finally, because data technologies are used to target people as members of specific groups rather than individuals, they are increasingly threatening group members' privacy rather than individual privacy. In addition to the issues that arise when a person as a member of a specific group is targeted by the information derived from a group profile, there are also risks to the privacy of a group *qua* group because revealing the information about a group increases the risk of discrimination against that group.

To address the issues, we need to rethink and expand our current moral vocabulary and legal frameworks for dealing with information technology. Broadening the scope of information protected by the right to privacy and data protection to include information primarily attributed to a person because of their membership in a specific group is one way to address the shortcomings of current privacy conceptions in relation to big data analytics (for more information, see Vedder's (1999) definition of categorical privacy). Furthermore, Henschke (2017) and Kammourieh et al. (2017) propose the protection of metadata, the valuable information that can be inferred from datasets, rather than raw data, as a way to address privacy issues.

To protect groups as such, I agree with Floridi's (2014) argument that measures must be taken that are not reduced to measures taken at the individual level. I also agree with Mantelero's (2017) idea that group privacy is required to limit the potential harm that can result from invasive and discriminatory data processing. However, unlike Floridi and Mantelero, I argued that a group right cannot be recognised as a means to protect a clustered group against discrimination through protecting the privacy of the group because a group right to privacy cannot theoretically be ascribed to a clustered group using traditional approaches. Moreover, the group right to privacy cannot be practically constructed by members of the group using society-based constructivism. In this regard, I recommend taking other approaches or moral

principles rather than the predominant standard approaches to protect a clustered group against discrimination.

Part III

Chapter 6:

Design for Embedding the Value of Privacy in Personal Information Management Systems

Abstract

Personal Information Management Systems (PIMS) aim to facilitate the sharing of personal information and protect privacy. Efforts to enhance privacy management, aligned with established privacy policies, have led to guidelines for integrating transparent notices and meaningful choices within these systems. Although discussions have revolved around the design of privacy-friendly systems that comply with legal requirements, there has been relatively limited philosophical discourse on incorporating the value of privacy into these systems. Exploring the connection between privacy and personal autonomy illuminates the instrumental value of privacy in enabling individuals to live their lives autonomously, highlighting the importance of intentionally embedding the value of privacy into these systems. To translate the value of privacy into concrete design requirements, this study constructs a value hierarchy consisting of values, norms, and design requirements. After analysing the relationships between privacy and autonomy and identifying norms, the design requirements translated from the norms associated with the components of personal autonomy are specified at the lowest layer. These requirements include a design to prevent unauthorised access and dark patterns and to provide effective and efficient notices and choices. The findings contribute to expanding the requirements for designing the aspect of privacy as a legal requirement to incorporate the value of privacy into systems.

Keywords: personal autonomy; personal information management systems; value of privacy; values hierarchy

This chapter is a modified version of the following publication:

Asgarina, H. (2024). Design for Embedding the Value of Privacy in Personal Information Management Systems. *Journal of Ethics and Emerging Technologies*, 33(1), Article 1. <https://doi.org/10.55613/jeet.v33i1.129>

6.1. Introduction

For the development, deployment, and use of artificial intelligence (AI) systems based on machine-learning (ML), a large quantity of data are collected to train ML algorithms and develop a model capable of processing new, untrained data (Al-Rubaie & Chang, 2019). Machine-learning models are deployed and used in diverse fields, such as healthcare (Freeman et al., 2021; Klein et al., 2021), urban informatics, meteorology, and crime prevention (Chen et al., 2014). Although the required information for the development, deployment, and use of ML models is diverse, such as sourcing data from the physical environment (van der Burg et al., 2021), this chapter specifically focuses on personal information based on the following definition by Henschke (2017, p. 186): ordered and meaningful data about a person. Personal information is the foundation for constructing datasets to train ML algorithms, which is crucial for developing valid models. Furthermore, personal information plays a pivotal role in the deployment and use of ML models. When deploying ML models to make predictions from new, untrained data, it is necessary to provide personal information as input to the model. For example, to build an ML model to detect breast cancer (Mazo et al., 2020), mammography images of a patient need to be sent to the model as input to provide information about the likelihood of the patient developing breast cancer. For the comprehensive development, deployment, and use of AI systems relying on ML, the collection and use of personal information are indispensable.

Those engaged in developing an ML model, such as engineers, and those involved in deploying an ML model, such as practitioners, require personal information to be shared with them. To enable and ensure individuals control the sharing of their personal information, technologies can be used to mediate the relationship between them and developers or deployers, with PIMS being a notable example (Asgarinia et al., 2023). Hence, PIMS enables individuals to manage the sharing of their personal data, which is essential for the development and functioning of ML models. The term ‘PIMS’ broadly represents a category of technology that enables individuals to decide what information about them is collected, when it is collected, how it is collected, and with whom it is shared. Personal data stores, personal data vaults, personal information management services, and personal data spaces all fall under the umbrella term of PIMS (Janssen & Singh, 2022). More recently, improved versions of PIMS, such as self-sovereign identity models, have been developed to enable individuals to mediate, monitor, and exert control over the access, usage, and sharing of their personal data (Asgarinia et al., 2023).

The PIMS approach promotes privacy self-management (Janssen & Singh, 2022). The objective is to make the processing of personal data transparent and to enable individuals (i.e., data subjects) to make decisions about their data. Two pivotal elements of privacy self-management are providing individuals with information about the data collected about them and how they are used (notice), as well as affording them the authority to decide whether they accept such data collection and usage (choice). This approach is commonly referred to as ‘notice and choice’ (Barocas & Nissenbaum, 2009; Solove, 2013). It involves providing transparent notice to individuals and enabling individuals to manage their own privacy preferences and interests (Solove, 2013).

There have been proposals to enhance the transparency of privacy notices, both in terms of content and the design of user interfaces. Transparency regarding content is commonly understood as a form of meaningful notice about the collection and usage of data (Barocas & Nissenbaum, 2009). Measures to enhance the transparency of privacy policy notices include presenting information about data usage in an understandable way. This approach can be achieved, for example, by shortening and simplifying privacy policies, as suggested by Calo (2012). Furthermore, beyond content and readability, to enhance the transparency of privacy notices regarding the design of user interfaces, measures have been introduced by Waldman (2018) to emphasise the design and aesthetics of content. These measures include elements such as font, size, colour backgrounds, and the use of charts or icons within notices, all aimed at effectively conveying information to individuals (Waldman, 2018).

To provide individuals with a real choice, one effective tactic is to employ an opt-in (instead of opt-out) mechanism as the default choice, ensuring individuals have the opportunity to decide how their data are used. In this regard, pre-checked checkboxes should not be equated with individuals having a real choice, and their use should be limited (Hoepman, 2022); individuals should have the opportunity to decide which data are collected about them, how they are used, and with whom they are shared. In addition, to enable individuals to control their personal information, dynamic choices using dynamic consent strategies (Budin-Ljøsne et al., 2017) should be provided, instead of a one-time choice at the beginning of personal information collection.

In addition to informing individuals and enhancing their ability to manage their privacy, the practical implications of privacy self-management involve imposing demands on entities or companies engaged in information collection or use to disclose their collection and usage practices explicitly and to demonstrate compliance with legal and regulatory requirements. The process of developing privacy notices triggers

internal adjustments within a company, fostering awareness regarding data collection and use practices (Barocas & Nissenbaum, 2009; Solove, 2013).

The use of PIMS as a means for sharing personal information and relying on these notice and choice approaches is effective for specific types of information—those related to or about an isolated person, rather than of a collective or group nature. As highlighted by Widdows (2013), genetic information makes individual choices inadequate because a person can only choose their own desires, preferences, and interests, whereas genetic information is interconnected with others. Regarding sharing information of a group nature, such as genetic information, alternative approaches or schemes become necessary. In these cases, data-sharing pools or data-cooperative schemes, rather than personal data sovereignty or PIMS, may work better. These approaches enable individuals to unite their data and collectively oversee their information (Asgarinia et al., 2023). Since this chapter focuses on PIMS and to avoid encountering the system's limitations, the type of personal information discussed in this research does not have a group nature.

Despite discussions regarding preserving privacy by implementing proper notice and choice, especially concerning privacy policies (Grannis, 2015; Waldman, 2016), there has been limited discourse on embedding the value of privacy into PIMS. Conducting philosophical investigations to integrate privacy into PIMS reveals the shortcomings of approaches that predominantly rely on notice and choice and suggests a more comprehensive approach for embedding privacy into the system. This chapter aims to address these shortcomings by proposing design requirements to incorporate the value of privacy into PIMS thoroughly. In this way, PIMS contributes to the value of privacy, which designers and developers intentionally embed in the technology.

The purpose of this chapter is to incorporate the value of privacy into PIMS; to do so, I draw a value hierarchy to translate the value of privacy into design requirements. As van de Poel (2013) explains, a value hierarchy consists of three layers of values, norms, and design requirements, in which higher-level elements are translated into lower-level ones. In this way, moving from the top layer to the bottom, abstract values are translated into tangible design requirements. In the values hierarchy this chapter proposes, the instrumental value of privacy is described in connection with a person's autonomy. From this perspective, specific design requirements are derived by translating norms that are aimed at promoting autonomy.

6.1.1. Conceptualising a Value Hierarchy for Privacy

Several approaches have been proposed in recent years to embed desired values into technology (Donia & Shaw, 2021; Iversen et al., 2012; Knobel & Bowker, 2011; Spiekermann & Winkler, 2020). The embedded values approaches claim that there can and should be an ethics of technology that is separate from the ethics of using it. These approaches hold that artefacts and technologies are never morally neutral. It is possible to identify tendencies in technologies to promote or de-promote the realisation of particular moral values and norms. Such tendencies are called embedded, embodied, or built-in moral values or norms. This implies that both the design and usage of technologies can have moral consequences (Brey, 2010).

Value Sensitive Design (VSD) is one of the most comprehensive, impactful embedded-value approaches (Friedman et al., 2008). The goal of VSD is to consider and incorporate moral values comprehensively throughout the design process. The approach provides guidelines for designing and developing technological products that promote the values desired by the various stakeholders whom these technologies may impact (Brey, 2010; van de Poel, 2009).⁴⁰

An essential stage in VSD is translating values into tangible design requirements. To do so, van de Poel (2013) introduces the notion of the value hierarchy, according to which values and design requirements have a hierarchical structure. The top layer of a value hierarchy consists of values; the intermediate layer consists of norms; and the most concrete layer involves design requirements. As van de Poel suggests, by moving from the upper to lower layers in a hierarchy, we can effectively translate abstract values into concrete design requirements.

Following van de Poel (2009, 2013), to construct a value hierarchy, leading to the intentional design of PIMS for the value of privacy, the following steps are essential: first, conceptualise how the value of privacy is understood or conceptualise the understanding of the value of privacy; second, translate the value of privacy into general and specific norms; and third, formulate design requirements through the translation of norms.

⁴⁰ For more information on the recent approach of Ethics by Design for AI (EbD-AI), which aims to embed ethical considerations in the design and development of AI systems and specifies actions that need to be taken at different design phases of AI development, see Brey and Dainow (2023). To explore what values are, the conceptualisation and specification of values, and issues related to conflicting values, see van de Poel (2021a).

Regarding conceptualising the value of privacy dedicated to the first layer of a value hierarchy, this chapter discusses the value of privacy in connection with a person's autonomy, in which privacy is considered valuable for the sake of autonomy. Since privacy is understood as a means to an end for autonomy, following the principle of the instrumental reason that instructs us to take the means to our ends (Korsgaard, 2009), if we want to live our lives autonomously, we must desire privacy to achieve that end. Simultaneously, the instrumental value of privacy depends on autonomy; the value of privacy is realised in metaphorical or symbolic spaces in which a person can develop and exercise their autonomy, enabling them to live their lives autonomously (Rössler, 2005).

Given the relationship between the instrumental value of privacy and autonomy, it is important to analyse the concept of autonomy. Drawing on different theories on autonomy, including moral autonomy (Korsgaard, 2009), relational autonomy (Christman, 2004; Mackenzie, 2008; Oshana, 1998), and personal autonomy (Rössler, 2005), I adopt Rössler's account of personal autonomy as developed in her book *The Value of Privacy* (2005). According to Rössler, an autonomous person is one who asks themselves practical questions and attempts to live accordingly. Practical questions include which of one's conflicting desires or convictions one wants to identify with, how to assess specific desires or preferences in their genesis, and what fundamental life projects play a part in assessing this identification. Personal autonomy thus consists of three necessary and sufficient components: authenticity and identification, the genesis of desires, and goals and projects (Rössler, 2005, p. 65).

I choose Rössler's theory over theories focusing on moral autonomy for two reasons. First, personal autonomy is not exclusively bound by a strong rationality; and second, unlike moral autonomy, personal autonomy incorporates personal components (see Sections 6.2.1.1 and 6.2.1.3). These two features imply that a person managing their privacy needs to have their own good reasons for doing so in specific ways, thereby being the author of their actions, regardless of whether others accept these reasons. Additionally, a person's desire, as a personal element, is required to manage their own privacy and to determine the flow of information.

Moreover, I adopt Rössler's (2005) conception of personal autonomy because social components are integrated into the concept. Although Rössler acknowledges social conditions in her account, she does not explore them in detail. Therefore, while taking Rössler's account as a basis for analysing autonomy, I modify it based on relational autonomy theories. In this conceptualisation, autonomy includes both primarily personal and thoroughly relational elements; one's autonomy is grounded

in certain affective attitudes towards oneself, constituted by intersubjective relationships and social conditions (Mackenzie, 2008; see Sections 6.2.1.2 and 6.2.2). The improvement that has occurred in the concept of autonomy, shifting from its association with isolated individuals detached from society to relational autonomy embedded in social conditions, has resonated within privacy discourses. As argued by Rössler and Mokrosinska (2013), privacy extends beyond individual control by emphasising the significance of intersubjective elements, particularly the purposes of relationships (see Section 6.3.1).

Concerning the second layer of a hierarchy dedicated to norms, in addition to the norm of reflection, I specify the norms associated with each component of the concept of personal autonomy. These components include authentication and identification, the genesis of desires, and goals and projects. The norms pertaining to the first component include exercising control over personal information to establish and maintain various social relationships; being aware of the types of relationships they are involved in, which helps them decide which part of their information to share; and considering social circumstances that provide a basis for recognition (see Section 6.3.1). The norms associated with the second component involve enabling a person to exercise control over their personal information to become less susceptible to manipulation and prevent manipulation to enable them to share their personal information as intended (see Section 6.3.2). Regarding the norms linked to the third component, they encompass the ability to contemplate and evaluate different alternatives for sharing information, ultimately choosing the one that aligns with one's objectives (see Section 6.3.3).

Concerning the third layer of a hierarchy centred on design requirements, I suggest the following design requirements regarding the value of privacy in PIMS: design for reflection through using friction, which obstructs a person in the completion of tasks typically performed without conscious thought, to stimulate imagination (translated for the norm of reflection; see Section 6.4); design to restrict unauthorised access by implementing encryption, considering the execution of contracts, and ultimately, employing blockchain technology to fulfil contract needs and apply encryption (translated from the first component of the concept of personal autonomy; see Section 6.4.1); design for effective notices, and design against dark patterns to prevent certain cognitive biases occurring (translated from the second component of the concept of autonomy; see Section 6.4.2); and design for effective and efficient notice and choice (translated from the third component of the concept of personal autonomy; see Section 6.4.3). As the proposed design requirements suggest,

embedding the value of privacy into PIMS involves more than just designing for notice and choice, as privacy policies emphasise. Additional requirements must be articulated and considered in the design of PIMS.

The findings of this chapter highlight that, although PIMS is primarily designed to protect privacy using the notice and choice approach that privacy policies regulate—design for meaningful notice and transparent choice—this approach must be completed by incorporating other elements, such as inclusiveness for diverse audiences (see Section 6.4.3). Furthermore, this chapter emphasises that the current privacy design in PIMS does not fully promote the realisation of the instrumental value of privacy, as it mainly addresses one component related to this value (i.e., goals and projects; see Section 6.4.3). However, other components, such as authenticity and identification, and the genesis of desires (see Sections 6.4.1 and 6.4.2), also require consideration in the design of PIMS. Therefore, the approach governed by privacy policies must be completed and also expanded to thoroughly incorporate the instrumental value of privacy. The main aim of this chapter is to conduct philosophical investigations that articulate design requirements for embedding the value of privacy into PIMS.

In the following sections, three parts are presented, each dedicated to a layer in a value hierarchy, namely values, norms, and design requirements. These layers are described in Sections 6.2 to 6.4, respectively. In addition to the guidelines and strategies developed to facilitate the design of privacy-friendly systems to ensure compliance with legal requirements and privacy policies, the proposed design requirements ensure that PIMS is built to realise the value of privacy.

6.2. The Layer of Values: Privacy and Autonomy

Following van de Poel's (2009, 2013) approach, the top layer of a hierarchy includes values. Since this chapter proposes a hierarchy for privacy, the top layer focuses on privacy. Therefore, to construct a hierarchy that facilitates proposing design requirements to incorporate the value of privacy into PIMS, the first step is to understand the value of privacy.⁴¹

⁴¹ In my view, various approaches describe the value of privacy, and one common approach distinguishes between intrinsic and instrumental value, for example, Rössler (2005). A second approach, influenced by Kagan (1998), suggests an object can have intrinsic value based on its instrumental value. A third approach, based on Korsgaard's (1983) categorisation of goodness, argues

A few debates in the literature on privacy focus on the idea that privacy is intrinsically valuable. Scanlon's (1975) thesis underscores the idea that we value a conventionally⁴² defined zone of privacy because, within such a realm, we need not be constantly vigilant against potential observations. Various social norms can delineate the boundaries of our zone of privacy in diverse ways. What matters most is that, within these realms or dimensions, we can engage in our activities without the need to be constantly alert for potential observers or listeners (Scanlon, 1975). A well-defined zone can be valued primarily for its own sake.

However, it has commonly been assumed by privacy scholars that privacy is valuable for the sake of something else, deriving its worth from other sets of moral values, principles, or commitments. Although the instrumental value of privacy has been discussed from different perspectives, from its relationship to social cohesion (Solove, 2008) to political values, such as power (Véliz, 2021) and democracy (Henschke, 2021), I focus on the value of privacy in relation to personal autonomy, as PIMS is developed with the primary aim of promoting one's autonomy.

The claim that the private realms or dimensions are valuable for their own sake does not in itself mean we do not also value them for the functions they fulfil. Although Scanlon's (1975) theory responds to the questions of why we value privacy and what would be lost if we were to lose it, functional theories can provide a deeper explanation. Most definitions and explanations of privacy in the literature are functional or, at the very least, can be interpreted as such. These functional approaches differ fundamentally from one another, with each one valuing privacy based on a distinct function it is intended to realise (Rössler, 2005).

In the scholarly literature, different theories have been developed to explain the value of privacy in relation to autonomy, e.g., Goffman (1959), Riesman (1952), and Rössler (2005). Goffman (1959) argues that privacy should be understood as a form of autonomy. According to Riesman (1952), the value of privacy stems from its

that something that is conditionally valuable (extrinsically good) when meeting certain conditions might be valued as an end instead of merely having instrumental value. Applying Korsgaard's categorisation results in describing privacy as having mixed value: it is extrinsically (conditionally) valuable as an end—for its own sake—as well as being extrinsically (conditionally) valuable as a means—for the sake of something else. Although the approaches inspired by Kagan and Korsgaard offer valuable discussions, this chapter adopts the more commonly employed approach of distinguishing between intrinsic and instrumental value.

⁴² Scanlon emphasises the conventional nature and, to a certain extent, the arbitrariness of the boundaries. For Scanlon, the mere existence of such boundaries, however varied they may be, is an indication that the value of privacy should be considered intrinsic (Rössler, 2005).

protection of individuals' autonomy, as privacy preserves a space around individuals, within which they can direct their lives and behaviour irrespective of social pressures. Rössler (2005) highlights that, in liberal societies, privacy is functionally valuable for the sake of a person's autonomy, of living autonomously. In short, we value privacy because we want to be autonomous, and without privacy, autonomy cannot work.

6.2.1. Privacy and Autonomy

Regarding the above discussions about the instrumental value of privacy, privacy is deemed valuable for the sake of autonomy. The conception I adopt in this chapter is based on Rössler's (2005) account of autonomy. Rössler argues autonomy is not connected to the strong criterion of rationality, unlike moral autonomy, and she considers social conditions' role in forming autonomy, such as relational autonomy. First, Rössler delineates between moral autonomy and personal autonomy, with a particular emphasis on the latter in terms of the functional value of privacy in furthering it. Each facet of autonomy necessitates a nuanced consideration of the distinct principles or conditions that underline it. Although Rössler does not explain the similarities and differences between moral and personal autonomy, I begin by briefly discussing moral autonomy and how it is often used following Kant. Second, Rössler identifies three sufficient and necessary components of personal autonomy and conceptualises personal autonomy in a way that depends on and is bound up with an intersubjective network. Therefore, before exploring the components of autonomy in Rössler's view, I discuss relational autonomy as the concept of personal autonomy with reference to intersubjective relations. Hence, in what follows, I provide an analysis of moral autonomy and discuss relational autonomy to facilitate an understanding of personal autonomy.

6.2.1.1. Moral Autonomy

As Korsgaard (2009) highlights, according to Kant, being autonomous means being governed by the principles of one's own causality—one's own maxims. The categorical imperative is a rule for constructing maxims (Korsgaard, 2009, p. 81). In *Groundwork*, the first formulation of the categorical imperative (i.e., formula of universality) is that you should act only according to that maxim, through which you can, at the same time, will that it can become a universal law (4:421). According to Kant, acting autonomously entails ensuring the maxim guiding one's action is one you could will to be a universal law. Autonomy identifies with the universalizability of one's own

maxims. Hence, to be autonomous means to act in conformity with the principle of *practical reason* (the categorical imperative; Korsgaard, 2009, pp. 71–80).

According to Korsgaard (2009), the reasons embodied in universal maxims must be understood as public or shareable:⁴³ reasons with normative force for all rational beings. Instead of merely thinking that, if I have a reason to do action-A in circumstances-C, then I must be able to grant that you also would have a reason to do the same (which relates to the universalizability requirement regarding the private conception of reasons), the public conception of reason indicates that universalizability commits me to the view that, if I have a reason to do action-A in circumstances-C, I must be able to will that you should do the same, because your reasons are normative for me. It is only regarding the public conception of reasons that a universalizability requirement leads us into moral territory—conformity with Kant’s law of humanity; by adopting other’s reasons as our own, with normative force for us, we treat them as an end in themselves (Korsgaard, 2009, pp. 191–192).

For an appropriate response to one’s incentives and to determine what actions are worth doing, a person must place themselves in the space between their incentives and their responses and reflect upon the principles that guide them in determining what an appropriate response is and what the situation demands. Korsgaard calls this space ‘reflective distance’ (Korsgaard, 2009, p. 116). The guiding principles must conform with the principle of practical reason, understood as public reason. Putting oneself in reflective distance that enables a person to exercise their moral autonomy.

Two features characterise Kant’s conception of moral autonomy, as adopted by Korsgaard (2009): first, rationality, which involves acting in a way that conforms with the principle of practical reason (i.e., the categorical imperative), understood as public and sharable reason; and second, focusing on the form of the maxim that must serve as a law without investigating the subjective source and the content of the maxims.

⁴³ There are other approaches to public reason. In Rawls’s doctrine of public reason to achieve justice, for example, the content and form of public reason are based on the notion of original position and the idea of reflective equilibrium (Rawls, 1997). The original position yields universally acceptable principles. Reflective equilibrium is a process in which each agent reflects on their considered beliefs while also considering the beliefs of others. Rawls moved moral reflection from the first-person perspective to public deliberation (Bird-Pollan, 2009). As Brandstedt and Brännmark (2020) highlight, reflective equilibrium serves as a valuable tool for public reasoning about practical problems, aiming to facilitate shared solutions. Although the relationship between Rawls’s and Korsgaard’s approaches to public reason merits exploration, it is beyond the scope of this chapter.

6.2.1.2. Relational Autonomy

Relational autonomy does not refer to a single account but to accounts shared under the assumption that ‘persons are socially embedded and that agents’ identities are formed within the context of social relationships’ (Mackenzie & Stoljar, 2000, p. 4). Thus, the focus of relational approaches is to emphasise certain social circumstances allowing a person to develop their autonomy (Oshana, 1998), interpersonal and social factors as conceptually necessary for autonomy (Christman, 2004), and social conditions necessary for the constitution of affective attitudes towards oneself (Mackenzie, 2008).

Marina Oshana defends and develops an influential account of social autonomy, emphasising that autonomy should be conceived as a ‘socio-relational’ phenomenon (Oshana, 1998, p. 94). In her account, it is social conditions that enable a person to self-determine that mark autonomy; autonomy is obtained only when the social conditions surrounding an individual meet certain standards. In cases in which basic opportunities for self-determination are denied due to strict obedience or subservience, such as in cases of voluntary slavery, the subservient woman, the conscientious objector, or the monk, then even if a person meets the condition of authenticity and chooses to enter or continue in certain conditions, the surrounding social conditions in which they reside do not allow them to be autonomous. According to Oshana (1998), a person who resides under oppressive social conditions cannot be autonomous.

Christman (2004) critiques Oshana’s view, arguing that, insofar as a person authentically embraces even an oppressive social status or subservient roles, they can still be considered autonomous. In Christman’s account, for a person to be autonomous, they must adequately reflect on their social conditions, including conditions of strict obedience. Rather than defending autonomy in idealised situations, breaking away from social norms that have influence over them, and pursuing their goals differently from those norms, as defended by Oshana’s account of autonomy, Christman states a person who can reflect adequately—in the sense that they can imagine choosing otherwise to value that alternative position—is autonomous. A slave, according to Christman, can consider themselves autonomous when they can see themselves doing otherwise, under at least some imaginable conditions, without needing to reject those conditions (Christman, 2004).

In contrast to Oshana (1998), Christman (2004) contends his view is consistent with the idea that selves are constituted by the social and interpersonal dynamics that surround them. In Christman’s view, insofar as the self is socially constituted, it is

counterintuitive to claim that such a self is only autonomous when they can break away from those very social conditions that constitute its being. As long as a person maintains the ability to reflect adequately on those conditions and embraces them, Christman argues we should continue to label them as autonomous (Christman, 2004).

Mackenzie (2008) critiques Christman's (2004) view by arguing that oppressive social conditions might undermine autonomy. According to Mackenzie, a person's practical identity may be shaped by false norms, beliefs, and distorted values arising from oppressive social conditions. This situation can lead to cultivating destructive affective attitudes towards oneself, such as a lack of self-respect or mistrusting one's own judgement.

Mackenzie (2008) advances a concept of relational autonomy that can be characterised as weak compared with Oshana's strong account, which defends a strong account of relational autonomy in the sense that abusive or oppressive social relationships necessarily undermine autonomy. However, Mackenzie (2008) argues that such conditions impair autonomy only when they fail to provide individuals with the recognitive basis necessary to maintain certain attitudes towards themselves.

Mackenzie (2008) develops a recognition-based account of relational autonomy. According to her, an autonomous person must be able to reflect on certain attitudes towards themselves, attitudes constituted by society and in intersubjective relationships. Drawing on the insights of Benson (1994) and McLeod (2002), Mackenzie highlights particular attitudes towards oneself as attitudes of self-respect, self-trust, and self-esteem. These affective attitudes are constituted by society and in intersubjective relationships. Thus, practical identity is first-person identity aligned with Henschke's (2017) idea of self-regarding identity, which is constituted thoroughly in intersubjective relationships and depends upon the mutual recognition in socio-relational situations.

Relational theorists have rejected the individualistic conception of autonomy that typically tends to understand practical identity as being formed by one's own desires, values, and commitments independently of social influence. Instead, these theorists have argued that practical identity is shaped by the social and interpersonal aspects of one's life. Oshana (1998) advocates a strong condition for autonomy, arguing that, to exercise autonomy properly, a person must reject abusive and oppressive social relations that contribute to the formation of their practical identity. Christman (2004) defends the autonomy of those who, for religious or ideological reasons, authentically embrace subservient relationships. Mackenzie (2008) defends weak relational

autonomy, in contrast to Oshana, arguing that, to exercise autonomy, the social environment should facilitate intersubjective recognition, which constitutes an affective attitude towards oneself.

I adopt Mackenzie's (2008) account of autonomy because I believe that those social relationships that do not provide a person with the recognitive basis necessary to sustain their affective attitudes towards themselves are inimical to autonomy, rather than advocating the strong idea that oppressive social conditions undermine autonomy or even the idea that a person is autonomous insofar as they adequately reflect on social conditions and embrace them.

6.2.1.3. Personal Autonomy

Instead of discussing moral autonomy, Rössler (2005) focuses on personal autonomy, referring to it as general personal self-determination concerning how a person wants to lead their life. A person must be able to ask themselves practical questions about how they want to live, what sort of person they want to understand themselves as, and what kind of life is good for them. Additionally, a person must be able to make decisions from this perspective and live in accordance with such decisions. Instead of reflecting on the reasons for actions, a person must reflect on their own life. To ask oneself practical questions and to live accordingly is to be autonomous.

Three features characterise Rössler's (2005) theory. First, unlike moral autonomy, personal autonomy is not exclusively bound to a strong notion of rationality. A person is autonomous in the sense of having their own good reasons if they can understand themselves as the author of an action. However, this point need not simultaneously mean that other people accept these reasons, nor does it imply that reasons must be public or shareable. A person's choice or action incorporates personal feelings, obligations, memories, and biographical influences that may not appear equally sensible or convincing to everyone.

Second, regarding personal conditions, Rössler's (2005) conception of autonomy consists of the genesis of desires, goals, and projects (see Section 6.2.2). As highlighted by Williams (1976), Kantian moral philosophy focuses on principles that apply universally, regardless of personal desires or the particular circumstances in which a person is situated. However, it is important to recognise that different people have different sets of desires, concerns, or projects for living their lives. It is not through having one's project affirmed by anyone that the person will have earned their place in the world; rather, a person will have made a distinctive contribution to the world if their distinctive project is realised (Williams, 1976). Williams's objection to Kantian

moral autonomy does not encompass personal autonomy, as the personal autonomy presented by Rössler (2005) incorporates personal elements within itself.

Third, Rössler (2005) also emphasises that intersubjectivity is generally intrinsic to the process of autonomy in various respects, concerning both the genesis of autonomy and the question of what aims and projects a person wants and is able to pursue. In this regard, Rössler extends the concept of personal autonomy to include relational autonomy.

6.2.2. Three Components of Personal Autonomy

Rössler's (2005) analysis focuses on the necessary and sufficient components for the concept of autonomy. Rössler posits that an autonomous person is one who asks oneself the practical question, which involves considering how one is to behave in certain situations given certain desires, one's own history, and one's convictions. This approach means asking oneself which desires or convictions one wants to identify with, how to assess specific desires or preferences in their genesis, and which fundamental life projects are involved in the evaluation and appraisal of this identification. The concept of personal autonomy comprises three components: authenticity and identification, the history and genesis of desires, and goals and projects. I adopt these three necessary and sufficient components of the concept of autonomy proposed by Rössler (2005) as a basis for analysing autonomy, and I refine them by considering relational autonomy.

Using van de Poel's (2009, 2013) methodology, the analysis reveals the link between autonomy and privacy is in the top layer of a value hierarchy. The connection between these values becomes more detailed by analysing the components of autonomy, a task undertaken in the following section. Section 6.3 discusses how privacy connects to autonomy by identifying privacy norms that should be met to promote autonomy.

6.2.2.1. Authenticity and Identification

The first component of the concept of autonomy is authenticity and identification. The authenticity condition specifies that, for a person to be autonomous, their beliefs, desires, value commitments, decisions, or actions must be authentically theirs, in the sense that they can identify with their desires, goals, and values as their own (Henschke, 2017; Rössler, 2005; Williams, 1976).

A person is autonomous if their desires and subsequent actions are their own, meaning they are authentically theirs and do not feel alienated from them (Mackenzie, 2008). To achieve this, a person must reflect on their desires, motivations, and values. Hence, a person must be able to, and in a position to, reflect on certain desires and, based on such reflections, accept, reject, or modify them. Therefore, the act of identification must be understood as evaluative (Rössler, 2005). There are different ways to establish the connection between autonomy and an agent's identity or evaluative first-person perspective, for example, the reflective endorsement of one's desires and values (Korsgaard, 1996) and identification with one's will (Frankfurt, 1971).

Reflective endorsement is motivated by the underlying intuition about the centrality of the first-person perspective. According to Korsgaard's (1996) account of reflective endorsement, when an agent reflects on the things to which they are inclined, they can either endorse or reject the authority of those inclinations and act accordingly. To determine what incentives, motives, or inclinations will count as reasons for us, and thereby endorse them, Korsgaard (2009) argues that principles, specifically rational principles, are required. This argument implies that the reasons should be endorsed by everyone, not just by the person who considers the reasons for their own actions or decisions (see Section 6.2.1.1). Korsgaard claims that reason is born in the space of reflective distance. It is from the reasons we legislate, and our actions and choices, that we construct our practical identities as individual human agents (Korsgaard, 2009). Thus, the agent reflectively endorses their motives, incentives, values, and inclinations that guide their actions, which endorses their practical identity. Similarly, Henschke introduces the notion of 'Self-Regarding Identity; I am who I perceive myself to be', which arises from the endorsement and rejection of particular traits (Henschke, 2017, p. 163).

Frankfurt's (1971) analysis of the concept of a person identifies a distinguishing mark of the human condition as the capacity to form 'second-order volitions' or 'volitions of the second order', or as being able to take an evaluative attitude towards the desires that incline them to act. A person has second-order volitions when they want certain desires to be their will, to be identified with (Frankfurt, 1971, p. 10). There may also be a series of higher and higher-order desires and volitions, and a person might identify themselves with their desires higher than the second order. As Frankfurt (1971) admits, there is no theoretical limit to the length of the series of desires and volitions. However, he argues that when a person decisively identifies themselves with one of their desires, they decide to stop forming desires or volitions of

higher orders. Frankfurt clarifies the meaning of the verb ‘to decide’ in his paper titled ‘Identification and Wholeheartedness’ (1987) to address the problem of infinite regression. Frankfurt highlights that ‘to decide’ is synonymous with ‘to make up one’s mind’, which is like ‘creating an orderly arrangement’ in such a way that the sequence of evaluation can be terminated, and the person as a whole can identify wholeheartedly with the decision (Frankfurt, 1987, p. 41).

Following the above discussion, the endorsement of one’s desires or values or the alignment between lower-order and higher-order elements of one’s volitions secures authenticity. Despite some philosophical differences between these accounts, the element uniting them is reflection; they emphasise an agent’s reflection on their own desires or wills. However, the process of establishing authenticity and identification is not entirely free from social influences; it occurs within intersubjective relations. Nevertheless, this does not mean the social conception of self and the personal conception of autonomy are contradictory. Autonomy requires the internal integration of one’s self, and since the self is constituted by social factors, a person acquires autonomy by reflecting on aspects of their character defined in the external relations they have with others (Christman, 2004).

Following the recognition-based relational view of Mackenzie (2008), an agent’s autonomy depends on intersubjective relationships that provide a basis for one’s recognition. Similarly, to emphasise the social and intersubjective aspects of identity, Henschke introduced ‘Other-Regarding Identity’, which is ‘who X perceives Y to be’, in which Y is another person; that is, my identity is who you perceive me to be (Henschke, 2017, p. 166). A person is autonomous if they can, and are in a position to, reflect on a practical identity or self-conception underpinned by certain affective attitudes constituted by society and developed in intersubjective relationships (Mackenzie, 2008).

6.2.2.2. The Genesis of Desires

Authentic identification with a desire does not necessarily guarantee a person is genuinely autonomous, as the desire might be a product of manipulation. A person is autonomous regarding beliefs, desires, value commitments, or decisions only if, were they to reflect on the historical process of their formation, they would learn they are not products of manipulation. Considering this point, the historical component becomes integral to the conception of autonomy. Hence, both authenticity and identification and the genesis of desires are sufficient and necessary components of autonomy (Rössler, 2005).

Reflecting on the genesis of desires helps a person avoid self-deception and manipulation, enabling them to develop a non-manipulative relationship with themselves (Rössler, 2005). Moreover, concerning the intersubjective and social conditions under which autonomy is learned or acquired, reflecting on the formation of certain attitudes towards oneself is required to prevent a person becoming involved in oppressive and abusive interpersonal relationships. Although it might be demanding to be free from manipulative external circumstances in the strong sense defended by Oshana (1998), it is necessary in a weaker sense. A person must live in social conditions that do not deny them a form of recognition, as defended by Mackenzie (2008). Hence, reflection is necessary to prevent a person from engaging in social relationships that do not grant them appropriate recognition.

6.2.2.3. Goals and Projects

In addition to authenticity and identification and the genesis of desires, the third component of the conception of autonomy concerns a person's goals and projects. To be autonomous, a person must have the ability and be in a position to form goals and design projects and to pursue these in practice (Rössler, 2005).

Rössler's (2005) argument does not explicitly refer to the diachronic dimensions of autonomy; however, it involves conceiving of autonomy in reference to personal history and the genesis of desires, which has retrospective elements. The argument is also related to making plans about the component of goals and projects, which have prospective elements. Considering both retrospective and prospective elements endorse that autonomy (specifically the self-governing dimension of it)⁴⁴ is a diachronic, temporally extended process. This claim is defended by Bratman (2007), Christman (2009), and Mackenzie (2023).

As Mackenzie (2023) states, one way of conceptualising the diachronic dimension of autonomy is Bratman's (2007) planning account.⁴⁵ In that account, the temporally extended structure of autonomy is defined by a person perceiving their agency as extending both backward into the past and forward into the future. Considering the past through memories and envisioning the future through intentions and plans, a

⁴⁴ According to Mackenzie, the concept of autonomy comprises three distinct but interacting dimensions: self-determination, self-governance, and self-authorisation (Mackenzie, 2023, p. 375). To avoid complexity, I do not compare the three components of autonomy presented by Rössler (2005) with the three dimensions of autonomy presented by Mackenzie.

⁴⁵ Mackenzie (2023) argues for the alternative account known as a narrative constitution account.

person establishes connections that bind their present to both their past and future. Mackenzie emphasises that, according to the planning account, a person establishes these connections and organises their activities over time by forming intentions, planning the means to realise those intentions, and enacting prior intentions (Bratman, 2007; Mackenzie, 2023). Autonomy is shaped over time by a person's intentions and plans orientating their reflections. To be autonomous is to form intentions, make plans, and direct one's life in accordance with those plans.

The temporally extended dimension of autonomy is shaped in relation to contingencies, social relations, and the social environment (Mackenzie, 2023). The contributory role of social conditions in this component of autonomy can be explained in two ways. First, the goals, projects, and ways of life available to a person are determined by specific cultural assumptions and social contexts. An autonomous person can reflect on how they are situated in cultural, social, and intimate contexts and incorporate this reflection into forming part of their goals and projects (Rössler, 2005). Second, involvement with particular other people might be one of the kinds of projects that figure in a person's life. An autonomous person can develop and sustain relationships with others with whom certain affective attitudes are formed.

The functional value of privacy is realised when a person can live their life autonomously. The conception of autonomy is explained precisely by analysing the three necessary and sufficient components. Thus, the value of privacy is understood as a means to protect and promote the three components of autonomy, which is the aim of the top layer of a value hierarchy—to explore relationships between values. The intermediate layer explores the relationships between privacy and autonomy by identifying norms that promote autonomy.

6.3. The Layer of Norms: Norms for Promoting Personal Autonomy

Following van de Poel (2009, 2013), the intermediate layer of a hierarchy comprises norms translated from the upper layer of values. Based on the previous section, the functional value of privacy depends on autonomy. In this section, I go beyond the commonly recognised norm of reflection within the three components of autonomy—reflection on self-conception, reflection on the genesis of desires, and reflection on goals and projects. I specifically identify additional norms that must be met for an individual to be autonomous, focusing on autonomy's components to realise the functional value of privacy. By delineating these norms, I explore the connection between privacy and each component of autonomy in detail.

6.3.1. Authenticity and Identification

According to Mackenzie's (2008) view, a person must perceive themselves as the legitimate source of authority over their decisions and actions. This normative authority is grounded in one's attitudes towards themselves, which are intertwined with interpersonal relationships and the social structures of mutual recognition (Mackenzie, 2008). Consequently, promoting autonomy involves fostering the interpersonal and social conditions necessary for its development and exercise. Furthermore, to foster autonomy, social circumstances should provide a basis for recognition that enables a person to realise their autonomy.

To promote autonomy, regarding relational autonomy, which emphasises that autonomy is developed and sustained intersubjectively, a person must be able to situate themselves in a network of intersubjective relationships governed by various social norms, in which they see themselves in different roles, such as a friend, colleague, and wife. Moreover, as Mackenzie (2008) argues, for a person to be recognised within their social network, a series of interconnected obligations on the part of the social network must be fulfilled. These obligations include treating a person as someone with a conception of themselves and for whom certain things matter, as well as understanding the subjective perspective regarding one's situation.

To explore the relationships between privacy and autonomy, moving away from the conceiving of autonomy as detachment from social life to viewing it as socially embedded helps to explain privacy discourses. Scholars in privacy studies who recognise relational or social autonomy have argued that not only does privacy protect autonomy by preserving engagement in social interactions, but it also facilitates the social relationships required for a person to be able to exercise their autonomy (Rössler & Mokrosinska, 2013).

In philosophical literature, privacy is commonly defined as control over access to oneself or one's information (Rössler, 2005; Westin, 1967). A person who has control over their information can determine what they disclose to others and what they conceal from them. Given that relationships between people can be differentiated according to the degree of personal information they share, the ability of individuals to disclose and conceal information to and from others enables them to form various social relationships. Hence, privacy serves to regulate and facilitate the enactment of social relationships (Rössler, 2005; Rössler & Mokrosinska, 2013).

In accordance with Rachels's (1975) perspective, Rössler and Mokrosinska (2013) highlight that informational management within relationships comprises two aspects: a subjective aspect, linked to an individual's ability to control information, and an

intersubjective aspect imposed by the type of relationship. Intersubjectively shared standards grounded in the purpose of social relationships determine the relevance of information to those relationships. What others, such as students or bankers, know about me is largely determined by the specific kind of relationship I am engaged in and the roles assumed within that relationship. Therefore, privacy involves an individual's control over access to information, with the degree of control the individual possesses depending on the character of the social roles they perform and the nature of the social relations in which they participate (Rössler & Mokrosinska, 2013).

Privacy is understood as a means of promoting a person's autonomy by fostering various social relationships and cultivating the social conditions required for the development and exercise of autonomy. Therefore, norms aimed at promoting autonomy, considering the first component of autonomy (i.e., authenticity and identification), involve, first, that individuals should have control over others accessing their information to maintain different social relationships. This control enables individuals to decide whether to disclose some information to certain people or conceal it from others. Second, individuals should be aware of the type of relationships they will be involved in to decide which information to share. Third, as mentioned earlier in this section about the intersubjective aspect of privacy, social circumstances should provide a basis for one's recognition by others in intersubjective relationships.

6.3.2. The Genesis of Desires

The historical component of the conception of autonomy is necessary to prevent a person from falling into self-deception and manipulation, enabling them to escape the external circumstances that underpin destructive attitudes towards themselves. Unlike Oshana (1998), who maintains the strong view that a person must be free from manipulative external circumstances to be autonomous, and unlike Christman (2004), who argues that, in certain circumstances, we may accept a desire or approve certain ways of acting or behaving even once we understand they resulted from manipulation, Mackenzie's (2008) view posits that those social circumstances that erode one's normative authority over their decisions and actions compromise autonomy.

Manipulation involving personal information might occur in two cases. First, manipulation arises when personal information about an individual is used in a way that prompts that person to take a particular action. The case I am discussing is similar to those instances in which a company, having accumulated significant private data on a person, uses this information to manipulate them, for instance, through targeted

advertising. Although manipulation can also occur in cases in which a person is forced to do things they might not otherwise do, such as in blackmail cases, these cases differ from the ones to which I refer. The manipulation I discuss here occurs because of the detailed information others have obtained about a person, not because of disinformation.

Second, manipulation might arise from software and user interface designs that afford certain actions, particularly the act of sharing personal information. These designs can manipulate a person by triggering cognitive biases, leading them to divulge more information than they intend to. One strategy to manipulate a person in this way is to present the information—about what happens to shared data and who accesses them—in such a way, both in terms of content and design, that it prompts them to share personal information they would not otherwise disclose.

As highlighted by Nissenbaum (2010), the relationship between privacy and autonomy is not restricted to one's ability to reflect on principles of actions and having the freedom to act according to them. The relationship also involves one's ability to carry out those actions without being manipulated by others or circumstances, which can influence the shaping of one's choices and actions (Nissenbaum, 2010). In the first case, the manipulation that deprives one's autonomy occurs due to the absence (or invasion) of one's privacy. In this regard, privacy is required to mitigate the problem of manipulation. A person can only exercise control over their personal information when they know what is being done with their information,⁴⁶ meaning they will be less susceptible to such manipulation.

In the second case, the person is exposed to information that triggers a specific cognitive bias, known as the metacognitive decision-making process (see Section 6.4.2), manipulating them into sharing their personal information (Waldman, 2020). To prevent manipulation that erodes one's autonomy, measures should be taken to prevent system designers or developers from providing information that triggers cognitive biases. The content of the information presented in privacy notices and how it is presented are important (see Sections 6.1 and 6.4.3) for preventing manipulation and, ultimately, protecting one's autonomy.

Based on the two cases discussed above, two norms related to the historical component of autonomy regarding the mitigation or prevention of manipulation are derived. First, a person should be able to exercise control over their information to become less susceptible to manipulation. Second, the action of sharing personal

⁴⁶ For more information about the relationship between privacy and purpose, see Asgarinia (2023).

information should be guided by one's intention or authentic desires rather than being caused by (external) factors triggering certain cognitive bias; the action of sharing should be formed by this reflective process rather than being a mere reaction to the conditions prompting a person to share their information. Thus, privacy notices (see Sections 6.1 and 6.4.3) should prevent a person from being trapped by cognitive bias, thereby avoiding the unintended divulgence of their information.

6.3.3. Goals and Projects

The third component of the conception of autonomy, according to Rössler (2005), emphasises that, to be autonomous, a person must be able to form goals and design projects by considering the social context in which they are situated and including the development of relationships with others. Furthermore, being autonomous is not solely about the ability to form intellectual plans; rather, an autonomous person can pursue their goals and projects in practice as well.

Regarding Bratman's (2007) planning account, autonomy is developed over time through a temporally extended process that involves the formation of intentions, the planning of means to realise those intentions, and the enactment of prior intentions that guide deliberation. To guide deliberation effectively, intentions and plans must meet certain norms, specifically means-end coherence. Means-end coherence helps guide deliberation by concentrating one's planning activities. For example, if a person aims to achieve an end, such as improving their fitness, this requires them to figure out the best means of doing so, meaning it is necessary to develop plans and subplans (Mackenzie, 2023).

Regarding the sharing of personal information using PIMS, if a person intends to share such information as part of a plan, perhaps to enhance their health, they should assess whether PIMS is an effective means for achieving that end. PIMS, which enables a person to share their information, is an effective means to realise their end if the purpose for which the information is collected using the system aligns with the person's intended purpose. For a person to decide whether PIMS coheres with their goals, privacy notices embedded in the system should clearly state the purpose for which information is collected; for example, the collected information is fed as input into an ML model developed to detect a disease, helping individuals to decide about whether to use PIMS as a means to realise their goals (see Section 6.4.3).

Moreover, individuals should be able to assess and consider different alternatives for sharing information and choose the one that aligns with their objectives. Based on these alternatives, individuals can make meaningful decisions about sharing their

information for specific purposes. Thus, privacy choices (see Sections 6.1 and 6.4.3) should enable a person to pursue their own choices.

Section 6.3 explored the second layer of the hierarchy, focusing on the norms translated from the upper layer of values. This section specified the norms that must be met for a person to be autonomous based on the three components of autonomy.

6.4. The Layer of Design Requirements: Design for the Value of Privacy

Having outlined the norms in the second layer, the final step is to translate these norms into design requirements, comprising the lowest layer of the hierarchy. As I mentioned in Section 6.3, reflection is a common norm among the three components of autonomy, encompassing reflection on self-conception, reflection on the genesis of desire, and reflection on goals and projects. Therefore, a deliberate approach to designing for reflection is required to align with these overarching norms. Terpstra et al. (2019) suggest that *designing for reflection* can enable individuals to reflect.

As highlighted by Terpstra et al. (2019), reflection can be triggered by the introduction of friction.⁴⁷ These scholars emphasise that deliberately incorporating friction into a design enables individuals to escape the habits of thought and behaviour to reflect critically on their actions and decisions. Friction is commonly understood as anything that obstructs a user in the completion of the tasks they typically perform without conscious thought, thereby instigating reflective thinking. Designers can embed friction into their designs by pre-emptively discerning a user's habitual behaviour and devising strategies counter to it (Terpstra et al., 2019). Mackenzie (2000) highlights that representational or imagistic thinking is integral to the process of self-reflection and deliberation. Representational imagining can open up a space within which a person can step away from their habitual modes of understanding themselves and their relationships with others. Within this mental space, a person can explore different possibilities for themselves. At its core, the ability to imagine ourselves in different ways plays an important role in practical reflection and deliberation about the self (Mackenzie, 2000).

Regarding the above discussions, friction can be used to trigger imagination, enabling a person to reflect on their desires, actions, decisions, and what matters to them. Thus, friction, understood as affording reflection and imagination, is a way of promoting autonomy. For instance, friction can be achieved by asking specific

⁴⁷ The way Terpstra et al. (2019) refer to friction differs from 'ontological friction' defined by Floridi as the forces that oppose the information flow within the infosphere (Floridi, 2005, p. 186).

questions that prompt individuals to imagine themselves in certain situations in the present or even the future, while sharing their information with others. The inclusion of specific questions is a high-level requirement for incorporating friction (to trigger imagistic thinking) into design. The specific content of such questions, the ways they are presented, and the provision of explanations to users regarding why such questions are asked necessitate empirical investigation.

In addition to designing for reflection to meet the common norm in the three components of personal autonomy, the norms identified in Section 6.3 need to be translated into design requirements, which is the focus of the remaining sections.

6.4.1. Authenticity and Identification

The norms related to promoting autonomy concerning its first component (i.e., authenticity and identification) are that individuals should have control over access to their information; they should be aware of the types of relationships they are involved in to share their information accordingly; and the social network people participate should treat them as people for whom certain things matter, for example, which pieces of their personal information are shared with whom for which purpose. In this section, these norms are translated into design requirements.

Privacy regarding control over access to information protects against unwanted or unauthorised access to information (Rössler, 2005). Given that unauthorised access to one's information leads to a loss of control and infringement of one's privacy, the sufficient condition—though not necessarily a necessary one—for losing control is unauthorised access. A measure to ensure a person maintains control over their information involves restricting unauthorised access to that information. Encryption is a valuable measure for protecting the sharing of information from unauthorised access. Employing encryption to protect privacy was proposed by Miller and Bossomaier (2021). Encryption works by converting data into a code that can be deciphered only by individuals who possess the correct decryption key. When data are encrypted, even if they are intercepted by a third party, that party should not be able to understand or make use of them without the decryption key (Coron, 2006).

To fulfil the next two norms, which share the intersubjective element, I suggest using contracts, which identify the purpose of sharing information and the person or parties with whom that information will be shared, as well as providing instructions for caring for the data (Christidis & Devetsikiotis, 2016). The purpose of contracts in contexts in which information is shared is to record how parties care about shared personal data and to serve as a reference to guide parties' activities. The contract

emphasises factors such as how caring about shared data matters for the person who shared them and how receivers care for something senders (or data subjects) care about. These aspects allow for a distributed consensus on a transaction and the sharing of data, ultimately facilitating mutual recognition.

PIMS should implement measures to restrict unauthorised access by third parties seeking to process user data. This system ensures that, even if an unauthorised third party gains access to encrypted data, they cannot decipher them without the proper cryptography key. Moreover, PIMS should execute contracts (or smart contracts). One way of meeting the needs of contracts and utilising encrypted data is through the use of blockchain. Blockchain is a technology using cryptographic hash functions to store and distribute sensitive data (Hölbl et al., 2018; Khezr et al., 2019), and it has a feature that can execute smart contracts (Christidis & Devetsikiotis, 2016); therefore, PIMS should employ blockchain to fulfil the aforementioned norms.

6.4.2. The Genesis of Desires

The two norms described in Section 6.3.2 are associated with the historical component of the conception of autonomy. First, a person should be able to exercise control over their personal information to be less susceptible to manipulation. This norm necessitates that a person should know what is being done with their personal information. Second, the act of sharing information should be guided by one's intentions or authentic desires rather than being caused by or reacting to external factors. One way to realise this norm is to design privacy notices to prevent certain cognitive biases. These biases might otherwise manipulate individuals into unintentionally sharing their personal information. The design requirements derived from the control norm—which necessitates awareness of the purpose for which data are collected—are detailed in Section 6.4.3. This section outlines the design requirements associated with the second norm.

Platforms usually employ design tactics to manipulate users into disclosing more information than they initially plan to share, sustaining an information-driven business model.⁴⁸ This manipulation frequently arises from implementing what are commonly referred to as 'dark patterns' in platform design. Designers employ such patterns to coerce and deceive users into disclosure and to trigger cognitive biases that prompt users to divulge information they might otherwise withhold (Waldman, 2020).

⁴⁸ For more information on surveillance capitalism, see Zuboff (2019).

Using dark patterns can trigger a cognitive bias known as the metacognitive decision-making process (Waldman, 2020). This bias hinders individuals' abilities to make choices that align with their preferences. When individuals encounter challenging decisions, some interpret the complexity as an indication of its importance, motivating them to engage in thoughtful deliberation when making choices. However, when individuals view difficulty as a signal that the task is impossible, they tend to be more likely to give up on their choices. This second approach indicates that, when individuals face challenges in making choices about sharing their personal information due to complex notices, as many often do, they become more inclined to avoid limiting the disclosure of their information (Waldman, 2020). Designing to promote the metacognitive decision-making process can lead to the manipulation of users, prompting them to share more information than they desire or intend to.

To design PIMS to counter dark patterns, designers should present information using a plain and transparent design and language, encouraging users to make decisions by reading notices. To achieve this aim, designers must provide meaningful notices that are concise and easily understood by the majority of individuals, not just legal experts. Transparent design methods, such as using tables with appropriate fonts and colours (as explained in Section 6.1), can aid in this effort. Moreover, designers need to encourage users to manage their privacy by providing feedback that clearly illustrates how user choices impact the real world. As privacy choices are a process, the system must offer clear and timely feedback that reflects the most recent user actions, indicating that privacy settings have been modified in accordance with their latest choices (Feng et al., 2021).

6.4.3. Goals and Projects

At this stage, the specific norms concerning goals and projects should be translated into design requirements. These norms encompass two key aspects: first, individuals should be made aware of why their personal information is collected and shared with others; and second, individuals should be able to assess and consider different alternatives for sharing information, and based on this assessment, they should choose and pursue their goals. The effective (in the sense of the information provided to individuals about the data collected about them and their usage) design of the notice fulfils the first norm, and the efficient (in the sense of enabling individuals to manage their own privacy preferences and interests) design of the choice aids in realising the second norm.

The effective design of notices discourages users from habitually accepting the notices without considering their content, helping users pay attention to data practices. The content, presentation, inclusiveness for different audiences, and integration of notices into PIMS are all crucial factors for achieving effective notices. Regarding content, well-designed notices should notify individuals about the data practices of PIMS. This aspect includes specifying what data are being collected about individuals, for which purposes, with whom they are shared, why, and how long they are stored (Schaub et al., 2018). Furthermore, information should be presented in a manner that effectively and transparently communicates these data-collection and data-sharing purposes to individuals, helping them to decide about sharing their information. When the purpose is clearly stated in the notice, users are aware that, by sharing their information, they can achieve their desired outcome. Notices should also inform users about the options available to control or prevent certain data-sharing practices.

For the effective design of notices, the audience and how the notices are presented should be considered. Regarding the audience, effective notices need to consider a wide range. Notices are typically conveyed through text, images, or icons, and it is important to incorporate auditory methods to inform the visually impaired community. Notices are often presented separately and detached from the individual's interaction with the system, such as being placed at the bottom of a page. However, to maximise the effectiveness of a design, notices should be seamlessly integrated into PIMS, so individuals do not need to seek them out but encounter and engage with them as part of their interaction with the system and read them (Schaub et al., 2017).

In addition to the design of notices, the design of choices should be structured to provide individuals with control over certain aspects of data practices and accommodate diverse user preferences. Rather than presenting a binary choice design that restricts individuals' abilities to express their preferences, designers should employ multiple choices. For example, mobile platforms, such as Android and iOS, offer users various ways to decide how they want to allow apps to access the location data collected by their devices, including 'always', 'while using the app', 'never', and, more recently, iOS has introduced 'just once' (Feng et al., 2021). If the different options are clearly explained, they do not create a cognitive load for people to follow and understand what each option means, meaning it is less confusing for people to decide.

Individuals should also have the opportunity to negotiate their decisions and choices, refining how and why their shared information can be used, specifying who can receive the shared information, or adjusting the level of information they share

(Schaub et al., 2018). The goal should be to provide individuals with the means to pursue their specific aims instead of presenting them with a binary choice of take-it-or-leave-it.

Timing can significantly influence the effectiveness of a notice, impacting how individuals engage in decision-making. One potential relationship between notice and choice is integration, wherein the system communicates notices to individuals either simultaneously or sequentially. For instance, one option for the integrated approach is just-in-time notifications, in which privacy choices are presented to individuals when a particular data practice is about to occur, often integrated within relevant privacy notices. For example, when a mobile application seeks to access specific data, such as location on a mobile device for the first time, the mobile platform asks the user for permission through a pop-up dialogue box (Feng et al., 2021).

Notices and choices can follow dynamic consent approaches while employing strategies to avoid consent fatigue. Dynamic consent is a strategy developed to facilitate the consent process and the ongoing communication between researchers and research participants (Budin-Ljøsne et al., 2017). However, a high frequency of interaction with individuals, overburdening them with too much information each time, can be cognitively draining and counterproductive, leading to consent fatigue. To remedy this fatigue, four measures can be taken. First, prior to the collection of personal data, the lawful basis for processing must be identified by the entities or agents collecting the data. Second, consent from individuals is obtained in cases in which the processing of personal data is likely to pose risks to the rights and freedoms of individuals. Third, instead of requesting systematic (re-)consent, transparency should be improved. Finally, a privacy label should provide the required information in an easily understandable manner, making privacy notices easier to read (Montezuma & Taubman-Bassirian, 2019).

Section 6.4 discussed incorporating design for reflection into PIMS to realise the common norm of reflection associated with the three components of personal autonomy. To achieve the specific norms of the three components, consideration should be given to designing a system to prevent unauthorised access, to counter dark patterns, and to employ effective and efficient notices and choices in the design of PIMS. Furthermore, this section has demonstrated that the notice and choice design, translated from norms relevant to the component of goals and projects of autonomy, is more detailed than the one focused solely on privacy policies.

A summary of Sections 6.2–6.4, each dedicated to a different layer of a value hierarchy, is illustrated in Figure 6.1. The values are in dark grey in the top layer, and

the connection between the instrumental value of privacy and the three components of autonomy is depicted in this layer. The middle layer is dedicated to the norms associated with each component of autonomy, displayed in light grey. Since reflection is a common norm, it appears in each component. The bottom layer is dedicated to the design requirements (in white) derived from the translation of the norms. Although some design requirements are linked immediately to the upper level—for example, imagination can be triggered by friction—in some cases, there is no immediate link; for instance, the inclusiveness of the audience is connected to the effective notices placed two levels above it.

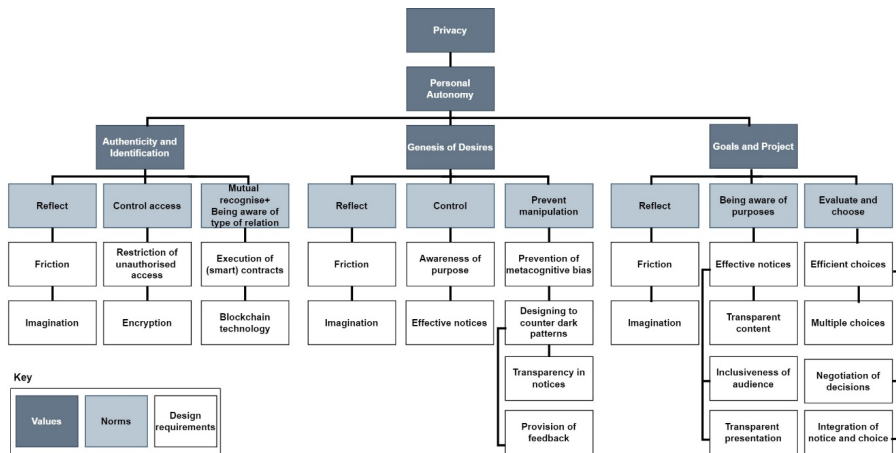


Figure 6.1. Possible Values Hierarchy for Privacy

6.5. Conclusion

This research proposed design requirements for embedding the value of privacy into PIMS. To achieve this goal, a three-layered value hierarchy was constructed. The first layer, dedicated to values, elucidated the connection between privacy and personal autonomy; privacy is functionally valuable for the sake of autonomy. In accordance with the three components of autonomy, namely authentication and identification, the genesis of desires, and goals and projects, the functional value of privacy was discussed. The second layer, dedicated to norms, identified commons and specific norms concerning the components of autonomy, considering that the value of privacy is realised when a person’s autonomy is protected or promoted. In the third layer, design requirements were derived by translating the identified norms. Regarding the common norm, the design for reflection should be incorporated into PIMS. Concerning the specific norms, designing to prevent unauthorised access, to counter

dark patterns, and to provide effective and efficient notices and choices related to the three components of personal autonomy, respectively, should be considered in the design of PIMS. The findings from this study contribute to the literature on privacy by design, emphasising the incorporation of value into the design of PIMS and elevating it beyond mere legal compliance and privacy policy adherence throughout the system-development lifecycle.

Chapter 7: Conclusion

This dissertation had an overarching concern regarding privacy and machine learning-based artificial intelligence (ML-based AI). To address this concern, the study adopted a privacy impact assessment (PIA) as the overall methodology, which encompasses analytical, legal, and technical stages. Accordingly, this dissertation was divided into three parts, each corresponding to one stage of a PIA. This dissertation conducted analytical investigations into privacy and ML-based AI to respond to the first research question (RQ): RQ1. ‘How does an AI system affect privacy?’ Moreover, the legal investigations responded to RQ2. ‘How effectively does the General Data Protection Regulation (GDPR) assess and address privacy issues concerning both individuals and groups?’ Finally, the technical investigations aimed to respond to RQ3. ‘How can the value of privacy be embedded into systems?’

The analytical investigations filled the gap in the literature regarding how inference impacts privacy. Through legal investigations, this dissertation highlighted the limitations of the GDPR in addressing privacy issues about groups designed by algorithms (clustered groups) and individuals as members of a group, and it further suggested how to mitigate those issues. Finally, this dissertation conducted technical investigations to propose design requirements to embed the value of privacy into systems.

This chapter is structured into three sections. Section 7.1 summarises the findings of this dissertation by referring to the main RQs, sub-questions (SQs), and corresponding responses. Section 7.2 highlights the limitations of this dissertation. Finally, Section 7.3 suggests recommendations for future research.

7.1. Overview of Research Findings

To address the gap in the literature concerning the impact of ML-based AI, with a particular focus on inference and privacy, and to highlight the limitations of the GDPR in addressing privacy issues, along with proposing requirements to be considered when designing systems to protect privacy, this dissertation was divided into three parts. Each was dedicated to a specific goal that was aligned with these overarching objectives.

Part I

Part I of this dissertation concerned the analysis of the impacts of ML-based AI on privacy, with a particular focus on inference. The main RQ that I articulated in this part is RQ1: ‘How does an ML-based AI system affect privacy?’ I considered inference to include inferred information, AI models’ performance, and access to information uncovered by an AI model. Moreover, I distinguished between two aspects of privacy: descriptive (i.e., concerning the definition of privacy) and normative, and I further divided the normative aspect into the value of privacy and the right to privacy. Accordingly, I articulated three SQs, each related to inferred information and the definition of privacy, the AI model’s performance and the value of privacy, access to information uncovered by an AI model and the right to privacy. After discussing the responses to the SQs, I return to RQ1.

SQ1: How does inferred information affect the definition of privacy?

In chapter 2, I focused on the most recent descriptive accounts of privacy: source control by Menges (2020a, 2020b) and actual access by Macnish (2018, 2020). I examined these accounts regarding inferred information and argued that specific sets of inferred information present counter-examples to these accounts; it was therefore necessary to revise them to interpret such cases correctly.

The initial definition of the source control account of privacy (Menges, 2020a, 2020b) states that ‘A’s privacy is lost iff: A has lost source control over the personal information P about agent A, if information flows at all’. However, this account does not correctly consider cases of ‘information inferred from once-private information that A has intentionally shared but which does not count as a piece of information that A intended to share’. To address such counter-examples, I revised the initial definition to include the loss of source control over personal information ‘due to the action(s) of agent B, who obtains or infers information contrary to A’s preferences’.

In parallel, the initial definition of the access account of privacy (Macnish, 2018, 2020) states that ‘A’s privacy is lost iff: B actually accesses personal information P about A’. Yet, this account also fails in cases of ‘information inferred from once-private information that A has intentionally shared and which itself counts as a piece of information that A intended to share’. In response, I revised the initial definition to include the consideration ‘B has reason to think that A intends to keep it private’ when describing that B actually accesses personal information.

Ultimately, my comparative analysis demonstrated that, although the revised versions of the descriptive accounts of privacy differ in their underlying rationales, they are extensionally equivalent. That is, the two formulas arrive at the identical result in the same instances: a loss of privacy. Metaphorically, the differences between them can be explained by referring to different sides of the same mountain.

In conclusion, inferred information challenges the current definition of privacy, and my revisions provide a more comprehensive understanding of privacy loss regarding inferred information. This exploration addressed SQ1 by showing how inferred information necessitates a nuanced approach to defining privacy.

SQ2: How does the performance of an AI model affect the social value of privacy?

In Chapter 3, I argued that the performance of AI models—specifically, their accuracy—is crucial in maintaining trust, and thereby in constituting privacy as a social value. Drawing on the insights of Waldman (2015), I argued that privacy is a social value constituted by interactions between different individuals based on trust. This emphasises the role of a person acting as the trustee (that is, the one who is trusted) in constituting privacy. A person (B) who employs an AI system to respond to another person's (A) question (p) is trustworthy iff they fulfil the norms of trust. One norm that B must fulfil to be trustworthy is competence (Hawley, 2019). B is competent only if they are *ex post* justified in believing whether p. I argued that justifying the computational belief involves either B themselves testing and gaining inductive support for the accuracy of the instrument, or the developer of the ML model testifying to some level of accuracy for the model. Therefore, I concluded that the justification of B's belief that p partly relies on the accuracy of the AI model.

In conclusion, accuracy is a feature of an AI model's performance that contributes to the justification of B's belief in p. Given that being justified in believing what the instrument delivers is required for B to be competent in what they assert, the accuracy of an AI model affects B's competence. Trustworthiness requires competence, and so an AI system impacts trust relationships between A and B. Since privacy is constituted by trust-based relationships, the effects of an AI system on trust impact privacy, as well. As trust requires an accurate AI system, privacy does also. In this way, I emphasised not only the role of the trustee, but also an AI system, which enables the trustee to adhere to trust norms while constituting privacy.

SQ3: What impact does accessing information uncovered by AI models have on the privacy of groups, and how can group privacy be respected?

In Chapter 4, I argued that accessing certain pieces of anonymous information about a clustered group, thus enabling the group to be easily identified and targeted—which would likely be used to harm the group in morally objectionable ways—harms the privacy of the group.

To protect the group privacy, I proposed that limitations be imposed regarding access to such information. The predominant approach, as seen in the work of scholars (e.g., Floridi, 2014, 2017; Mantelero, 2017; and van der Sloot, 2017) advocates for the recognition of a right to privacy for the group. Correspondingly, the group right to privacy places a duty on others, for example, private companies, institutions, and governments, to refrain from accessing such information. However, in this chapter, I argued that a clustered group cannot have a right to privacy. I suggested that the moral principle of protecting the vulnerable can be employed to protect group privacy. As a result, I concluded that the duty to respect group privacy is not entailed in the group right to privacy, but rather in the moral principle of protecting vulnerable groups within an ethics of vulnerability.

My argument demonstrating that clustered groups cannot have a right to privacy had two premises; they are presented below.

First, drawing on Raz's (1988) and Réaume's (1988) respective concepts of inherent public and participatory good, I argued that clustered group privacy cannot be conceived of as either an inherent or a participatory good. As a result, if there is a right to clustered group privacy, that right cannot be either a collective or a corporate right.

Second, Raz (1988) and Newman (2004) argue, respectively, that group rights protect or further certain aggregative and non-aggregative group interests. I maintained that a clustered group cannot have either aggregative or non-aggregative interests in clustered group privacy. As a result, a cluster group cannot have the kinds of interests required to establish a group right. I therefore concluded that, on these approaches, a cluster group cannot have a right to privacy.

In conclusion, to protect clustered group privacy, I suggested the moral principle of protecting the vulnerable (Goodin, 1986) within an ethics of vulnerability that can be employed to articulate obligations to limit access to certain pieces of information about a clustered group.

Having responded to the SQs, I return to RQ1. To address RQ1, I summarise insights that resulted from answering SQ1-SQ3.

RQ1: How does an ML-based AI system affect privacy?

1. Effects of inferred information on the definition of privacy: SQ1 examined how inferred information affects the definition of privacy. My analysis revealed that inferred data presents significant challenges to current definitions of privacy. The definitions have been revised to interpret cases involving different sets of inferred information correctly as a loss of privacy.
2. Impacts of an AI model's performance on the social value of privacy: SQ2 focused on the impact of an AI model's performance on the social value of privacy. I argued privacy as a social value, which is constituted in the social context of intersubjectivity based on trust, is affected by an AI model's performance. The accuracy of these models affects the social value of privacy.
3. Impacts of accessing information uncovered by an AI model on the privacy of groups: SQ3 addressed the impact of accessing information uncovered by an AI model on group privacy. My findings indicated that accessing specific anonymous information about clustered groups raises significant concerns about the privacy of these groups. I proposed that while the right to privacy cannot be recognised for such groups to impose limitations on accessing the information, but the principle of protecting the vulnerable can.

In conclusion, ML-based AI affects the descriptive (i.e., the definition of privacy) and normative (i.e., the value of and the right to privacy) aspects of privacy in multifaceted ways: it challenges the current definitions of privacy, influences its social value, and raises new concerns regarding the privacy of groups.

Part II

The second part of the dissertation was dedicated to legal investigations aimed at highlighting the limitations of the GDPR in addressing privacy issues and providing suggestions to mitigate those issues. The RQ that I articulated to be addressed in this part was as follows:

RQ2: How effectively does the GDPR assess and address privacy issues concerning both individuals and groups?

In Chapter 5, I answered this question by examining privacy issues concerning two different cases: first, cases in which a person is targeted as a member of a group, and second, cases in which a group as a whole is targeted for policies and decision-making.

In the first case, I argued inferring group characteristics to identify a person threatens the privacy of the person, as the person might intend to keep that information private, or the information may be inferred contrary to the person's preference. In the second case, I argued accessing certain information about a group that might provide others with information to target the group in morally objectionable ways raises concerns about the privacy of the group.

I argued that the GDPR cannot address the identified privacy issues, as the scope of privacy and data protection is limited to personal information and does not consider information ascribed to or inferred from personal information within that scope—concerning the first case. Moreover, since the GDPR merely focuses on protecting pre-existing, intuitive groups, such as ethnicity and race, it does not consider groups that are designed algorithmically—concerning the second case.

Consequently, I concluded that the GDPR has limitations with respect to protecting individuals as members of a group and the group itself. To address these limitations, I suggested expanding the scope of information protected by the right to privacy and data protection to include information primarily attributed to a person because of their membership in a specific group and to broaden the scope to consider inferred data, as well. Moreover, I recommended establishing principles to respect group privacy and protect vulnerable clustered groups.

Part III

The third part of this dissertation was dedicated to technical investigations that proposed design requirements for protecting privacy in systems by embedding privacy into the design of those systems. This part responded to the following RQ:

RQ3: How can the value of privacy be embedded into systems?

In Chapter 6, drawing on van de Poel's (2013) notion of a value hierarchy, I focused on a hierarchical structure of values, norms, and design requirements, according to which, to incorporate a value into a system, that value must be translated into tangible design requirements derived from the translation of norms. Accordingly, to respond to RQ3, I argued that a value hierarchy for privacy must be constructed in which the value of privacy is translated into norms, and design requirements are formulated based on the translation of norms. In this chapter, the hierarchy was constructed in three stages: first, conceptualising the value of privacy; second, identifying norms translated from the value of privacy; and third, articulating design requirements

through the translation of norms. The design requirements derived from the following stages intentionally embed the value of privacy into systems.

1. Conceptualising the value of privacy: In the top layer of the value hierarchy, I discussed the instrumental value of privacy in connection with one's autonomy; privacy is understood as a means to an end for autonomy. Given the significance of autonomy in conceptualising the value of privacy, I analysed autonomy, adopting Rössler's (2005) account of personal autonomy, with reference to intersubjective relations and social conditions (Mackenzie, 2008). Rössler defines an autonomous person as one who asks themselves practical questions and attempts to live accordingly. Rössler identifies three components of the concept of autonomy: authenticity and identification, the history and genesis of desires, and goals and projects. Thus, the instrumental value of privacy is realised when the three components of autonomy are protected and promoted.
2. Identification of norms translated from the value of privacy: In the middle layer of the hierarchy, I identified specific norms aimed at promoting and protecting the three components of autonomy.
3. Articulation of design requirements through the translation of norms: In the lowest layer of the hierarchy, I proposed design requirements intended to embed the value of privacy into systems.

The specific norms and design requirements related to each component of the conception of autonomy are detailed in Table 7.1.

Table 7.1. *Specific Norms and Design Requirements*

Components of the value of autonomy	Norms	Design requirements
Authenticity and identification	<ul style="list-style-type: none"> • Exercising control over personal information to establish and maintain various social relationships • Being aware of the types of relationships a person is involved in, which helps them decide which part of their information to share • Considering social circumstances that provide a basis for forming affective attitudes towards themselves 	<ul style="list-style-type: none"> • Design to restrict unauthorised access by implementing encryption • Consider the execution of contracts • Employ blockchain technology to apply encryption and fulfil contract needs
History and genesis of desires	<ul style="list-style-type: none"> • Enabling a person to exercise control over their personal information to become less susceptible to manipulation • Preventing manipulation to enable a person to share their personal information as intended 	<ul style="list-style-type: none"> • Design for effective notices • Design against dark pattern to prevent certain cognitive biases occurring
Goals and projects	<ul style="list-style-type: none"> • Enabling a person to contemplate and evaluate different alternatives for sharing information • Providing a person options to choose the one that aligns with their objectives 	<ul style="list-style-type: none"> • Design for effective notices • Design for efficient choice

By following a hierarchal structure that moves between values, norms, and design requirements, which provides a structured way to translate the abstract value of privacy into concrete design requirements that uphold and promote this value in

practical applications, the chapter illustrated how the instrumental value of privacy, particularly in relation to autonomy, can intentionally be embedded into the design of systems.

This dissertation had an overarching objective regarding privacy and ML-based AI. To address this aim, it formulated three RQs that covered philosophical, legal, and technical investigations. The first RQ was further divided into three SQs focusing on the definition of privacy, its value, and the right to privacy. Having responded to the RQs and SQs, this dissertation has fulfilled its objectives. These objectives included conducting a philosophical investigation of how ML-based AI impacts privacy, a legal investigation into the effectiveness of the GDPR in addressing inference, and a technical investigation proposing design requirements to incorporate the value of privacy into systems. These objectives fulfilled the overarching objective regarding privacy and ML-based AI.

7.2. Limitations of the Research

In this section, I identify the limitations of this research regarding the focus and scope of its investigations, as well as the overall methodology adopted. The focus of this dissertation was on inference; the scope included philosophical, legal, and technical investigations; and the overall methodology was the privacy impact assessment (PIA).

First, this dissertation focused on inference as a process associated with ML-based AI systems. It explored how inference impacts the definition of privacy, its value, and the right to privacy. It highlighted the limitations of the GDPR in addressing inference. Nonetheless, focusing solely on inference—a gap identified in the literature—limited the dissertation’s exploration of existing theories about privacy and AI, as well as its comprehensive assessment of the GDPR regarding the discussed privacy issues.

Second, this dissertation has a philosophical-conceptual limitation, as it was limited to a specific account of privacy—personal privacy—and did not consider other accounts, such as institutional or political accounts of privacy. As proponents such as Henschke (2020) and Véliz (2021) have noted, these accounts concern power, and they focus on privacy and how it limits institutional and governmental power. The limitation of concentrating on the personal account of privacy stems from its conceptualisation of privacy. The personal account adopts a horizontal perspective concerned with a person’s control or access. In contrast, institutional or political accounts conceptualise privacy from a vertical perspective, emphasising power and

the dominance of governments and institutions. Consequently, the findings have limitations regarding the vertical perspective of privacy and its associated conceptualisation.

Third, this dissertation has a limitation regarding the legal scope because it did not extend its analysis to the EU AI Act (EU Parliament, 2023). This Act may offer different perspectives, particularly at the group level, which could mitigate the issues that this dissertation identified. It remains an open question, to which this dissertation did not respond, whether the EU AI Act considers the risk of AI systems that potentially impact group privacy, which would require specific rules to mitigate these issues. Therefore, the generalisability of the findings of this dissertation to the context of the EU AI Act is uncertain, and further investigation into this area is warranted to understand the alignment or divergence between the EU AI Act and the legal framework adopted in this dissertation.

Fourth, this dissertation presented a limitation in terms of its technological scope, as it merely focused on ML-based AI systems. Emerging complex, multifunctional virtual worlds such as the metaverse, which integrate various technologies like augmented reality, fall outside the scope of this study. As Brey discusses in his presentation *Virtual Reality and the Metaverse: Ontology and Ethics* (2023), privacy issues in the metaverse include not only information privacy, but also bodily, behavioural, and mental privacy. Since this dissertation was limited to investigating the impacts of AI systems, which heavily depend on personal information, on privacy, it only addressed information privacy. Consequently, the findings are primarily applicable to the impacts of ML-based AI on information privacy and may not extend to these broader and more multifaceted virtual worlds.

Fifth, this dissertation presented a limitation in the analytical stage of the overall methodology. It adopted a PIA that encompassed the analytical, legal, and technical stages. The analytical stage of a PIA is commonly employed to evaluate how specific software or systems impact privacy. However, this dissertation adopted the analytical stage of a PIA regarding a specific process associated with ML-based AI (i.e., inference). Adopting this stage guided me in navigating issues regarding privacy and inference in a comprehensive manner, irrespective of the specific applications of AI systems. This adaptation, although unique and thorough, may not completely align with the conventional use of the first stage of a PIA, potentially limiting its applicability to the standard practices of a PIA.

7.3. Recommendations for Future Research

This section provides recommendations for future works aimed at further expanding upon the insights of each chapter in this dissertation.

7.3.1. Developing a Normative Account Based on the Foundations of the Descriptive Account of Privacy

In chapter 2, according to Gavison's (1980) and Powers's (1996) views, I argued that concentrating on a descriptive conception of privacy is required, as it enables us to build a layer on top of it using criteria to determine how much privacy is good. Given the in-depth exploration of the descriptive conception of privacy through the lens of source control (Menges, 2020a, 2020b) and actual access (Macnish, 2018, 2020) accounts, future research could benefit from developing a normative account which is based on these insights and which goes beyond mere description to examine the moral implications of privacy losses. This involves assessing how descriptive accounts align with or inform moral values concerning privacy.

Future fundamental research on privacy could provide a more comprehensive grasp of the topic, not merely as a neutral concept but also as a moral value. The benefits of such research would include determining the circumstances under which a loss of privacy, as discussed in Chapter 2, would be morally wrong. Additionally, the benefits would extend to the broader objective of developing a theory that clarifies the relationship between a loss and a violation of privacy.

7.3.2. Relationships between Privacy and Trust

In Chapter 3, drawing on the insights of Waldman (2015), I assumed that privacy is constituted by the interaction of different individuals in the social context of intersubjectivity based on trust, and I analysed how AI impacts trust and, eventually, privacy. To examine the assumption I made in this chapter, future research could further explore the relationship between privacy and trust. The importance of such research would be to identify norms to cultivate trust in a context in which personal information is shared and revealed, making that context private.

Regarding the relationships between privacy and trust, the important aspect is to distinguish between necessity and dependence.

1. Necessity vs. Dependence: Future research could investigate whether the relationship between privacy and trust is a necessity, as Fried (1984), Rachels

(1975), and Waldman (2015) argue, or a dependence in the sense that trust is an essential property of privacy in a context in which information is shared and revealed.

2. **Metaphysical or Contingent Claims:** The idea that privacy is constituted by trust can be studied in two ways. First, the idea can be understood as a metaphysical claim, such that trust is seen as essentially part of privacy. Second, it can be understood as a contingent claim about privacy. Such a view considers trust-based relationships as a constitutive element of privacy, while viewing such relationships as evolving, changing, or varying over time or based on circumstances. Accordingly, privacy should be conceived as a concept with variability and contingency.⁴⁹

These areas of exploration would significantly contribute to the ongoing discussion of privacy and trust by providing valuable insights into how privacy and trust interact and form each other in ethical and social contexts. The benefits of elucidating the relationship between privacy and trust involve regulating the trust-promoting norms that govern the relational duties of the person who is trusted regarding how to build trust-based relationships with the one who trusts, thereby making the context suitable for disclosure.

7.3.3. Exploring the Group Right to Privacy on Social Networking Sites

In light of the discussion in Chapter 4 about groups in which there is no tie between the members and the conclusion that such groups cannot have a right to privacy, future research should consider a right to privacy for different groups with ties between members, such as groups on social networking sites. As Barocas and Nissenbaum (2014) have highlighted, on social networking sites, there are meaningful or recognised relations among the individuals who are grouped together, and inferences are made based on confirmed relations among users.

1. **Nature of the Group Right to Privacy:** Future research could investigate the nature of the potential group right on social networking sites, or the group right to ‘networked privacy’, the concept introduced by boyd (2011). Specifically, research should explore whether this right is merely a collection of individual rights or if it is a right to a group as a whole.

⁴⁹ Christman’s (2004) idea of the relationship between autonomy and social relations inspired me to explore dependency in the relationship between privacy and trust.

2. **Comparative Analysis with the Individual Right to Privacy:** Future research should also involve a comparative analysis to understand how a group right to privacy, if one exists, on social networking sites interacts with and differs from the individual right to privacy. This would involve assessing the issues and considerations that arise regarding data-processing and data inferences regarding groups on these platforms.
3. **Implications for Duty-Bearers:** If a group right to privacy is recognised, it is important to identify the corresponding duties that this right would engender for duty-bearers. This includes determining who the duty-bearers are and what specific duty they have to uphold the group right to privacy.

Conducting the proposed research would contribute to our understanding of privacy and the right to privacy in the increasingly interconnected and networked environment of social networking sites.

7.3.4. Evaluating New Data Governance Schemes in Addressing Privacy Issues

Building on the findings of Chapter 5, future research could explore whether and how new data governance schemes can address privacy concerns at both the individual and group levels. Such schemes include data sharing pools, data co-operative, public data trusts, and personal data sovereignty. These schemes are increasingly recognised for their potential to empower data subjects to control and manage their personal data (European Commission, Joint Research Centre et al., 2021).

1. **Personal Data Sovereignty and Control of Inferred Information:** An important area of exploration is how personal data sovereignty schemes enable individuals to control not only their explicit personal data, but also inferred information and information ascribed to them due to their membership in specific groups. This research should investigate the effectiveness of these schemes in empowering individuals to manage data that are derived or inferred from provided information.
2. **Data Co-operative Schemes and Group Privacy:** Another critical aspect to investigate is the applicability of data co-operative schemes to the privacy of clustered groups. Current schemes might primarily focus on groups in which there is interaction among members. Future research should assess how these schemes can be improved or adapted to address the privacy issues of clustered groups.

Future research on the aforementioned issues would provide valuable insights into the potential of new data governance schemes to address contemporary privacy issues. It would also contribute to the development of more robust and inclusive privacy protection principles and measures that account for group privacy and inferred data.

7.3.5. Empirical and Technical Investigations

Building upon the theoretical design requirements proposed in Chapter 6, future research should extend to empirical and technical investigations. As detailed in the following paragraphs, these investigations are necessary to ensure the needs and expectations of stakeholders are considered in the design of systems, and that the proposed design requirements are technically feasible to include in the design process.

1. **Empirical Investigations:** Future research should involve empirical studies to understand the needs and expectations of stakeholders (Friedman et al., 2008). This stage is necessary to:
 - determine stakeholders' needs concerning privacy and systems,
 - explore how technological artefacts are expected to be used in real-world cases, and
 - assess the impacts of technological artefacts on users and other stakeholders.
2. **Technical Investigations:** Future study should involve conducting technical investigations to evaluate:
 - the effectiveness of systems in supporting the value of privacy and
 - the practical feasibility of the proposed design requirements.
3. **Prototype Testing:** An essential component of technical investigations is prototype testing (Umbrello & van de Poel, 2021). This process involves creating and testing prototypes that embody the design requirements established in Chapter 6. The aim is to:
 - identify any unforeseen side effects or challenges in implementing the proposed design requirements and
 - consider the necessity of incorporating additional values or modifying the existing requirements based on practical findings (Umbrello & van de Poel, 2021).

Expanding investigations to include empirical and technical investigations is necessary for a comprehensive understanding of how the value of privacy can be effectively embedded into systems. In this way, the gap between theoretical concepts and

practical implications can be filled, ensuring that the resulting technologies are aligned with stakeholders' needs and expectations.

Looking ahead, a world suffused by AI highlights the imperative to recognise privacy as a social value, requiring collective efforts to realise and preserve it. The encroachment on privacy by AI technologies presents a significant challenge not only to individuals but also to groups and society as a whole, transforming from a concern for a person into a societal issue that demands collective action to preserve privacy. Thus, it is crucial to transition from an individualistic perspective to a broader social viewpoint, conceptualising privacy to include the privacy of groups and establishing social norms to preserve it.

References

- Abualganam, O., Al-Khatib, S., & Hiari, M. (2022). Data Mining Model for Predicting Customer Purchase Behavior in E-Commerce Context. *International Journal of Advanced Computer Science and Applications*, 13, 421. <https://doi.org/10.14569/IJACSA.2022.0130249>
- Akselrod-Ballin, A., Chorev, M., Shoshan, Y., Spiro, A., Hazan, A., Melamed, R., Barkan, E., Herzel, E., Naor, S., Karavani, E., Koren, G., Goldschmidt, Y., Shalev, V., Rosen-Zvi, M., & Guindy, M. (2019). Predicting Breast Cancer by Applying Deep Learning to Linked Health Records and Mammograms. *Radiology*, 292(2), 331–342. <https://doi.org/10.1148/radiol.2019182622>
- Al-Rubaie, M., & Chang, J. M. (2019). Privacy-Preserving Machine Learning: Threats and Solutions. *IEEE Security & Privacy*, 17(2), 49–58. <https://doi.org/10.1109/MSEC.2018.2888775>
- Altman, I. (1975). *The Environment and Social Behavior: Privacy, Personal Space, Territory, Crowding*. Brooks/Cole Publishing Company.
- Annas, G. J., & Grodin, M. A. (Eds.). (1995). *The Nazi Doctors and the Nuremberg Code: Human Rights in Human Experimentation*. Oxford University Press.
- Aouad, L. M., Le-Khac, N., & Kechadi, M. (2007). Lightweight Clustering Technique for Distributed Data Mining Applications. *Industrial Conference on Data Mining*. https://doi.org/10.1007/978-3-540-73435-2_10
- Asgarinia, H. (2023). Convergence of the source control and actual access accounts of privacy. *AI and Ethics*. <https://doi.org/10.1007/s43681-023-00270-z>
- Asgarinia, H., Chomczyk Penedo, A., Esteves, B., & Lewis, D. (2023). “Who Should I Trust with My Data?” Ethical and Legal Challenges for Innovation in New Decentralized Data Management Technologies. *Information*, 14(7)(351). <https://doi.org/10.3390/info14070351>
- Baier, A. (1986). Trust and Antitrust. *Ethics*, 96(2), 231–260. <https://doi.org/10.1086/292745>
- Barocas, S., & Nissenbaum, H. (2009). *On Notice: The Trouble with Notice and Consent* (SSRN Scholarly Paper 2567409). <https://papers.ssrn.com/abstract=2567409>
- Barocas, S., & Nissenbaum, H. (2014). *Big data's end run around anonymity and consent*. <https://doi.org/10.1017/CBO9781107590205.004>

- Bengtsson, L., Lu, X., Thorson, A., Garfieldarfier, R., & Schreeb, J. (2011). Improved Response to Disasters and Outbreaks by Tracking Population Movements With Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. *PLoS Medicine*, 8, e1001083. <https://doi.org/10.1371/journal.pmed.1001083>
- Benito-León, J., Castillo, M. D. del, Estirado, A., Ghosh, R., Dubey, S., & Serrano, J. I. (2021). Using Unsupervised Machine Learning to Identify Age- and Sex-Independent Severity Subgroups Among Patients with COVID-19: Observational Longitudinal Study. *Journal of Medical Internet Research*, 23(5), e25988. <https://doi.org/10.2196/25988>
- Benson, P. (1994). Free Agency and Self-Worth. *The Journal of Philosophy*, 91(12), 650–668. <https://doi.org/10.2307/2940760>
- Bird-Pollan, S. (2009). Rawls: Construction and Justification. *Public Reason*, 1(2).
- Bloustein, E. J. (1964). Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser. *New York University Law Review*, 39, 962.
- Bloustein, E. J. (2019). *Individual and Group Privacy* (N. J. Pallone, Ed.). Routledge.
- boyd, danah. (2011). *Networked Privacy*. Personal Democracy Forum. New York, NY. <https://www.danah.org/papers/talks/2011/PDF2011.html>
- Brandom, R. (1983). Asserting. *Noûs*, 17(4), 637–650. <https://doi.org/10.2307/2215086>
- Brandstedt, E., & Brännmark, J. (2020). Rawlsian Constructivism: A Practical Guide to Reflective Equilibrium. *The Journal of Ethics*, 24(3), 355–373. <https://doi.org/10.1007/s10892-020-09333-3>
- Bratman, M. E. (2007). *Structures of Agency: Essays*. Oxford University Press.
- Brey, P. (2000). Disclosive Computer Ethics. *Acm Sigcas Computers and Society*, 30(4), 10–16. <https://doi.org/10.1145/572260.572264>
- Brey, P. (2010). Values in technology and disclosive computer ethics. In L. Floridi (Ed.), *The Cambridge Handbook of Information and Computer Ethics* (pp. 41–58). Cambridge University Press. <https://doi.org/10.1017/CBO9780511845239.004>
- Brey, P. (2023, June 5). *Virtual reality and the metaverse: Ontology and ethics* [PowerPoint Slides].
- Brey, P., & Dainow, B. (2023). Ethics by design for artificial intelligence. *AI and Ethics*. <https://doi.org/10.1007/s43681-023-00330-4>

- Brunton, F., & Nissenbaum, H. F. (2015). *Obfuscation: A user's guide for privacy and protest*. MIT Press.
- Budin-Ljøsne, I., Teare, H. J. A., Kaye, J., Beck, S., Bentzen, H. B., Caenazzo, L., Collett, C., D'Abramo, F., Felzmann, H., Finlay, T., Javaid, M. K., Jones, E., Katić, V., Simpson, A., & Mascalzoni, D. (2017). Dynamic Consent: A potential solution to some of the challenges of modern biomedical research. *BMC Medical Ethics*, 18(1), 4. <https://doi.org/10.1186/s12910-016-0162-9>
- Calo, M. (2012). Against Notice Skepticism in Privacy (and Elsewhere). *Notre Dame Law Review*, 87(3), 1027.
- Carter, J. A., & Simion, M. (2021). The Ethics and Epistemology of Trust. In *Internet Encyclopaedia of Philosophy*. <https://iep.utm.edu/trust/>
- Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589(7840), 82–87. <https://doi.org/10.1038/s41586-020-2923-3>
- Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, 19(2), 171–209. <https://doi.org/10.1007/s11036-013-0489-0>
- Chowriappa, P., Dua, S., & Todorov, Y. (2014). Introduction to Machine Learning in Healthcare Informatics. In *Intelligent Systems Reference Library* (Vol. 56, pp. 1–23). https://doi.org/10.1007/978-3-642-40017-9_1
- Christidis, K., & Devetsikiotis, M. (2016). Blockchains and Smart Contracts for the Internet of Things. *IEEE Access*, 4, 2292–2303.
- Christman, J. (2004). Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves. *Philosophical Studies*, 117(1), 143–164. <https://doi.org/10.1023/B:PHIL.0000014532.56866.5c>
- Christman, J. (2009). The historical conception of autonomy. In *The Politics of Persons: Individual Autonomy and Socio-historical Selves* (pp. 133–163). Cambridge University Press. <https://doi.org/10.1017/CBO9780511635571.007>
- Citron, D., & Pasquale, F. (2014). The Scored Society: Due Process for Automated Predictions. *Washington Law Review*, 89(1), 1.
- Coady, C. A. J. (1973). Testimony and Observation. *American Philosophical Quarterly*, 10(2), 149–155.
- Coeckelbergh, M. (2020). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, 26(4), 2051–2068. <https://doi.org/10.1007/s11948-019-00146-8>

- Cohen, J. E. (2012). *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice*. New Haven: Yale University Press.
- Committee on Digital Economy & Policy (CDEP). (2022). *OECD Framework for the Classification of AI systems: OECD Digital Economy Papers*. OECD Secretariat. <https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm>
- Copp, D. (1995). *Morality, Normativity, and Society*. Oxford University Press.
- Coron, J.-S. (2006). What is cryptography? *IEEE Security & Privacy*, 4(1), 70–73. <https://doi.org/10.1109/MSP.2006.29>
- Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Daniels, N. (2007). *Just Health: Meeting Health Needs Fairly*. Cambridge Core; Cambridge University Press. <https://doi.org/10.1017/CBO9780511809514>
- de Andrade, N. N. G. (2011). Data Protection, Privacy and Identity: Distinguishing Concepts and Articulating Rights. In S. Fischer-Hübner, P. Duquenoy, M. Hansen, R. Leenes, & G. Zhang (Eds.), *Privacy and Identity Management for Life* (Vol. 352, pp. 90–107). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20769-3_8
- Dodds, S. (2013). Dependence, Care, and Vulnerability. In C. Mackenzie, W. Rogers, & S. Dodds (Eds.), *Vulnerability: New Essays in Ethics and Feminist Philosophy*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199316649.003.0008>
- Donia, J., & Shaw, James. A. (2021). Ethics and Values in Design: A Structured Review and Theoretical Critique. *Science and Engineering Ethics*, 27(5), 57. <https://doi.org/10.1007/s11948-021-00329-2>
- Dwork, C. (2008). Differential Privacy: A Survey of Results. In M. Agrawal, D. Du, Z. Duan, & A. Li (Eds.), *Theory and Applications of Models of Computation* (Vol. 4978, pp. 1–19). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-79228-4_1
- EU Parliament. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) Doc Type: R Usr_lan: En. OJ L. Vol. 119. <http://data.europa.eu/eli/reg/2016/679/oj/eng>*
- EU Parliament. (2023). *Artificial Intelligence Act*. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)

- European Commission, Joint Research Centre, Craglia, M., Scholten, H., Micheli, M., & et al. (2021). *European Commission, Joint Research Centre: Digitranscope – The governance of digitally-transformed society*. Publications Office.
- Fazlioglu, M. (2023, March 22). Most consumers want data privacy and will act to defend it. *Iapp*. <https://iapp.org/news/a/most-consumers-want-data-privacy-and-will-act-to-defend-it/>
- Faulkner, P. (2015). The attitude of trust is basic. *Analysis*, 75(3), 424–429. <https://doi.org/10.1093/analys/anv037>
- Faulkner, P. (2017). The Problem of Trust. In P. Faulkner & T. Simpson (Eds.), *The Philosophy of Trust* (p. 0). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198732549.003.0007>
- Feng, Y., Yao, Y., & Sadeh, N. (2021). A Design Space for Privacy Choices: Towards Meaningful Privacy Control in the Internet of Things. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3411764.3445148>
- Finn, R. L., Wright, D., Finn, R., Wright, D., Friedewald, M., & Isi, F. (2013). Seven Types of Privacy. In *European Data Protection: Coming of Age, Edited By*, 3–32.
- Floridi, L. (2005). The Ontological Interpretation of Informational Privacy. *Ethics and Information Technology*, 7(4), 185–200. <https://doi.org/10.1007/s10676-006-0001-7>
- Floridi, L. (2014). Open Data, Data Protection, and Group Privacy. *Philosophy & Technology*, 27(1), 1–3. <https://doi.org/10.1007/s13347-014-0157-8>
- Floridi, L. (2017). Group Privacy: A Defence and an Interpretation. In L. Taylor, L. Floridi, & B. van der Sloot (Eds.), *Group Privacy: New Challenges of Data Technologies* (pp. 83–100). Springer. https://doi.org/10.1007/978-3-319-46608-8_5
- Fodor, J. A. (1974). Special Sciences (Or: The Disunity of Science as a Working Hypothesis). *Synthese*, 28(2), 97–115.
- Frankfurt, H. G. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, 66(23), 829–839. <https://doi.org/10.2307/2023833>
- Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1), 5–20. <https://doi.org/10.2307/2024717>
- Frankfurt, H. G. (1987). Identification and wholeheartedness. In F. Schoeman (Ed.), *Responsibility, Character, and the Emotions New Essays in Moral Psychology* (pp. 27–45). Cambridge University Press.

- Freeman, K., Geppert, J., Stinton, C., Todkill, D., Johnson, S., Clarke, A., & Taylor-Phillips, S. (2021). Use of artificial intelligence for image analysis in breast cancer screening programmes: Systematic review of test accuracy. *BMJ*, *374*, n1872. <https://doi.org/10.1136/bmj.n1872>
- Freeman, M. (1995). Are there Collective Human Rights? *Political Studies*, *43*(1), 25–40. <https://doi.org/10.1111/j.1467-9248.1995.tb01734.x>
- French, P. A. (1984). *Collective and Corporate Responsibility*. Columbia University Press.
- Fricker, E. (2018). *Trust and Testimonial Justification*. 1–19.
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
- Fried, C. (1968). Privacy. A moral analysis. *Yale Law Journal*, *77*, 475–493.
- Fried, C. (1984). Privacy [a moral analysis]. In F. D. Schoeman (Ed.), *Philosophical Dimensions of Privacy* (1st ed., pp. 203–222). Cambridge University Press. <https://doi.org/10.1017/CBO9780511625138.008>
- Friedman, B., Jr, P. H. K., & Borning, A. (2008). Value Sensitive Design and Information Systems. In K. E. Himma & H. T. Tavani (Eds.), *The Handbook of Information and Computer Ethics* (pp. 69–101). John Wiley & Sons, Inc.
- Garrett, L. (1996). *The coming plague: Newly emerging diseases in a world out of balance*. Penguin.
- Gavison, R. (1980). Privacy and the Limits of Law. *The Yale Law Journal*, *89*(3), 421–471. JSTOR. <https://doi.org/10.2307/795891>
- Gavison, R. (1984). Privacy and the limits of law. In F. D. Schoeman (Ed.), *Philosophical Dimensions of Privacy* (1st ed., pp. 346–402). Cambridge University Press. <https://doi.org/10.1017/CBO9780511625138.017>
- Gellner, E. (1959). Holism versus Individualism in History and Sociology. In P. L. Gardiner (Ed.), *Theories of History: Readings from Classical and Contemporary Sources* (pp. 489–503). Free Press.
- Goffman, E. (1959). *The presentation of self in everyday life*. Doubleday.
- Goldman, A. I. (1979). What is Justified Belief? In G. S. Pappas (Ed.), *Justification and Knowledge: New Studies in Epistemology* (pp. 1–23). Springer Netherlands. https://doi.org/10.1007/978-94-009-9493-5_1
- Goldman, A. I. (2015a). The Structure of Justification. In A. I. Goldman & M. McGrath, *Epistemology: A Contemporary Introduction* (pp. 3–24). Oxford University Press.

- Goldman, A. I. (2015b). Two Debates About Justification: Evidentialism vs. Reliabilism and Internalism vs. Externalism. In A. I. Goldman & M. McGrath, *Epistemology: A Contemporary Introduction* (pp. 25–50). Oxford University Press.
- Goodin, R. E. (1986). *Protecting the Vulnerable: A Re-Analysis of our Social Responsibilities*. University of Chicago Press.
- Grabenwarter, C. (2014). *The European Convention on Human Rights: A Commentary* (Eerste editie). Hart Publishing.
- Grannis, A. (2015). You Didn't Even Notice: Elements of Effective Online Privacy Policies. *Fordham Urban Law Journal*, 42(5), 1109–1170.
- Grindrod, J. (2019). Computational beliefs. *Inquiry*, 1–22. <https://doi.org/10.1080/0020174X.2019.1688178>
- Haldar, P., Pavord, I. D., Shaw, D. E., Berry, M. A., Thomas, M., Brightling, C. E., Wardlaw, A. J., & Green, R. H. (2008). Cluster Analysis and Clinical Asthma Phenotypes. *American Journal of Respiratory and Critical Care Medicine*, 178(3), 218–224. <https://doi.org/10.1164/rccm.200711-1754OC>
- Hardin, R. (2002). *Trust and trustworthiness* (pp. xxi, 234). Russell Sage Foundation.
- Hart, H. L. A. (1982). Legal Rights. In his *Essays on Bentham*. Oxford: Clarendon Press, pp. 162–193.
- Hawley, K. (2014). Trust, Distrust and Commitment. *Nous*, 48(1), 1–20. <https://doi.org/10.1111/nous.12000>
- Hawley, K. (2019). *How To Be Trustworthy* | *Oxford Academic*. Oxford University Press.
- Held, V. (2006). *The Ethics of Care: Personal, Political, and Global*. Oxford University Press, USA.
- Henschke, A. (2017). *Ethics in an Age of Surveillance: Personal Information and Virtual Identities*. Cambridge University Press. <https://doi.org/10.1017/9781316417249>
- Henschke, A. (2020). Privacy, the Internet of Things and State Surveillance: Handling Personal Information Within an Inhuman System. *Moral Philosophy and Politics*, 7(1), 123–149. <https://doi.org/10.1515/mopp-2019-0056>
- Henschke, A. (2021). From need to share to need to care: Information aggregation and the need to care about how surveillance technologies are used for counter-terrorism. In S. Miller, A. Henschke, & J. F. Feltes, *Counter-Terrorism* (pp. 156–168). Edward Elgar Publishing. <https://doi.org/10.4337/9781800373075.00019>

- Hieronymi, P. (2008). The reasons of trust. *Australasian Journal of Philosophy*, 86(2), 213–236. <https://doi.org/10.1080/00048400801886496>
- Hinchman, E. S. (2005). Telling as Inviting to Trust. *Philosophy and Phenomenological Research*, 70(3), 562–587. <https://doi.org/10.1111/j.1933-1592.2005.tb00415.x>
- Hoepman, J.-H. (2022). *Privacy Design Strategies (The Little Blue Book)*.
- Hölbl, M., Kompara, M., Kamišalić, A., & Nemeč Zlatolas, L. (2018). A Systematic Review of the Use of Blockchain in Healthcare. *Symmetry*, 10(10), Article 10. <https://doi.org/10.3390/sym10100470>
- Holton, R. (1994). Deciding to trust, coming to believe. *Australasian Journal of Philosophy*, 72(1), 63–76. <https://doi.org/10.1080/00048409412345881>
- Hughes, K. (2015). The social value of privacy, the value of privacy to society and human rights discourse. In B. Rössler & D. Mokrosinska (Eds.), *Social Dimensions of Privacy: Interdisciplinary Perspectives* (pp. 225–243). Cambridge University Press.
- Inness, J. C. (1992). *Privacy, Intimacy, and Isolation*. Oxford University Press.
- Inness, J. C. (1996). *Privacy, Intimacy, and Isolation*. Oxford University Press. <https://doi.org/10.1093/0195104609.001.0001>
- Irti, C. (2022). Personal Data, Non-personal Data, Anonymised Data, Pseudonymised Data, De-identified Data. In R. Senigaglia, C. Irti, & A. Bernes (Eds.), *Privacy and Data Protection in Software Services* (pp. 49–57). Springer. https://doi.org/10.1007/978-981-16-3049-1_5
- Islam, M. S., Kuzu, M., & Kantarcioglu, M. (2014). Inference attack against encrypted range queries on outsourced databases. *Proceedings of the 4th ACM Conference on Data and Application Security and Privacy*, 235–246. <https://doi.org/10.1145/2557547.2557561>
- Iversen, O. S., Halskov, K., & Leong, T. W. (2012). Values-led participatory design. *CoDesign*, 8(2–3), 87–103. <https://doi.org/10.1080/15710882.2012.672575>
- Janssen, H., & Singh, J. (2022). Personal Information Management Systems. *Internet Policy Review*, 11(2), 1–6. <https://doi.org/10.14763/2022.2.1659>
- Jones, K. (1996). Trust as an Affective Attitude. *Ethics*, 107(1), 4–25.
- Jones, P. (1999). Group Rights and Group Oppression. *Journal of Political Philosophy*, 7(4), 353–377. <https://doi.org/10.1111/1467-9760.00081>

- Jones, P. (Ed.). (2009). *Group Rights (The International Library of Essays on Rights)* (1st edition). Routledge.
- Jones, P. (2010). Cultures, group rights, and group-differentiated rights. In M. Dimova-Cookson & P. Stirk (Eds.), *Multiculturalism and Moral Conflict*, pp. 38–57.
- Jones, P. (2013). Groups and human rights. In C. Holder & D. Reidy (Eds.), *Human Rights: The Hard Questions* (pp. 100–114). Cambridge University Press.
- Kagan, S. (1998). Rethinking Intrinsic Value. *The Journal of Ethics*, 2(4), 277–297.
- Kammourieh, L., Baar, T., Berens, J., Letouzé, E., Manske, J., Palmer, J., Sangokoya, D., & Vinck, P. (2017). Group Privacy in the Age of Big Data. In *Group Privacy: New Challenges of Data Technologies* (pp. 37–66). Springer International Publishing.
- Kant, I. (1993). *Grounding for the metaphysics of morals; with, On a supposed right to lie because of philanthropic concerns* (J. W. (James W. Ellington, Trans.). Indianapolis: Hackett Pub. Co.
- Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics*, 23(2), 141–146. <https://doi.org/10.1038/ejhg.2014.71>
- Khezr, S., Moniruzzaman, M., Yassine, A., & Benlamri, R. (2019). Blockchain Technology in Healthcare: A Comprehensive Review and Directions for Future Research. *Applied Sciences*, 9(9), Article 9. <https://doi.org/10.3390/app9091736>
- Klein, C., Zeng, Q., Arbaretaz, F., Devèvre, E., Calderaro, J., Lomenie, N., & Maiuri, M. C. (2021). Artificial intelligence for solid tumour diagnosis in digital pathology. *British Journal of Pharmacology*, 178(21), 4291–4315. <https://doi.org/10.1111/bph.15633>
- Knobel, C., & Bowker, G. C. (2011). Values in design. *Communications of the ACM*, 54(7), 26–28. <https://doi.org/10.1145/1965724.1965735>
- Korsgaard, C. M. (1983). Two Distinctions in Goodness. *The Philosophical Review*, 92(2), 169–195. <https://doi.org/10.2307/2184924>
- Korsgaard, C. M. (1996). *The Sources of Normativity* (O. O'Neill, Ed.). Cambridge University Press.
- Korsgaard, C. M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.

- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5805. <https://doi.org/10.1073/pnas.1218772110>
- Lackey, J. (2008). *Learning From Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press.
- Lever, A. (2012). *On Privacy*. New York: Routledge.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors (Basel, Switzerland)*, 18(8), 2674. <https://doi.org/10.3390/s18082674>
- Lippert-Rasmussen, K., & Aastrup Munch, L. (2021). Price Discrimination in the Digital Age. In C. Véliz (Ed.), *The Oxford Handbook of Digital Ethics*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198857815.013.24>
- List, C., & Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199591565.001.0001>
- Liu, J., Yang, B., Cheung, W. K., & Yang, G. (2012). Malaria transmission modelling: A network perspective. *Infectious Diseases of Poverty*, 1, 11. <https://doi.org/10.1186/2049-9957-1-11>
- Loi, M., & Christen, M. (2020). Two Concepts of Group Privacy. *Philosophy & Technology*, 33(2), 207–224. <https://doi.org/10.1007/s13347-019-00351-0>
- Mackenzie, C. (2000). Imagining Oneself Otherwise. In C. Mackenzie & N. Stoljar (Eds.), *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. Oup Usa.
- Mackenzie, C. (2008). Relational Autonomy, Normative Authority and Perfectionism. *Journal of Social Philosophy*, 39(4), 512–533. <https://doi.org/10.1111/j.1467-9833.2008.00440.x>
- Mackenzie, C. (2023). Autonomous agency, we-agency, and social oppression. *The Southern Journal of Philosophy*, 61(2), 373–389. <https://doi.org/10.1111/sjp.12521>
- Mackenzie, C., & Stoljar, N. (2000). Introduction. In C. Mackenzie & N. Stoljar (Eds.), *Relational Autonomy Feminist Perspectives on Autonomy, Agency, and the Social Self*. Oxford University Press.

- Macnish, K. (2012). Unblinking eyes: The ethics of automating surveillance. *Ethics and Information Technology*, 14(2), 151–167. <https://doi.org/10.1007/s10676-012-9291-0>
- Macnish, K. (2018). Government Surveillance and Why Defining Privacy Matters in a Post-Snowden World. *Journal of Applied Philosophy*, 35(2), 417–432. <https://doi.org/10.1111/japp.12219>
- Macnish, K. (2019). Informed Consent. In C. Veliz (Ed.), *Data, Privacy and the Individual*. IE University Press.
- Macnish, K. (2020). Mass Surveillance: A Private Affair? *Moral Philosophy and Politics*, 7(1), 9–27. <https://doi.org/10.1515/mopp-2019-0025>
- Macnish, K., & Asgarinia, H. (2023). Privacy and the media. In C. Fox & J. Saunders (Eds.), *The Routledge Handbook of Philosophy and Media Ethics (Routledge Handbooks in Applied Ethics)*. Routledge.
- Mainz, J. T., & Uhrenfeldt, R. (2021). Too Much Info: Data Surveillance and Reasons to Favor the Control Account of the Right to Privacy. *Res Publica*, 27(2), 287–302. <https://doi.org/10.1007/s11158-020-09473-1>
- Mantelero, A. (2017). From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era. In B. van der Sloot, L. Floridi, & L. Taylor (Eds.), *Group Privacy: New Challenges of Data Technologies*. Springer.
- Mariner, W. K. (2007). *Mission Creep: Public Health Surveillance and Medical Privacy* (SSRN Scholarly Paper ID 1033528). Social Science Research Network. <https://papers.ssrn.com/abstract=1033528>
- Marmor, A. (2015). What Is the Right to Privacy? *Philosophy & Public Affairs*, 43(1), 3–26. <https://doi.org/10.1111/papa.12040>
- Mazo, C., Kearns, C., Mooney, C., & Gallagher, W. M. (2020). Clinical Decision Support Systems in Breast Cancer: A Systematic Review. *Cancers*, 12(2), E369. <https://doi.org/10.3390/cancers12020369>
- McDonald, M. (1991). Should Communities Have Rights? Reflections on Liberal Individualism. *Canadian Journal of Law & Jurisprudence*, 4(2), 217–237. <https://doi.org/10.1017/S0841820900002915>
- McGee, P., Murphy, H., & Bradshaw, T. (2020, April 28). Coronavirus apps: The risk of slipping into a surveillance state. *Financial Times*. <https://www.ft.com/content/d2609e26-8875-11ea-a01c-a28a3e3fbd33>
- McKenna, M., & Coates, D. J. (2004). *Compatibilism*. <https://plato.stanford.edu/archives/spr2020/entries/compatibilism/>

- McLeod, C. (2002). *Self-Trust and Reproductive Autonomy*. MIT Press. <https://mitpress.mit.edu/9780262537230/self-trust-and-reproductive-autonomy/>
- McLeod, C. (2021). Trust. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2021/entries/rust/>
- Menges, L. (2020a). A Defense of Privacy as Control. *The Journal of Ethics*, 25(3), 385–402. <https://doi.org/10.1007/s10892-020-09351-1>
- Menges, L. (2020b). Did the NSA and GCHQ Diminish Our Privacy? What the Control Account Should Say. *Moral Philosophy and Politics*, 7(1), 29–48. <https://doi.org/10.1515/mopp-2019-0063>
- Merton, R. K. (1968). *Social Theory and Social Structure*. Free Press.
- Mikkelsen, D., Soller, H., & Strandell-Jansson, M. (2020). *Data privacy in the pandemic* | *McKinsey*. <https://www.mckinsey.com/business-functions/risk-and-resilience/our-insights/privacy-security-and-public-health-in-a-pandemic-year>
- Miller, S. (2001). *Social Action: A Teleological Account*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511612954>
- Miller, S., & Bossomaier, T. (2021). Privacy, Encryption and Counter-Terrorism. In A. Henschke, A. Reed, S. Robbins, & S. Miller (Eds.), *Counter-Terrorism, Ethics and Technology: Emerging Challenges at the Frontiers of Counter-Terrorism* (pp. 139–154). Springer International Publishing. https://doi.org/10.1007/978-3-030-90221-6_9
- Montezuma, L. A., & Taubman-Bassirian, T. (2019). *How to avoid consent fatigue*. <https://iapp.org/news/a/how-to-avoid-consent-fatigue/>
- Moore, A. D. (2010). *Privacy Rights: Moral and Legal Foundations*. Pennsylvania State University Press.
- Morley, J., Cows, J., Taddeo, M., & Floridi, L. (2020). Ethical guidelines for COVID-19 tracing apps. *Nature*, 582(7810), 29–31. <https://doi.org/10.1038/d41586-020-01578-0>
- Mühlhoff, R. (2021). Predictive privacy: Towards an applied ethics of data analytics. *Ethics and Information Technology*, 23(4), 675–690. <https://doi.org/10.1007/s10676-021-09606-x>
- Munch, L. A. (2021). Privacy Rights and “Naked” Statistical Evidence. *Philosophical Studies*, 178(11), 3777–3795. <https://doi.org/10.1007/s11098-021-01625-0>

- Naehrig, M., Lauter, K., & Vaikuntanathan, V. (2011). Can Homomorphic Encryption be Practical? *Proceedings of the 3rd ACM Workshop on Cloud Computing Security Workshop*, 124. <https://doi.org/10.1145/2046660.2046682>
- Nehf, J. (2003). Recognizing the Societal Value in Information Privacy. *Washington Law Review*, 78.
- Newlands, G., Lutz, C., Tamò-Larrioux, A., Villaronga, E. F., Harasgama, R., & Scheitlin, G. (2020). Innovation under pressure: Implications for data privacy during the Covid-19 pandemic. *Big Data & Society*, 7(2), 2053951720976680. <https://doi.org/10.1177/2053951720976680>
- Newman, D. (2011). *Community and Collective Rights: A Theoretical Framework for Rights Held by Groups* (1st edition). Hart Publishing.
- Newman, D. G. (2004). Collective Interests and Collective Rights. *The American Journal of Jurisprudence*, 49(1), 127–163. <https://doi.org/10.1093/ajj/49.1.127>
- Nissenbaum, H. F. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford Law Books.
- OECD (Ed.). (2021). *Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers*. OECD. <https://doi.org/10.1787/eb61fd29-en>
- Ohm, P. (2009). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 57(6), 1701–1777.
- O’Neill, O. (2003). Some limits of informed consent. *Journal of Medical Ethics*, 29(1), 4–7. <https://doi.org/10.1136/jme.29.1.4>
- Orwell, G. (1949). *Nineteen Eighty-Four (1984)*. Penguin UK.
- Oshana, M. A. L. (1998). Personal Autonomy and Society. *Journal of Social Philosophy*, 29(1), 81–102. <https://doi.org/10.1111/j.1467-9833.1998.tb00098.x>
- Parent, W. A. (1983). Privacy, Morality, and the Law. *Philosophy & Public Affairs*, 12(4), 269–288. JSTOR.
- Parfit, D. (2011). *On What Matters: Volume One*. Oxford University Press.
- Post, R. C. (1989). The Social Foundations of Privacy: Community and Self in the Common Law Tort. *California Law Review*, 77(5), 957–1010. <https://doi.org/10.2307/3480641>
- Powers, M. (1996). A Cognitive Access Definition of Privacy. *Law and Philosophy*, 15(4), 369–386. JSTOR. <https://doi.org/10.2307/3505032>

- Preda, A. (2012). Group Rights and Group Agency. *Journal of Moral Philosophy*, 9, 229–254. <https://doi.org/10.1163/174552412X625736>
- Puri, A. (2022). *The group right to privacy* [University of St Andrews]. <https://research-repository.st-andrews.ac.uk/handle/10023/25152>
- Rachels, J. (1975). Why Privacy is Important. *Philosophy and Public Affairs*, 4(4), 323–333.
- Rawls, J. (1997). The Idea of Public Reason Revisited. *The University of Chicago Law Review*, 64(3), 765–807. <https://doi.org/10.2307/1600311>
- Raz, J. (1988). *The Morality of Freedom*. Oxford University Press.
- Réaume, D. (1988). Individuals, Groups, and Rights to Public Goods. *The University of Toronto Law Journal*, 38(1), 1–27. <https://doi.org/10.2307/825760>
- Regan, P. M. (1995). *Legislating Privacy: Technology, Social Values, and Public Policy*. Chapel Hill: University of North Carolina Press.
- Reiman, J. H. (1976). Privacy, Intimacy, and Personhood. *Philosophy & Public Affairs*, 6(1), 26–44. JSTOR.
- Reiman, J. H. (1995). Driving to the Panopticon: A Philosophical Exploration of the Risks to Privacy Posed by the Highway Technology of the Future. *Santa Clara Computer and High-Technology Law Journal*, 11, 27.
- Richards, N., & Hartzog, W. (2017). Privacy's Trust Gap: A Review. *The Yale Law Journal*, 126: 1180, 45.
- Richards, N. M. (2008). Intellectual Privacy. *Washington U. School of Law*, 87, 387–445.
- Richards, N. M., & Hartzog, W. (2020). A Relational Turn for Data Protection? *European Data Protection Law Review*, 6(4), 492–497. <https://doi.org/10.21552/edpl/2020/4/5>
- Riesman, D. (1952). *Faces in the Crowd: Individual Studies in Character and Politics*. Yale University Press.
- Rodríguez, J.-V., Rodríguez-Rodríguez, I., & Woo, W. L. (2022). On the application of machine learning in astronomy and astrophysics: A text-mining-based scientometric analysis. *WIREs Data Mining and Knowledge Discovery*, 12(5), e1476. <https://doi.org/10.1002/widm.1476>
- Rogers, W. (2013). Vulnerability and Bioethics. In C. Mackenzie, W. Rogers, & S. Dodds (Eds.), *Vulnerability: New Essays in Ethics and Feminist Philosophy*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199316649.003.0003>

- Ross, C. (2020, April 15). 5 burning questions about tech efforts to track Covid-19 cases. *STAT*. <https://www.statnews.com/2020/04/15/coronavirus-digital-contact-tracing-tech-questions/>
- Rössler, B. (2005). *The Value of Privacy* | Wiley. Polity.
- Rössler, B., & Mokrosinska, D. (2013). Privacy and social interaction. *Philosophy & Social Criticism*, 39(8), 771–791. <https://doi.org/10.1177/0191453713494968>
- Rumbold, B., & Wilson, J. (2019). Privacy Rights and Public Information. *Journal of Political Philosophy*, 27(1), 3–25. <https://doi.org/10.1111/jopp.12158>
- Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM J. Res. Dev.* <https://doi.org/10.1147/rd.33.0210>
- Scanlon, T. (1975). Thomson on Privacy. *Philosophy & Public Affairs*, 4(4), 315–322. JSTOR.
- Schaub, F., Balebako, R., & Cranor, L. F. (2017). Designing Effective Privacy Notices and Controls. *IEEE Internet Computing*, 21(3), 70–77. <https://doi.org/10.1109/MIC.2017.75>
- Schaub, F., Balebako, R., Durity, A. L., & Cranor, L. F. (2018). A Design Space for Effective Privacy Notices. In E. Selinger, J. Polonetsky, & O. Tene (Eds.), *The Cambridge Handbook of Consumer Privacy* (1st ed., pp. 365–393). Cambridge University Press. <https://doi.org/10.1017/9781316831960.021>
- Sharon, T. (2020). Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers. *Ethics and Information Technology*. <https://doi.org/10.1007/s10676-020-09547-x>
- Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership Inference Attacks Against Machine Learning Models. *2017 IEEE Symposium on Security and Privacy (SP)*, 3–18. <https://doi.org/10.1109/SP.2017.41>
- Solove, D. J. (2008). *Understanding Privacy* (SSRN Scholarly Paper ID 1127888). Social Science Research Network. <https://papers.ssrn.com/abstract=1127888>
- Solove, D. J. (2013). Introduction: Privacy Self-Management and the Consent Dilemma. *Harvard Law Review*, 126:1880, 1880–1903.
- Solove, D. J. (2015). The meaning and value of privacy. In B. Rössler & D. Mokrosinska (Eds.), *Social Dimensions of Privacy: Interdisciplinary Perspectives* (pp. 71–81). Cambridge University Press.

- Sosa, E. (2006). Knowledge: Instrumental and Testimonial. In J. Lackey & E. Sosa (Eds.), *The Epistemology of Testimony* (pp. 116–124). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199276011.003.0006>
- Sosa, E. (2010). How Competence Matters in Epistemology. *Philosophical Perspectives*, 24(1), 465–475. <https://doi.org/10.1111/j.1520-8583.2010.00200.x>
- Spiekermann, S., & Winkler, T. (2020). *Value-based Engineering for Ethics by Design*.
- Steeves, V. (2009). Reclaiming the Social Value of Privacy. In I. Kerr, V. Steeves, & C. Lucock (Eds.), *Lessons from the Identity Trail: Anonymity, Privacy, and Identity in a Networked Society* (pp. 109–208). New York: Oxford University Press.
- Strauß, S. (2017). Privacy Analysis – Privacy Impact Assessment. In S.-O. Hansson (Ed.), *The Ethics of Technology – Methods and Approaches* (pp. 143–156). Rowman and Littlefield International.
- Strawson, P. F. (1974). Freedom and resentment. In *Freedom and resentment, and other essays* (First Edition, pp. 1–25). Methuen [distributed in the USA by Harper & Row, Barnes & Noble Import Division].
- Sundquist, C. P. (2021). Pandemic Surveillance Discrimination. *SETON HALL LAW REVIEW*, 51, 13.
- Tancock, D., Pearson, S., & Charlesworth, A. (2010). The emergence of privacy impact assessments. *HP Laboratories Technical Report*, 63. Scopus.
- Tatem, A., Qiu, Y., Smith, D. L., Sabot, O., Ali, A. S., & Moonen, B. (2009). The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. *Malaria Journal*. <https://doi.org/10.1186/1475-2875-8-287>
- Taylor, L. (2017). Safety in Numbers? Group Privacy and Big Data Analytics in the Developing World. In L. Taylor, L. Floridi, & B. van der Sloot (Eds.), *Group Privacy: New Challenges of Data Technologies*. Springer.
- Taylor, L., Floridi, L., & van der Sloot, B. (2017). Introduction: A New Perspective on Privacy. In L. Taylor, L. Floridi, & B. van der Sloot (Eds.), *Group Privacy: New Challenges of Data Technologies* (pp. 10–22). Springer International Publishing.
- Taylor, L., van der Sloot, B., & Floridi, L. (2017). Conclusion: What Do We Know About Group Privacy? In L. Taylor, L. Floridi, & B. van der Sloot (Eds.), *Group Privacy: New Challenges of Data Technologies* (pp. 225–237). Springer. https://doi.org/10.1007/978-3-319-46608-8_12

- Terpstra, A., Schouten, A. P., Rooij, A. de, & Leenes, R. E. (2019). Improving privacy choice through design: How designing for reflection could support privacy self-management. *First Monday*. <https://doi.org/10.5210/fm.v24i7.9358>
- Thomson, J. J. (1975). The Right to Privacy. *Philosophy & Public Affairs*, 4(4), 295–314. JSTOR.
- Tourani, R., Misra, S., Mick, T., & Panwar, G. (2018). Security, Privacy, and Access Control in Information-Centric Networking: A Survey. *IEEE Communications Surveys & Tutorials*, 20(1), 566–600. <https://doi.org/10.1109/COMST.2017.2749508>
- Umbrello, S., & van de Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. *AI and Ethics*, 1(3), 283–296. <https://doi.org/10.1007/s43681-021-00038-3>
- United Nations. (1948). *Universal Declaration of Human Rights*. United Nations; United Nations. <https://www.un.org/en/about-us/universal-declaration-of-human-rights>
- van de Poel, I. (2009). Values in Engineering Design. In *Philosophy of Technology and Engineering Sciences* (pp. 973–1006). Elsevier. <https://doi.org/10.1016/B978-0-444-51667-1.50040-9>
- van de Poel, I. (2013). Translating Values into Design Requirements. In D. P. Michelfelder, N. McCarthy, & D. E. Goldberg (Eds.), *Philosophy and Engineering: Reflections on Practice, Principles and Process* (Vol. 15, pp. 253–266). Springer Netherlands. https://doi.org/10.1007/978-94-007-7762-0_20
- van de Poel, I. (2021a). Design for value change. *Ethics and Information Technology*, 23(1), 27–31. <https://doi.org/10.1007/s10676-018-9461-9>
- van de Poel, I. (2021b). Values and Design. In D. P. Michelfelder & N. Doorn (Eds.), *The Routledge Handbook of the Philosophy of Engineering* (pp. 300–314). Routledge - Taylor & Francis Group.
- van den Hoven, J. (1997). Privacy and the varieties of informational wrongdoing. *ACM Sigcas Computers and Society*, 27, 33–37. <https://doi.org/10.1145/270858.270868>
- van den Hoven, J. (2009). Information Technology, Privacy, and the Protection of Personal Data. In J. van den Hoven & J. Weckert (Eds.), *Information Technology and Moral Philosophy*. Cambridge University Press.
- van der Burg, S., Kloppenburg, S., Kok, E. J., & van der Voort, M. (2021). Digital twins in agri-food: Societal and ethical themes and questions for further research. *NJAS: Impact in Agricultural and Life Sciences*, 93(1), 98–125. <https://doi.org/10.1080/27685241.2021.1989269>

- van der Sloot, B. (2017). Do Groups Have a Right to Protect Their Group Interest in Privacy and Should They? Peeling the Onion of Rights and Interests Protected Under Article 8 ECHR. In L. Taylor, L. Floridi, & B. van der Sloot (Eds.), *Group Privacy: New Challenges of Data Technologies* (pp. 197–224). Springer.
- van Dyke, V. (1977). The Individual, the State, and Ethnic Communities in Political Theory. *World Politics*, 29(3), 343–369.
- Vedder, A. (1999). KDD: The challenge to individualism. *Ethics and Information Technology*, 1(4), 275–281. <https://doi.org/10.1023/A:1010016102284>
- Vedder, A. H. (2000). *Law and Medicine Current Legal Issues Volume 3* (M. Freeman & A. Lewis, Eds.). Oxford University Press.
- Véliz, C. (2021). *Privacy is Power: Why and How You Should Take Back Control of Your Data*. Bantam Press.
- Véliz, C. (2022). The Surveillance Delusion. In C. Véliz (Ed.), *The Oxford Handbook of Digital Ethics* (p. C30.P1-C30.N7). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198857815.013.30>
- Wachter, S. (2018). Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. *Computer Law & Security Review*, 34(3), 436–449. <https://doi.org/10.1016/j.clsr.2018.02.002>
- Wachter, S., & Mittelstadt, B. (2018). *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI* (SSRN Scholarly Paper ID 3248829). Social Science Research Network. <https://papers.ssrn.com/abstract=3248829>
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), 76–99. <https://doi.org/10.1093/idpl/ix005>
- Waldman, A. E. (2015). Privacy as Trust: Sharing Personal Information in a Networked World. *UNIVERSITY OF MIAMI LAW REVIEW*, 69, 72.
- Waldman, A. E. (2016). Privacy, Notice, and Design. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2780305>
- Waldman, A. E. (2018). Privacy, Notice, and Design. *STANFORD TECHNOLOGY LAW REVIEW*, 21:129.
- Waldman, A. E. (2020). Cognitive biases, dark patterns, and the ‘privacy paradox.’ *Current Opinion in Psychology*, 31, 105–109. <https://doi.org/10.1016/j.copsyc.2019.08.025>

- Warren, S. D., & Brandeis, L. D. (1890). The Right to Privacy. *Harvard Law Review*, 4(5), 193–220. JSTOR. <https://doi.org/10.2307/1321160>
- Wenar, L. (2023). Rights. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=rights>
- Westin, A. F. (1967). *Privacy and freedom*. Atheneum.
- Widdows, H. (2013). *The Connected Self: The Ethics and Governance of the Genetic Individual*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139051798>
- Williams, B. (1976). Persons, Character, and Morality. In J. Rachels (Ed.), *Moral Luck: Philosophical Papers 1973-1980*. Cambridge University Press.
- Wright, D., & de Hert, P. (2012). Introduction to Privacy Impact Assessment. In D. Wright & P. de Hert (Eds.), *Privacy Impact Assessment* (pp. 3–32). Springer Netherlands. https://doi.org/10.1007/978-94-007-2543-0_1
- Zhao, C., Zhao, S., Zhao, M., Chen, Z., Gao, C.-Z., Li, H., & Yu-an, T. (2018). Secure Multi-Party Computation: Theory, Practice and Applications. *Information Sciences*, 476. <https://doi.org/10.1016/j.ins.2018.10.024>
- Zuboff, S. (2019). *The Age of Surveillance Capitalism*. Public Affairs.

Summary

This dissertation consists of five chapters, each written as independent research papers that are unified by an overarching concern regarding information privacy and machine learning-based artificial intelligence (AI). This dissertation addresses the issues concerning privacy and AI by responding to the following three main research questions (RQs): RQ1. ‘How does an AI system affect privacy?’; RQ2. ‘How effectively does the General Data Protection Regulation (GDPR) assess and address privacy issues concerning both individuals and groups?’; and RQ3. ‘How can the value of privacy be embedded into systems?’

To respond to the RQs, this dissertation adopts the privacy impact assessment (PIA) as the overall methodology. A PIA encompasses three distinct stages; the first, the analytical stage, concerns the analysis of how AI (particularly focusing on inference as a process that includes inferred information, AI models’ performance, and accessing information uncovered by AI models) impacts privacy. Second, the legal assessment stage concerns whether AI that processes personal information and develops models complies with the GDPR. Finally, the design requirements stage features proposals for design requirements for systems aimed at protecting privacy. Accordingly, this dissertation is structured in three parts, each corresponding to a specific stage of a PIA and responding to one of the RQs.

Part I, which addresses the first stage of the PIA, comprises three chapters that altogether respond to RQ1. Chapter 2 analyses how AI impacts the descriptive aspect of privacy; this part argues that AI challenges the current definitions of privacy and that the ‘source control’ and ‘actual access’ definitions of privacy, once revised in the face of counter-examples involving inferred information, converge. Chapter 3, which considers how AI impacts the normative aspect of privacy, particularly the value of privacy, argues that AI affects the social value of privacy, which depends on trust, as this dimension of privacy is constituted when AI models perform accurately. Chapter 4 examines how AI impacts the normative aspect of privacy, particularly the right to privacy, and it argues that, although accessing information uncovered by AI models raises concerns about the privacy of algorithmically designed groups, the right to privacy cannot be recognised for such groups. This is the disruptive feature of AI that has led to consideration of new approaches other than the traditional one, which involves recognising the right to privacy, to protect the privacy of these groups. Instead of recognising the right to privacy for algorithmically designed groups, this chapter

suggests taking a moral principle for the moral obligation of protecting vulnerable groups within an ethics of vulnerability.

Part II concerns the second stage of the PIA and consists of one chapter that responds to RQ2. Chapter 5, in addition to evaluating whether AI that processes personal information and develops models complies with the GDPR, also assesses whether the GDPR adequately addresses the privacy issues raised by AI. It specifically focuses on group privacy and argues that GDPR has limitations in protecting the privacy of algorithmically designed groups and that the privacy of such vulnerable entities must be considered in the context of privacy and data protection.

Part III, which is related to the third stage of the PIA, also consists of one chapter that responds to RQ3. Chapter 6 proposes design requirements to protect privacy by integrating privacy into systems and argues that privacy is instrumentally valuable for the sake of autonomy. Accordingly, to embed the value of privacy into systems, design requirements are articulated through the translation of norms that promote and protect autonomy.

Samenvatting

Dit proefschrift bestaat uit vijf hoofdstukken, elk geschreven als onafhankelijke wetenschappelijke artikelen die een centrale zorg delen over informatieprivacy en op machine learning gebaseerde kunstmatige intelligentie (AI). Dit proefschrift behandelt de kwesties rond privacy en AI door antwoord te geven op de volgende drie onderzoeksvragen (RQ's): RQ1. 'Hoe beïnvloedt een AI-systeem de privacy?'; RQ2. 'Hoe effectief beoordeelt en adresseert de Algemene Verordening Gegevensbescherming (AVG) privacykwesties die zowel individuen als groepen aangaan?'; en RQ3. 'Hoe kan de waarde van privacy in systemen worden ingebed?'

Om in te gaan op de onderzoeksvragen, hanteert dit proefschrift de Privacy Impact Assessment (PIA) als algemene methodologie. Een PIA omvat drie verschillende fasen: de eerste, de analytische fase, betreft de analyse van hoe AI (met name gericht op inferentie als een proces dat afgeleide informatie behelst, de prestaties van AI-modellen en het verkrijgen van toegang tot informatie die door AI-modellen wordt ontdekt) de privacy beïnvloedt. Ten tweede gaat het in de juridische beoordelingsfase om de vraag of AI die persoonlijke informatie verwerkt en modellen ontwikkelt, voldoet aan de AVG. De fase van ontwerpvereisten bevat ten slotte voorstellen voor ontwerpvereisten voor systemen gericht op het beschermen van de privacy. Dienovereenkomstig is dit proefschrift opgebouwd uit drie delen, die elk overeenkomen met een specifieke fase van een PIA en antwoord geven op een van de RQ's.

Deel I behandelt de eerste fase van de PIA en bestaat uit drie hoofdstukken die samen een antwoord geven op RQ1. Hoofdstuk 2 analyseert hoe AI het beschrijvende aspect van privacy beïnvloedt; in dit deel wordt betoogd dat AI de huidige definities van privacy ter discussie stelt en dat de definities van 'broncontrole' en 'daadwerkelijke toegang' van privacy, wanneer herzien in het licht van tegenvoorbeelden met betrekking tot afgeleide informatie, in elkaar overgaan. Hoofdstuk 3, waarin wordt bekeken hoe AI het normatieve aspect van privacy beïnvloedt, en dan met name de waarde van privacy, betoogt dat AI de sociale waarde van privacy beïnvloedt. De sociale waarde hangt af van vertrouwen, aangezien deze dimensie van privacy tot stand komt wanneer AI-modellen accuraat presteren. Hoofdstuk 4 onderzoekt hoe AI het normatieve aspect van privacy beïnvloedt, met name het recht op privacy, en beargumenteert dat, hoewel toegang tot informatie die door AI-modellen wordt ontdekt zorgen oproept over de privacy van algoritmisch ontworpen groepen, het

recht op privacy voor dergelijke groepen niet erkend kan worden. Dit is het ontwrichtende kenmerk van AI dat heeft geleid tot het overwegen van nieuwe benaderingen, anders dan de traditionele, waarbij het recht op privacy wordt erkend om de privacy van deze groepen te beschermen. In plaats van het recht op privacy te erkennen voor algoritmisch ontworpen groepen, suggereert dit hoofdstuk om een moreel principe aan te nemen voor de verplichting om kwetsbare groepen te beschermen binnen een ethiek van kwetsbaarheid.

Deel II betreft de tweede fase van de PIA en bestaat uit één hoofdstuk dat inspeelt op RQ2. Hoofdstuk 5 beoordeelt niet alleen of AI die persoonlijke informatie verwerkt en modellen ontwikkelt, voldoet aan de AVG, maar ook of de AVG adequaat tegemoetkomt aan de privacykwesties die AI met zich meebrengt. Het richt zich specifiek op groepsprivacy en stelt dat de AVG beperkingen heeft bij het beschermen van de privacy van algoritmisch ontworpen groepen en dat de privacy van dergelijke kwetsbare entiteiten moet worden beschouwd in de context van privacy en gegevensbescherming.

Deel III, gerelateerd aan de derde fase van de PIA, bestaat ook uit één hoofdstuk dat inspeelt op RQ3. Hoofdstuk 6 stelt ontwerpvereisten voor om privacy te beschermen door deze in systemen te integreren en stelt dat privacy instrumenteel waardevol is omwille van de autonomie. Om de waarde van privacy in systemen te verankeren, worden ontwerpvereisten geformuleerd door normen te vertalen die autonomie bevorderen en beschermen.

List of Publications

Peer-Reviewed Journal Articles

- Asgarina, H. (2023). Convergence of the source control and actual access accounts of privacy. *AI and Ethics*. <https://doi.org/10.1007/s43681-023-00270-z>
- Asgarina, H. (2024). Adopting trust as an ex post approach to privacy. *AI and Ethics*, 3(4). <https://doi.org/10.1007/s43681-024-00421-w>
- Asgarina, H. (2024). Design for Embedding the Value of Privacy in Personal Information Management Systems. *Journal of Ethics and Emerging Technologies*, 33(1), Article 1. <https://doi.org/10.55613/jeet.v33i1.129>
- Asgarina, H. (2024). Limiting Access to Certain Anonymous Information: From the Group Right to Privacy to the Principle of Protecting the Vulnerable. *The Journal of Value Inquiry*. <https://doi.org/10.1007/s10790-024-09980-x>
- Asgarina, H., Chomczyk Penedo, A., Esteves, B., & Lewis, D. (2023). “Who Should I Trust with My Data?” Ethical and Legal Challenges for Innovation in New Decentralized Data Management Technologies. *Information*, 14(7)(351). <https://doi.org/10.3390/info14070351>

Book Chapters

- Asgarina, H. (2023). Big Data as Tracking Technology and Problems of the Group and its Members. In K. Macnish & A. Henschke (Eds.), *The Ethics of Surveillance in Times of Emergency*. Oxford University Press. <https://doi.org/10.1093/oso/9780192864918.003.0005>
- Macnish, K., & Asgarinia, H. (2023). Privacy and the Media. In C. Fox & J. Saunders (Eds.), *The Routledge Handbook of Philosophy and Media Ethics (Routledge Handbooks in Applied Ethics)*. Routledge. <https://doi.org/10.4324/9781003134749>

Other Publications

- Asgarina, H. (2020). The role of privacy impact assessments in shaping privacy-protective technical solutions. *Conference on Responsible Use of Technology and Health Data*. <https://fpf.org/wp-content/uploads/2020/10/40-Haleh-Asgarina-Privacy-Pandemic-Conference.pdf>

- Asgarinia, H. (2023). Ex-post Approaches to Privacy: Trust Norms to Realize the Social Dimension of Privacy. *International Conference on Computer Ethics, 1*(1), Article 1. <https://soremo.library.iit.edu/index.php/CEPE2023/article/view/287>
- Asgarinia, H., & Macnish, K. (2022). *Trustworthy Digital Twins in Intelligent Transport Systems*. Sopra-Steria, London. https://www.soprasteria.co.uk/docs/librariesprovider2/sopra-steria-uk-documents/thought-leadership/trustworthy-digital-twins-in-intelligent-transport-systems.pdf?sfvrsn=93f8d6dc_1
- Esteves, B., Asgarinia, H., Penedo, A. C., Mutiro, B., & Lewis, D. (2022). Fostering trust with transparency in the data economy era: An integrated ethical, legal, and knowledge engineering approach. *Proceedings of the 1st International Workshop on Data Economy*, 57–63. <https://doi.org/10.1145/3565011.3569061>

The Simon Stevin Series in Ethics of Technology is an initiative of the 4TU Centre for Ethics and Technology. 4TU.Ethics is a collaboration between Delft University of Technology, Eindhoven University of Technology, University of Twente, and Wageningen University & Research. Contact: info@ethicsandtechnology.eu

Books and Dissertations

Volume 1: Lotte Asveld, *'Respect for Autonomy and Technology Risks'*, 2008

Volume 2: Mechteld-Hanna Derksen, *'Engineering Flesh, Towards Professional Responsibility for 'Lived Bodies' in Tissue Engineering'*, 2008

Volume 3: Govert Valkenburg, *'Politics by All Means. An Enquiry into Technological Liberalism'*, 2009

Volume 4: Noëmi Manders-Huits, *'Designing for Moral Identity in Information Technology'*, 2010

Volume 5: Behnam Taebi, *'Nuclear Power and Justice between Generations. A Moral Analysis of Fuel Cycles'*, 2010

Volume 6: Daan Schuurbiens, *'Social Responsibility in Research Practice. Engaging Applied Scientists with the Socio-Ethical Context of their Work'*, 2010

Volume 7: Neelke Doorn, *'Moral Responsibility in R&D Networks. A Procedural Approach to Distributing Responsibilities'*, 2011

Volume 8: Ilse Oosterlaken, *'Taking a Capability Approach to Technology and Its Design. A Philosophical Exploration'*, 2013

Volume 9: Christine van Burken, *'Moral Decision Making in Network Enabled Operations'*, 2014

Volume 10: Faridun F. Sattarov, *'Technology and Power in a Globalising World, A Political Philosophical Analysis'*, 2015

Volume 11: Gwendolyn Bax, *'Safety in large-scale Socio-technological systems. Insights gained from a series of military system studies'*, 2016

Volume 12: Zoë Houda Robaey, *'Seeding Moral Responsibility in Ownership. How to Deal with Uncertain Risks of GMOs'*, 2016

Volume 13: Shannon Lydia Spruit, *'Managing the uncertain risks of nanoparticles. Aligning responsibility and relationships'*, 2017

- Volume 14: Jan Peter Bergen, *Reflections on the Reversibility of Nuclear Energy Technologies*, 2017
- Volume 15: Jilles Smids, *Persuasive Technology, Allocation of Control, and Mobility: An Ethical Analysis*, 2018
- Volume 16: Taylor William Stone, *Designing for Darkness: Urban Nighttime Lighting and Environmental Values*, 2019
- Volume 17: Cornelis Antonie Zweistra, *Closing the Empathy Gap: Technology, Ethics, and the Other*, 2019
- Volume 18: Ching Hung, *Design for Green: Ethics and Politics for Behavior-Steering Technology*, 2019
- Volume 19: Marjolein Lanzing, *The Transparent Self: a Normative Investigation of Changing Selves and Relationships in the Age of the Quantified Self*, 2019
- Volume 20: Koen Bruynseels, *Responsible Innovation in Data-Driven Biotechnology*, 2021
- Volume 21: Naomi Jacobs, *Values and Capabilities: Ethics by Design for Vulnerable People*, 2021
- Volume 22: Melis Baş, *Technological Mediation of Politics. An Arendtian Critique of Political Philosophy of Technology*, 2022
- Volume 23: Mandi Astola, *Collective Virtues. A Response to Mandevillian Morality*, 2022
- Volume 24: Karolina Kudlek, *The Ethical Analysis of Moral Bioenhancement. Theoretical and Normative Perspectives*, 2022
- Volume 25: Chirag Arora, *Responsibilities in a Datafied Health Environment*, 2022
- Volume 26: Agata Gurzawska, *Responsible Innovation in Business. A Framework and Strategic Proposal*, 2023
- Volume 27: Rosalie Anne Waelen, *The Power of Computer Vision. A Critical Analysis*, 2023
- Volume 28: José Carlos Cañizares Gaztelu, *Normativity and Justice in Resilience Strategies*, 2023
- Volume 29: Martijn Wiarda, *Responsible Innovation for Wicked Societal Challenges: An Exploration of Strengths and Limitations*, 2023
- Volume 30: Leon Walter Sebastian Rossmailer, *mHealth Apps and Structural Injustice*, 2024

Volume 31: Haleh Asgarinia, *Privacy and Machine Learning-Based Artificial Intelligence: Philosophical, Legal, and Technical Investigations*, 2024

Simon Stevin (1548-1620)

‘Wonder en is gheen Wonder’

This series in the philosophy and ethics of technology is named after the Dutch / Flemish natural philosopher, scientist and engineer Simon Stevin. He was an extraordinary versatile person. He published, among other things, on arithmetic, accounting, geometry, mechanics, hydrostatics, astronomy, theory of measurement, civil engineering, the theory of music, and civil citizenship. He wrote the very first treatise on logic in Dutch, which he considered to be a superior language for scientific purposes. The relation between theory and practice is a main topic in his work. In addition to his theoretical publications, he held a large number of patents, and was actively involved as an engineer in the building of windmills, harbours, and fortifications for the Dutch prince Maurits. He is famous for having constructed large sailing carriages.

Little is known about his personal life. He was probably born in 1548 in Bruges (Flanders) and went to Leiden in 1581, where he took up his studies at the university two years later. His work was published between 1581 and 1617. He was an early defender of the Copernican worldview, which did not make him popular in religious circles. He died in 1620, but the exact date and the place of his burial are unknown. Philosophically he was a pragmatic rationalist for whom every phenomenon, however mysterious, ultimately had a scientific explanation. Hence his dictum ‘Wonder is no Wonder’, which he used on the cover of several of his own books.

This dissertation explores how machine learning-based artificial intelligence (ML-based AI) impacts information privacy, particularly analysing how inference as a process associated with ML affects information privacy. Furthermore, this research highlights the limitations of the General Data Protection Regulation (GDPR) in addressing issues concerning inference, and suggests design requirements to embed the value of privacy into systems.

In its philosophical investigation, this dissertation distinguishes between various components and activities related to inference, including inferred information, AI models' performance, and accessing anonymous information uncovered by ML models. Two aspects of privacy are considered: the descriptive, which pertains to its definition, and the normative, which relates to its value and the right to privacy. The investigation explores how inferred information affects the definition of privacy, the influence of AI models' performance on the social value of privacy, and the implications of accessing information uncovered by ML models for group privacy, more precisely the group right to privacy.

In its legal investigation, this dissertation examines the GDPR's effectiveness in addressing privacy issues related to information inferred about or ascribed to a person as a member of a group, as well as information derived from inference about a group as a whole.

In its technical investigation, this research proposes design requirements to embed the social value of privacy into systems. It develops a value hierarchy for privacy in which the highest layer examines the relationships between privacy and social autonomy, the middle layer identifies norms regarding promoting or protecting social autonomy, and the lowest layer translates those norms into design requirements.

‘Wonder en is gheen wonder’

4TU ● Centre for Ethics
and Technology