Alma Mater Studiorum - Università di Bologna in cotutela con Katholieke Universiteit Leuven e Mykolas Romeris University

DOTTORATO DI RICERCA IN
LAW, SCIENCE AND TECHNOLOGY

Ciclo 35

Settore Concorsuale: 12/H3 –Filosofia Del Diritto
Settore Scientifico Disciplinare: IUS/20 – Filosofia Del Diritto

INFLUENCEABLE AUTONOMY AND PREDICTABLE FREEDOM IN THE IOE

Presentata da: Maximilian Gartner

Coordinatore Dottorato
Prof. Monica Palmirani

Supervisore
Prof. Giovanni Sartor

Supervisore
Prof. Mindaugas Kiškis

Supervisore
Prof. Anton Vedder

Esame finale anno 2023

FACULTY OF LAW AND CRIMINOLOGY.
Tiensestraat 41 box 3413
B-3000 LEUVEN, BELGIUM
tel. + 32 16 32 00800
maximilian.gartner@kuleuven.be
https://www.law.kuleuven.be/citip/en/staff-members/staff/00135645

Maximilian Gartner

INFLUENCEABLE AUTONOMY AND PREDICTABLE FREEDOM IN THE IOETITLE CONTINUED

February 2023

KU LEUVEN

FACULTY OF LAW AND CRIMINOLOGY

INFLUENCEABLE AUTONOMY AND PREDICTABLE FREEDOM IN THE IOE

Maximilian Gartner

Supervisor: Prof. Giovani Sartor
Supervisor: Prof. Mindaugas Kiškis
Supervisor: Prof. Anton Vedder

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor of Laws (PhD).

February 2023

This thesis was submitted on 15 February 2023.

Pages: 254
Words (without footnotes): 101949

Sources are referenced in accordance with the 3rd edition of the Modern Humanities Research Association citation style.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ITN EJD grant agreement No 814177.

Abstract

The present thesis investigates how individuals can develop, exercise, and maintain autonomy and freedom in the presence of information technology. It is particularly interested in how information technology can impose autonomy constraints.

There are three distinct parts within this thesis. The first part identifies a problem with current autonomy discourse: There is no agreed upon object of reference when bemoaning loss of or risk to an individual's autonomy. Believing this to be a detriment both to engage in useful discourse and to develop appropriate countermeasures against autonomy constraints that may be deemed inappropriate or unacceptable, the thesis introduces a pragmatic conceptual framework to classify autonomy constraints. This framework is informed by philosophical theory, privacy studies and the reflections of the notion of self-government in the legal domain, all of which are surveyed. In essence, the proposed framework divides autonomy in three categories: intrinsic autonomy, relational autonomy and informational autonomy.

The second part of the thesis investigates the role of information technology in enabling and facilitating autonomy constraints. The analysis identifies eleven characteristics of information technology, as it is embedded in society, so-called vectors of influence, that constitute risk to an individual's autonomy in a substantial way. These vectors are assigned to three sets that correspond to the general sphere of the information transfer process to which they can be attributed to, namely domain-specific vectors, agent-specific vectors and information recipient-specific vectors.

The third part of the thesis investigates selected ethical and legal implications of autonomy constraints imposed by information technology. It shows the utility of the theoretical frameworks introduced earlier in the thesis, that is the pragmatic account of autonomy and the concept of vectors of influence, when conducting an ethical analysis of autonomy-constraining technology. It also traces the concept of autonomy in the European Data Laws, showing that the European regulator is becoming more aggressive in ensuring autonomy protection within the digital domain. Finally, the thesis investigates the impact of cultural embeddings of individuals on efforts to safeguard autonomy and shows intercultural flashpoints of autonomy differences. Further analysis suggests the importance of timing and speed of technology introduction as well as digital literacy when it comes to incurring autonomy risks through intercultural technology deployment and suggests strategies to mitigate these risks.

In view of this, the thesis approaches the exercise and constraint of autonomy in presence of information technology systems holistically. It contributes to establish a common understanding of (intuitive) terminology and concepts within the scope of inquiry, connects this to current phenomena arising out of ever-increasing interconnectivity and computational power and helps operationalize the protection of autonomy through application of the proposed frameworks.

Acknowledgements

Writing a dissertation is already a strange experience. Writing a dissertation at the precipice between the domains of law, philosophy and technology is stranger still. Writing a dissertation during a pandemic may very well be the strangest of all. In navigating this journey, I was not alone.

Many people shaped this text. First are my three supervisors, Prof. Giovanni Sartor, Prof. Mindaugas Kiškis and Prof. Anton Vedder. They have provided insight where I did not know anything, guidance where I did not know enough and freedom where I knew what to do. I am grateful for their guidance and support.

I had the privilege of pursuing my doctoral degree as part of the Joint Doctorate in Law, Science and Technology cohort. My colleagues have proven to be an endless resource of expertise and companionship for which I am grateful. I am particularly indebted to Yannick Vogel and Pier Giorgio Chiara. Discussing legal theories, exchanging reading recommendations and laughing about the insanity of entering the world of academia during a global lockdown have made the last three years infinitely more enjoyable.

My time as a young researcher was generously supported under the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ITN EJD grant agreement No 814177. Studying under a MSCA grant is a privilege that has allowed me to travel throughout Europe to connect with my fellow colleagues and focus on my research full-time. Without this funding, there would be no research, no dissertation, no doctoral degree. I am grateful for having received this support. At this point, gratitude is owed also to the programme coordinator Monica Palmirani, who was the driving force behind this constellation.

Between the pressure of life as a researcher and the difficulties of life as a human, I have been lucky to draw on the support of my close friends and family. When borders were closed and people were sick, deadlines were moved and instructions unclear; they helped me stay sane. This was nice of them, and for that I am grateful.

Most of this text was written by me. But sometimes one of my two cats walked over the keyboard. This has prompted me to reconsider my text more than once. While I cannot recommend feline editing for everyone, I believe they made this thesis better in their own way.

Pursuing a doctoral degree does not an agreeable man make. As I oscillated between triumph and tribulation on every page, embedded in unpredictable assignments and last-minute travel arrangements, I could count on my wife Reagan Aylmer to keep me grounded and pointed towards the finish line.

Table of Contents

A Note on Scientific Integrity

This thesis has been the product of three years of research. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ITN EJD grant agreement No 814177. In line with the requirements of the applicable grant agreement, and with the goal of disseminating my research, some of the contents of this thesis have been previously published and presented in the following venues:

- An early outline of the thesis has been presented at the 7th Doctoral Consortium of the 32nd International Conference on Legal Knowledge and Information Systems (JURIX 2019) and subsequently published in its respective proceedings.[1]
- An abridged early draft of Sections 1-5 has been presented at the KU Leuven's 8th Assistentenconferentie (ACCA 2020) and subsequently published in the book Technology and Society: The Evolution of the Legal Landscape.[2]
- Exploratory research that connects to Section 13 has been presented at the Harvard Carr Center for Human Rights' Intercultural Digital Ethics Symposion 2021.
- Parts of Section 6-10 have been presented at the Oxford Internet Institute's Second Multidisciplinary International Symposium (MISDOOM 2021).
- Parts of Section 6-10 have been presented at the International Cartographic Association's International Cartographic Conference (ICC 2021) and published in its respective proceedings.[3]
- Parts of Section 12 have been published in the European Data Protection Law Review.[4]

Additional parts of this thesis are currently under review in form of articles and essays at the time of submission.

[1] Maximilian Gartner, 'Research Proposal: Influenceable Autonomy and Predictable Freedom in the IoE', in *Proceedings of the Seventh JURIX 2019 Doctoral Consortium Co-Located with 32nd International Conference on Legal Knowledge and Information Systems (JURIX 2019)*, ed. by Monica Palmirani (CEUR Workshop Proceedings, 2018).

[2] Maximilian Gartner, 'Fit for the Future: A Pragmatic Account of Human Autonomy to Understand Emerging Issues in The Internet of Everything', in *Technology and Society: The Evolution of the Legal Landscape*, ed. by Marie Bourguignon and others (Antwerpen: Gompel&Svacina, 2021).

[3] Maximilian Gartner, 'LBS as Vectors of Influence', in *Proceedings of the ICA*, 2021.

[4] M. Gartner, 'Regulatory Acknowledgment of Individual Autonomy in European Digital Legislation: From Meta-Principle to Explicit Protection in the Data Act', *European Data Protection Law Review*, 8.4 (2023), 462–73 <https://doi.org/10.21552/edpl/2022/4/6>.

1 Introduction

1.1 General Remarks on Purpose

"Human autonomy is under threat, and it is all due to the Internet" seems to be an increasingly popular warning of actors in the political and academic sector alike. At first sight, this seems like a plausible statement. The Internet as the compound connections between otherwise separated computer networks has enjoyed a steady rise from a specialized tool to an undisputed general-purpose technology.[5] The increasingly widespread adoption is two-pronged; not only are the sheer number of similarly connected devices increasing, but more types of devices are outfitted with connection capabilities.[6] This phenomenon has initially been dubbed the "Internet of Things" due to the focus on outfitting sensors to physical devices (i.e. "things"). However other terminology has been suggested more recently as the world moves ever closer towards ubiquitous computing (i.e. omnipresent data processing)[7] due to the sheer amount of data and ever-increasing level of connectivity. One of these new terms coined is the Internet of Everything, defined for example by CISCO as a network that "is bringing together people, process, data, and things to make networked connections more relevant and valuable than ever before-turning information into actions that create new capabilities, richer experiences, and unprecedented economic opportunity for businesses, individuals, and countries."[8] Generally, one may see the Internet of Everything as an umbrella term encompassing the Internet of Things as well as the phenomenon of the Internet of Health, Internet of People, Internet of Money and Internet of Data, or just the state of the internet necessary for ubiquitous computing.[9]

With this increasing interconnectivity, many are quick to point out alleged risks that are caused by emerging technologies. One reoccurring theme to this respect is the conceived threat of ubiquitous computing, or more specifically certain of its computing processes (hereafter called informational agents), to human autonomy. Informational agents are identified in the literature as a process that can "undermine consumers' sense of autonomy",[10] and "contribute to self-fulfilling prophecies and stigmatisation in targeted groups, undermining their autonomy and participation in society. […] Value-laden decisions made by algorithms can also pose a threat to the autonomy of data subjects."[11] According to the United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression their opacity "risks interfering with individual self-determination, or what is referred to […] as "individual

[5] See for a more rigorous investigation and the relevance of this statement for the larger inquiry at hand in particular Section 8.
[6] John Naughton, 'The Evolution of the Internet: From Military Experiment to General Purpose Technology', *Journal of Cyber Policy*, 1.1 (2016), 5–28 <https://doi.org/10.1080/23738871.2016.1157619>.
[7] In the meaning of Mark Weiser, 'The Computer for the 21 St Century', *ACM SIGMOBILE Mobile Computing and Communications Review*, 3.3 (1999), 3–11 <https://doi.org/10.1145/329124.329126>.
[8] CISCO, *The Internet of Everything - Global Public Sector Economic Analysis*, 2013.
[9] This grouping is adopted from the diction of the research programme that has spawned this thesis. See for other research in this area including other dissertations other output that is connected to the Marie Skłodowska-Curie ITN EJD grant agreement No 814177.
[10] Quentin André and others, 'Consumer Choice and Autonomy in the Age of Artificial Intelligence and Big Data', *Customer Needs and Solutions*, 5.1–2 (2018), 28–37 <https://doi.org/10.1007/s40547-017-0085-8>; Eliza Mik, 'The Erosion of Autonomy in Online Consumer Transactions', *Law, Innovation and Technology*, 8.1 (2016), 1–38 <https://doi.org/10.1080/17579961.2016.1161893>.
[11] Brent Daniel Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate', *Big Data and Society*, 3.2 (2016) <https://doi.org/10.1177/2053951716679679>.

autonomy and agency.[12] Similarly, the Committee of Ministers of the Council of Europe warns that "fine grained, sub-conscious and personalised levels of algorithmic persuasion may have significant effects on the cognitive autonomy of individuals and their right to form opinions and take independent decisions."[13] In the more accessible popular literature, Zuboff characterises informational agents with the phrase "what is to be killed here is the inner impulse toward autonomy and the arduous, exciting elaboration of the autonomous self as a source of moral judgment".[14] These warnings, while seeming justified from an intuitive standpoint, share a common opacity. What does it mean to be autonomous in this context? What does it mean to be constrained by informational agents? The term autonomy, after all the alleged victim of the informational agents is never truly established. Throughout both popular and academic literature dealing with emerging technology, the true nature of what exactly commentators regard as the autonomy under threat remains undefined. Believing this opacity to be a detriment to both research and practical applications, this thesis aims to address the issue at hand and investigates the interconnected concepts of human freedom[15] and autonomy, agency and self-determination as they are aided and undermined by characteristics of and agents within the Internet of Everything.

1.2 General Remarks on Research Questions and Methodology

This inquiry aims to investigate the intersection of individual and autonomy by approaching the question of what freedom and autonomy an individual can develop and maintain in presence of systems of the IoE. As this target is very broad, this thesis inverses the question, and focuses particularly on what type of obstacles there are to the development and maintaining of autonomy in this context. Research is conducted by casting the light on the following main interconnected research questions and sub-research questions:

- Q1: How can autonomy be conceptualized to be relevant and salient for investigating emerging issues of the IoE?
    - SQ1.1 Is there a current consensus on how to conceptualize autonomy in general, and in the domain of technology specifically?
    - SQ1.2 How are the concepts of privacy and autonomy connected?
    - SQ1.3 How are autonomy concepts already reflected in the legal domains generally?
    - SQ1.4: What are the main obstacles to conceptualize autonomy generally and in this context?
- Q2: What role does the information technology have in enabling and facilitating autonomy constraints?

[12] Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, *Seventy-Third Session Item 74 (b) of the Provisional Agenda** Promotion and Protection of Human Rights: Human Rights Questions, Including Alternative Approaches for Improving the Effective Enjoyment of Human Rights and Fundamental Freedoms*, 2018.

[13] Council of Europe, *Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes*, 2019.

[14] Shoshana Zuboff, *The Age of Surveillance Capitalism*, 2019 <https://doi.org/10.1017/CBO9781107415324.004>.

[15] Note that for the majority of this text the term autonomy is used to denote the set of autonomy and freedom as used colloquially together. Section 5 introduces an expansive conception of autonomy that covers the concepts behind most uses of the term freedom as well. The tension between these two terms is also explored in Section 2.4 . The use of freedom in the title of this dissertation is partly owed to administrative hurdles imposed by the underlying funding mechanism of this research making title-changes difficult, but the avid reader will agree that the present analysis can be fairly considered to encompass "predictable freedom" and "influenceable autonomy" alike.

─ SQ2.1: How does the existing field of research on persuasive technology connect to this inquiry?
─ SQ2.2: What are the characteristics of the IoE that are potentially relevant for understanding the risk of autonomy constraints?
─ SQ2.3: What are the characteristics of processes within the IoE (i.e. informational agents) that are potentially relevant for understanding the risk of autonomy constraints?
─ SQ2.4: What are typical patterns and phenomena that characterize risk to an individual's autonomy in the context of information receival?

-Q3: What are some of the ethical and legal boundaries pertaining to autonomy constraints imposed by informational agents?

─ SQ3.1: How and under consideration of which factors should an ethical analysis of technology-imposed autonomy constraints be conducted and to what extent is the pragmatic account of autonomy useful in this context?
─ SQ3.2: How is the concept of autonomy as exercised and constraint in presence of technology reflected in existing and upcoming European legislative instruments
─ SQ3.3: What is the impact of cultural embeddings of individuals on efforts to safeguard autonomy and how does this affect intercultural deployment of technology?

The text of the thesis follows the order of this research questions, with each chapter typically addressing one or two of the sub-questions.

The nature of the research proposed is interdisciplinary, consequently the methods undertaken will be heterogeneous. An overview is given in the following paragraphs.

With respect to questions of law and policy, the research methodology will be dictated by a (qualitative) black-letter approach to analyse the statutes and jurisprudence of relevant jurisdictions. Due to language abilities, focus will be laid on jurisdictions where legal sources or their elaborations are available in English or German. Where appropriate, comparative analysis of the legal situations within the jurisdictions in question will be conducted. Resources used will be both primary (i.e. laws, court cases) and secondary, where available. Similarly to the outlined above, instruments of soft law or similar (guidelines, reports, etc.) will be referenced and analysed.

Moral and ethical frameworks with respect to the area of inquiry will be analysed, interpreted and compared. A later section will also outline some boundaries of a potential ethical framework with respect to technology-imposed autonomy constraints. The research method for this is qualitative and descriptive.

Extrapolation from current scholarship with respect to agents deployed in the IoE and their potential to impact private autonomy and algorithmic influence is attempted. Starting from a consolidation of fundamental principles of function, uses and limitations of such agent, their potential impact will be assessed based on their technological capabilities; areas of risk with respect to autonomy / freedom originating from the underlying technical approach of such agents will be determined and highlighted. Agents will be assessed abstractly and in context; no model building will be attempted. The research method for these sections is qualitative and mostly descriptive.

Current findings in the field of the effects of algorithmic manipulations will be consolidated, compared and interpreted. Similarly, findings in the field of private autonomy and personal freedom will be consolidated and reviewed for their applicability to the topic of interest as laid

out above. Out of this research, a working model of human autonomy will be synthesized to be used in the dissertation, as to avoid ambiguities when elaborating on the subject. The research method in this section is qualitative and prescriptive.

1.3 Whose Autonomy is it Anyway?

In this thesis, individual autonomy and freedom are discussed, which begs the question of exactly who the subject of interest, i.e. the individual, is. As this text will reference existing scholarship, it is important to keep in mind that much of the cited research has been targeted at specific groups of individuals. Each of these groups have their own characteristics in which autonomy constraint may be particularly visible, measurable, or comparatively easier to impose (or protect against). For example, much of the current online domain accessible to individuals has a transactional context and information recipients are thus classified as potential customers and consumers. Much research exists hence to investigate consumer's behaviour in online environments, and much legal legwork exists to shape these efforts. By analysing the impact of technology on their choices and contrasting them with their preferences and under consideration of the legal postulation of what ought to be an inviolable part of their autonomy (and privacy!), we may denote the technological as an autonomy risk.

But as this is a holistic inquiry, analysis in this thesis aims to provide a cross-section of this research without commitment to a specific sub-group. Indeed, the text is built on the assumption that these findings can be abstracted to a certain extent. This has advantages and disadvantages. The disadvantage is that the analysis findings may need to undergo further concretizations to be reapplied to certain scenarios. The advantage is that this analysis is less concerned by difficulties of labelling certain individuals and their interactions within their context, but only identifying them within the conceptual superstructure that is outlined in this text.[16]

Because of this approach, the subject of interest in this thesis is the individual. Autonomy investigated in this thesis is the autonomy of the individual in its broadest form.[17]

1.4 General Remarks on Structure

This thesis analyses autonomy constraints imposed by informational agents and what kind of legal and ethical principles can be derived from this. It is divided into three parts.

First, the thesis introduces a pragmatic conception of autonomy to capture the above concerns of autonomy constraints in an interconnected world. This will be done as follows: Section 2 and 3 survey the existing debate about human autonomy and the somewhat related concept of privacy. Section 4 analyses how autonomy is reflected in different legal phenomenon. Based on this information, the thesis then introduces a pragmatic account of autonomy in Section 5 and provides context as to how this theory can account for many of the intuitive concerns voiced in current debate.

[16] In the concrete example given above, this avoids difficult weighing of an individual's status between user, consumer, participant, supplier and any other status they might have in the context of e.g. consumer protection law or privacy theory when applied to the digital domain.

[17] This is not to say that there cannot be contextualization. To this end, the thesis aims to provide ample examples of how the concepts of the conceptual superstructure outlined in this text are both derived and can apply to practical and contextual examples.

The second part of the thesis analyses how technology can impart constraints on an individual's autonomy. As a primer on existing debate, Section 6 gives an overview over the related but not fully overlapping debate on so called persuasive technology to provide further context and situate the analysis in this thesis, as well as outline existing research and debate. Section 7 outlines two theoretical concepts, the informational pipeline and vectors of influence respectively, that will be relied upon in the rest of the thesis and serve as descriptors of the domain investigated. Following the research logic, Section 8 investigates the IoE-domain and highlights its physical and meta characteristics that are relevant to the analysis at hand. Section 9 provides formal definitions for the concept of informational agents and highlights the most important agent characteristics relevant to their potential of constraining an individual's autonomy. Finally, Section 10 focuses broadly on interaction characteristics from the position of the information recipient, analysing which compound phenomena of information transfer situations are indicative of risk to an individual's autonomy, and thus completing the analysis of the suggested vectors of influence.

The third part considers ethical and legal questions that connect to the concerns raised in this thesis. Section 11 considers some ethical implications of autonomy-constraining technology and how the pragmatic account of autonomy and the concept of vectors of influence fit into ethical analysis. Section 12 traces the regulatory acknowledgment of individual autonomy in European Digital Legislation. Section 13 highlights that when deploying informational agents and similar technology, safeguarding an individual's autonomy is highly dependent on the cultural sphere in which the individual is embedded.

Finally, Section 14 revisits the research questions and contraposes them with the thesis' findings and offers concluding remarks.

1.5 Sketching the Path Ahead: Law from Ethics from Morals from Ends

For the first part which serves as the basis for the latter analysis, i.e. when it comes to the first goal of establishing a pragmatic account of autonomy, this inquiry takes recourse to three sources of meaning which are the general scholastic consensus and relevant theories in the field of philosophy and epistemology, moral and ethical frameworks,[18] and legal frameworks. As these choices are not immediately obvious compared to some of the other analysis conducted in this thesis, this section aims to justify the path taken as well as to familiarise readers with a background leaning into the technical or legal sides into the more philosophical and ethical aspects of the research conducted. As a welcome side effect, this elaboration may also prove useful to crystallize the meaning of certain terms used in this thesis in cases in which the terminology can have multiple valid objects of reference.

The reasoning for the approach taken is best explained when working backwards from the concrete and factual to the abstract by highlighting the connections between these domains. Legal frameworks, in short laws, structure the relations between people and other entities and objects, (e.g. other people, animals, material and immaterial objects, rights, etc.) of the system

[18] The terms "ethics" and "morals" are not interchangeable. Ethics are value systems of a certain group. Morality is then a subset of such ethics, namely a value system dealing first and foremost with the notions of "right" and "wrong". See notion is undermined by the fact that the term ethics is "commonly used interchangeably with 'morality" to quote John Deigh, 'Ethics', *The Cambridge Dictionary of Philosophy* (Cambridge University Press, 2015).

(i.e. society, the world) individuals are (inter-)acting in. By imposing enforceable rules and non-enforceable but authoritative endorsements, the legislative process steers individual system-members' behaviour. Compliance is enforced through governmental sanctions or incentives. These legal boundaries are set out to satisfy certain societal functions that are deemed important based on or at least influenced by a set of values adopted by some part of the society by the legislative entity.[19] Not accounting for the ephemeral concept of natural law, the legal domain ends at this point.[20] In other words, under this understanding, law is an inherently neutral set of tool, that is structured along the lines of an underlying value system; consequently, to understand its purpose and enactment, an analytical, teleological approach with knowledge of this underlying system is helpful.

Normative, i.e. rule-giving, frameworks which have not been legitimized by a legislative process are (pseudo-) ethical frameworks.[21] Compliance can be achieved through non-governmental means such as social or cultural pressure from members of the group applying the ethical framework and this compliance can be as widespread as compliance with a legal rule. Ethical frameworks can exceed the scope of a legal frameworks (i.e. a behaviour is legal but unethical) or can fall short of it (i.e. a behaviour is illegal but ethical). In the first case, the ethical framework fulfils a guiding function between multiple legal behavioural options or might be used to reason about situations that are plainly not covered by law yet.[22] In the second case, the legal framework is often characterized to be unfair and/or faulty. In both cases, if the divergence between the law and the underlying ethics is stark enough, a societal pressure is likely to grow, potentially leading to changes in the legal framework. A different approach of conceptualizing this is the notion that within a democracy the legitimacy and content of a legal system originates from its people; because these people are not ethically blank prior to the establishment of a legal system, the legal system will inevitably be infused with the ethical system of and propagated by the people.[23] At the same times, ethical frameworks are not uniform. Ethical systems, such as cultural or religious frameworks, can contradict each other. The same is true for ethical systems in the narrow sense (that is explicitly ethical frameworks),

[19] Indeed, the social functions of law might be so closely interwoven with the underlying moral or political imperatives that they are of "no use to anyone who does not completely and exclusively endorse them"; see Joseph Raz, *The Authority of Law: Essays on Law and Morality* (Oxford: Oxford University Press, 1979), p.166.

[20] A legal exclusive positivist position is assumed here. More precisely, it is assumed that morality does not affect validity of legal frameworks while it is acknowledged that morality factually influences both the creation and implementation of legal frameworks. In contrast, an inclusive positivist position would allow for invalidation of a legal framework based on moral considerations while a natural law position would require alignment with some sort of moral standard in order to be considered valid. Cf. Matthew Kramer, *Where Law and Morality Meet* (Oxford: Oxford University Press, 2004), p.3f; Willy Moka-Mubelo, *Reconciling Law and Morality in Human Rights Discourse* (Cham: Springer, 2017), p.53f, 86f.

[21] The distinction between ethics, religion, social conventions, cultural practices, etc. is a question of definition. When accepting all of these value systems as "ethics", more weight is put on the underlying question of morality. This latter option is appealing because it coincides with wide usage of the term; for example there is little value in labelling professional ethic-standards as pseudo-ethical. Consequentially, one must allow that professional ethic-standards might imply immoral rules, potentially making them an "immoral ethic framework". This view is contested, (see instead of many just the rebuke in Richard Paul and Linda Elder, *The Miniature Guide to Understanding the Foundations of Ethical Reasoning*, 3rd edn (Critical Thinking, 2005) p.7f.) but will be utilized here.

[22] This may very well be the case for many of the instances or phenomena discussed within this thesis, as legislation tends to lag behind emerging technology and its implication. Later in this text, the creeping explicit recognition of individual autonomy as described within this text and more theoretical literature by European legislation is explored further, see Section 12.

[23] For government structures that are not democratic, this conceptualization falls obviously short; however some sort of societal pressure for adopting changes to the legal structure is plausible to arise in any case, even without the direct representation of the populace within the legislative branch.

which is evidenced not only by the fact of the sheer multitude and variance but also by the meta-ethical disputes about their general viability and validity.[24] In order to be effective in effectuating legal frameworks, an ethical framework must either be widely adopted or intellectually appealing, be it through internal logic, through offering a solution to a contemporary problem or through prohibitive cognitive costs for not abiding by it (e.g. deeply rooted beliefs, disavowal of which conflicts directly with the identity of the agent).[25] To assess the flow of imperatives between these systems, consideration needs to be given to some of the underlying values of the ethical approach propagated; the proposition being that certain values will give stronger (factual or moral) justification to the ethical systems that depend on them. In other words, it is proposed here that the (factual) persuasiveness of underlying values determines the (factual) adoption of ethical frameworks that build upon them.

Persuasion aiding the ethical system is rooted in the values of the system itself, specifically in values that determine if an action is "right" or "wrong". Such underlying systems of values and corresponding principles concerned with (absolute) rights or wrongs are moral frameworks. How exactly these values derive authority (and subsequently grant authority to ethical frameworks based upon them) is arguably one of the core questions of the scientific discipline of ethics. Explanation attempts are broadly of two camps: deontological/duty-based and teleological/end-based assessment of an action. With the former, an action may be right if it is compliant with rules.[26] With the latter, an action is right when it promotes the right end and it is the best action to promote said end.[27] While only with the latter, the question of right or wrong is fully subordinate to a question of ends, these overarching moral principles still impact the framework of the other approaches.[28]

---

[24] Difficult questions arise, when trying to assess the existence of morality by itself and moral facts, the capability to address moral facts by moral statements and if or to what extent such is accessible through human inquiry at all. Generally, the meta-validity of ethical systems can be considered mostly with respect to (1) metaphysical and (2) epistemological positions. When it comes to metaphysical inquiry, e.g. nihilism/error theory maintain that there are no moral facts on which to base an ethical theory on, while subjectivism proclaims any moral fact to be inherently subjective. Within an epistemological inquiry, non-cognitivism characterizes all ethical concepts all expressions of an emotional state rather than an object of knowledge and as such without objective ethical/moral value. Of course even when finding that moral facts on which to base an ethical system exist, there is no universal consensus neither on their content nor on how to find or approximate them. Cf. for all this Deigh; Marcus Singer, 'Moral Epistemology', *The Cambridge Dictionary of Philosophy* (Cambridge University Pres.

[25] This distinction should not be construed to hide that intellectually appealing ethical frameworks are naturally suited to be or become widespread. However, it is conceivable that certain ethical guidelines might be transplanted into legal frameworks before they become widespread, or even despite the fact that they deal with a subset of situations that is just not considered widely enough to enter the general ethical discourse at all.

[26] E.g. rules that have a self-evident character (i.e. Intuitionism) or supernatural endorsement (i.e. Divine Command Theory), legitimation through a (hypothetical) form of social cooperation that would lead to their adoption/finding, i.e. through the fairness of such a process (i.e. Contractarianism) or that they stem from pure reason (i.e. Formalism), see Deigh.

[27] Deigh.

[28] From a contractarian viewpoint, an action is right or wrong depending on its compliance with a ruleset that is or would have been constructed through a fair process of people living together in fellowship and treating each other as equals. However, in that "constitutional" drafting process, the rules agreed upon will likely reflect the ends (e.g. happiness or pleasure of the individual or the collective). From a divine command theory-viewpoint, not only are actions wrong that violate the rules of the authoritative supernatural entity, but the supernatural entity is in itself the origin of ends; it seems to follow that relevant ends are congruent with (at least compliance with) the will or aim of the supernatural entity. From the viewpoint of formalism, pure reason leads to moral principles that satisfy the condition of being a universal law; all this to ensure that the moral agents are all individually an end in itself. Similar connections can be made for other ethical approaches. Cf. Jean Hampton, 'Contractarianism', *The Cambridge Dictionary of Philosophy* (Cambridge University Press, 1999). Philip

The above outlined considerations paint the picture of a "normative ladder" of moral and factual influence, in which certain principles originate as cornerstones of moral systems, and colour the frameworks higher up the ladder that are based upon them at the end of which are supposedly value-neutral legal systems. Different rungs of this ladder stand in a relationship of interconnectivity to each other. Importantly, incongruence between different rungs creates a sort of tension. Non-compliance of a framework (e.g. law) with lower-level framework (e.g. cultural norms) can undermine the adoption of the higher-level framework. At the same time, the lower-level framework or at least its perception may also be subject to (factual) change, brought in part by a rigid higher-level framework.

This interconnected structure aids the methodology of this thesis, and its integrity should have now become clear. Personal freedom and autonomy are inarguably considered and protected by legal frameworks on national and international level. This protection is reflecting a corresponding call to do so by ethical frameworks. The ethical frameworks in question are factually relevant due to the persuasiveness of their underlying moral principles, which mandate that actions undermining autonomy and personal freedom are (in general) morally wrong and protecting and fostering autonomy are (in general) morally right. Depending on the approach, this moral framework is more or less directly originating from an end of which autonomy or freedom is a necessary or helpful requirement, feature or consequence. In any case, these considerations root in an intuitive understanding of autonomy as something important. Indeed, autonomy is considered widely as something intrinsically valuable. The validity of the approach outlined at the beginning of this section, if not verifiable, can therefore be made plausible by observing the prevalence of rules that pertain to autonomy and freedom in legal and ethical or moral frameworks.

1.6 The Duality of Autonomy Research and Its Relevance for Legal Studies

The first part of the thesis, laying the groundwork for the subsequent parts, deals with autonomy as an ambiguous term in need of clarification. To give an informed delimitation of autonomy, highlight the relevant context for this inquiry and allow for a more cohesive terminology, as well as to lay the theoretical groundwork for the pragmatic model of autonomy in Section 5, it is necessary to survey the existing domain of fundamental autonomy research.

Conceptualizing autonomy can serve at least two distinct goals. First, autonomy can be seen as a yet to be fully explained phenomenon. Any further investigations bring with it the potential for a better understanding of autonomy as a whole or within specific context such as the domain of the IoE. The fact that not all is known (or agreed upon) about autonomy serves as the justification for further inquiry, and this goal is satisfied if more knowledge is gained by the inquiry. In the case of rather ephemeral concepts (as is the case with autonomy), the requirement of more knowledge can be seen as fulfilled, if a newly devised theory either deepens the understanding of existing conceptual approaches or if it suggests a new conceptual approach which (1) is consistent in itself and with a range of prerequisite concepts that can be considered state-of-the-art, and (realistically) (2) adds some sort of utility to the existing scientific discussion. In other words, the scientific question inherently justifies its answer. One

Quinn, 'Divine Command Ethics', *The Cambridge Dictionary of Philosophy* (Cambridge University Press, 1999). Steven Kuhn, 'Formalism', *The Cambridge Dictionary of Philosophy* (Cambridge University Press, 1999).

may call the quest to clarify, consolidate or compare these theories for the purposes of further insights into the nature of autonomy the *primary* goal of autonomy research.

There is, however, a *secondary* goal that autonomy research can fulfil, which is to undergird discussion and remedies of issues with autonomy. Assuming that autonomy carries with itself a certain value, and that this value ought to be protected, it is consequential that certain legal or factual measures may be taken to safeguard autonomy. One may assess these measures by using the insights of a model of autonomy with respect to their scope and efficacy. At the same time, not utilizing a clear model of autonomy hinders the inception of sufficiently scoped or efficient safeguards. For example, as laid out in the introduction, many commentators perceive the increasingly widespread use and increasing capabilities of informational agents as a risk to one's individual autonomy.[29] Taking this seriously, it then seems only consequential to engage in the construction of safeguards against such intrusions. For example, if data collection, and micro-targeted advertising by a profiling algorithm interferes with our decision-making capabilities, (and those capabilities are valuable), some sort of regulation of these processes may be warranted. However this practical application of conceiving legal measures against potential novel breaches of our autonomy is hindered by the opacity of the term autonomy in two ways. Firstly, most commentators seem unable to closely delimit what the autonomy one ought to protect actually is. Instead, most commentators seem to rely on a more intuitive understanding, locating some kind of autonomy conflict in connection with informational agents. It follows that the risks themselves are not clearly defined by virtue of their affected targets being not clearly defined. Secondly, any assessment of regulation, such as through legal means, requires some sort of control of the efficacy of the safeguards established. Establishing normative boundaries, such as laws, to the infringement of an individual's autonomy cannot or can only in a limited capacity be found to be either effective or ineffective, if the target is not clearly defined.[30] For example, banning the collection, storage or use of certain types of personalized data may suffice *prima facie* to alleviate our concerns with respect to targeted advertising because we feel like the quality of such information is inherently harmful to our vague understanding of autonomy. However, closer inspection can reveal that profiling algorithms can have the same or similar impact by utilizing data that is not traced back to an individual but to a whole group of people. This different IoE-process therefore has the very same impact of our autonomy, a fact one can only take into account if the very same autonomy is defined clearly enough from the onset. In other words, while regulation attempts might be conducted, their assessment is hindered by ambiguity of their aim. In the worst case, ambiguous language is carried forward into legislation in a misguided attempt to allow more judicial discretion or legal flexibility. This goal is rendered impossible as the judge or respective authority, even before being able to flexibly account for a case with all its details, must conduct

[29] This of course is rarely based on a deep systemic analysis of autonomy, but more often based on some sort of intuitive understanding of the concept.

[30] This should not be confused with the use of wide or undefined terms or concepts within the instrument establishing the normative boundary (i.e. law). For example, the GDPR strives for technological neutrality, explicitly to avoid a protective scope that is too narrow to grant adequate protection and would be susceptible to circumvention measures, as stated in its Recital 15. Similarly, many norms of fundamental character, be it in human rights regimes or on a domestic constitutional level often use undefined, sweeping terminology. This may actually aid the protection of their intended target, as courts may use methods of interpretation to find applicability of the wide (and potentially malleable) source text to issues that warrant it, as is seen in the "living instrument"-approach the ECtHR takes with the ECHR. However, it remains necessary to identify such issues that warrant application of normative boundaries narrow and wide alike. To continue the first example, it is not necessary to define the concept of data protection explicitly within the GDPR in order for this instrument's success to be assessed. However, there must be some idea of what data protection entails *somewhere*, if not directly in the instrument, then on the level of the assessing observer.

some sort of semantic analysis, the very task that is rendered impractical by such endeavours. This leads to the *secondary* goal of autonomy research which is to undergird discussion and remedies of issues with autonomy. I propose here that it is this secondary objective that makes the theoretical analysis of autonomy relevant and worthwhile for legal professionals and philosophical theorists alike. Assuming that autonomy carries with itself a certain value, and that this value ought to be protected, it seems only consequential that (a) certain legal or factual measures may be taken to safeguard autonomy; (b) one may assess these measures by using the insights of a model of autonomy with respect to their scope and efficacy; and conversely; (c) not utilizing a clear model of autonomy hinders the inception of sufficiently scoped or efficient safeguards and their control.

This thesis is concerned with both of these sets of questions. On the one hand, this thesis will suggest a new concept of autonomy, one which builds upon existing literature but adapts them to be both more holistic in general and more relevant for issues of the IoE. On the other hand, this very concept will then be used to understand and pinpoint conflicts with autonomy, which may have already been identified with a lesser degree of precision on the basis of intuition. For this second item, value is added, if the hereafter introduced model allows for a more granulated view of autonomy and its conflict with informational agents than (1) the ambiguous colloquial use of the term autonomy and (2) the existing theoretical frameworks. This added value will be demonstrated in the later parts of the thesis, when the implications of this research for conducting ethical analysis and determining its reflection in the legal domain are investigated.

Part I: Autonomy

The first part of this thesis deals with the question of how autonomy can be conceptualized to be relevant and salient for investigating emerging issues of the IoE. To begin this inquiry, the text dives into the most influential philosophical theories relating to individual autonomy to establish a theoretical baseline. Informed by this, the text then analyses the field of privacy theory and fundamental concepts within the legal domain to find further reflection of an individual's capacity for self-government and the protection we intuitively tend to award to it. On the basis of this, the thesis proposes a tripartite model of how to conceptualize autonomy and understand autonomy constraints. In essence, the suggested framework distinguishes between three types of autonomy: Intrinsic autonomy is maintained through an individual's mental states, such as preferences, to be aligned with one another and one's actions. Relational autonomy describes capacity to act unbarred by societal or normative obstacles or matter-of-fact (e.g. physical) factors. Finally informational autonomy depends on an individual's access to information, as well as the form and veracity of such.

The main argument of this section is this: A robust framework to understand autonomy is important both for its own sake and to engage in meaningful discussion on how to protect autonomy where appropriate. For such a framework to be useful it must reflect existing intuitive autonomy concerns. To account for these concerns, autonomy must be understood to be constituted through mental congruence, access to appropriate and sufficient information and in absence of factual, normative and societal constraints, with deviations from this to be understood as autonomy constraints.

2 The Concept of Freedom and Autonomy

2.1 Introduction and General Theory

The word autonomy is derived from the ancient Greek terms *αὐτο* (meaning "self") and *νόμος (*meaning rule or law). This term is used colloquially and therefore comes with considerable opaqueness; it can arguably be used to denote the characteristic of free will, the capacity or action of self-control or control over one's life, the state of freedom of coercion, or a right to or a value based on the above; all of which are dogmatically different.[31] At the same time, all of these permutations of autonomy are widely considered to be charged with some sort of intrinsic value, and as such are reflected in the way society and the rules that govern it are structured. To investigate the intersection of emerging technology with (our) autonomy, it follows that the object of inquiry must be staked out precisely. To map the issues the emerging IoE could impose, a thorough analysis of the concept of autonomy is then a necessary first step.[32]

This chapter aims to illuminate modern academic understanding of individual autonomy and contribute to the answer of research (sub-) question SQ1.1, asking if there is there a current consensus on how to conceptualize autonomy in general, and in the domain of technology specifically. It will do so as follows. First, in Sub-Section 2.1.1, the domain of philosophy is identified as a relevant basis for further inquiry into individual autonomy. In Sub-Section 2.1.2 I highlight the four major obstacles to research into autonomy, how they inhibit the analysis and comparison of competing theories of autonomy, and how I address them within this thesis. In Sub-Section 2.1.3 I give a short overview over historical accounts of autonomy theories, most importantly by Immanuel Kant, for sake of completeness and to serve as context for the rest of the chapter. Section 2.2 introduces the enormously influential hierarchical conceptions of autonomy by reference to the theories of Harry Frankfurt, Gerald Dworkin and Andrew Sneddon. Sub-Section 2.3 highlights contentious differences between autonomy theories, some compatible with, but also others potentially invalidating hierarchical theory; the most important factors being information access and factual constraints of individual agents. Some thoughts about distinguishing autonomy from concepts such as liberty, free will and agency are given in Section 2.4. A conclusion and preliminary application to the scope of this inquiry is given in 2.5.

2.1.1 Philosophy as the Basis for Autonomy Research

To begin understanding autonomy, the domain of philosophy is a useful starting point.[33] We can recall Section 1.1 to roughly locate the concept of autonomy as self-rule or self-governance.

[31] The term autonomy can also be charged with (unambiguous) meaning in other specialized domains, such as computer science, robotics, etc. This section deals with autonomy as pertaining to human agents in a broad setting.

[32] In close connection to the concept of autonomy stand the concepts of freedom and or liberty. Within the rest of the thesis beyond this chapter, the term autonomy is used as a sort of umbrella-term to denominate principles that pertain to autonomy, freedom or liberty at large. However, this means that findings pertaining to this perspective of autonomy do not necessarily translate to other conception of autonomy, freedom or liberty without friction. Some general thoughts on the distinction between these principles is given in Section 2.4.1.

[33] Compare the statement *„Erkenntnisse aus den Natur- und Sozialwissenschaften [müssen] zwar zu einer Theorie der Autonomie beitragen [...], diese Theoriebildung [soll] aber eine primär philosophische Aufgabe bleibt und kein rein natur- bzw. sozialwissenschaftliches Projekt ist. (Although findings from the natural and social sciences must contribute to a theory of autonomy, the formation of this theory [should] remain a*

Within the limits of this formulation, the nature of the term autonomy is technical, i.e. it has or is meant to have a clear meaning, which is defined, and which is recognized by experts within a certain domain (i.e. philosophy). Similarly, although autonomy as a concept is valued throughout different cultures, the term relies on different connotations and philosophical assumptions in different cultural contexts. [34] As a result and similar to colloquial use, different autonomy theories attach, as will be explored immediately below, at different semantic, dogmatic and substantive levels.[35]

2.1.2 Four Obstacles to Inquiry into Autonomy[36]

There are some obstacles that need to be addressed if the concept of autonomy is to be explored fruitfully. The nature of the first (and widest) obstacle to further inquiry is *semantic (in the narrow sense).*[37] To illustrate, one might recount an instance of how the concept of autonomy is currently communicated: the Routledge Encyclopaedia of Philosophy defines autonomy as the concept of sovereignty over oneself, self-governance or self- determination.[38] One can derive from there that excluding or undermining autonomy is lack of self-governance but also determination by other factors outside of oneself.[39] This approach lends itself to be understood to delimit autonomy as a "capacity" of an agent to act in compliance with certain requirements that ensure sovereignty of the action. An agent that has this capacity (and exercises it) can be described to have an autonomous state. Action or choices taken on the basis of this capacity are then themselves autonomous.[40] However, utilizing a different interpretation of this text one might also understand this as the factual condition of an individual being in a state of self-governance, or as a product of rules that sculpt and enable self-governance.[41] Despite these difficulties of precisely denominating the target of the term autonomy as a descriptor, it seems evident that most of these concepts are co-dependent upon each other; e.g. the factual state of self-governance is obviously connected to the capacity to self-govern, etc.

The nature of a related, but distinct second obstacle is *dogmatic.* Different accounts of autonomy, while using the same terminology, offer varying depths of their theoretical

---

primarily philosophical task and not purely a project of natural or social science.)" in Markus Christen, 'Autonomie – Eine Aufgabe Für Die Philosophie', *Schweizerische Zeitschrift Für Philosophie*, 66 (2007) <https://doi.org/10.24894/StPh-de.2007.66012>. Cf. *"It soon became apparent, however, that autonomy was being used (in both senses) in a rather vague and excessively broad fashion. The concept of autonomy required the same kind of care and detailed mapping that ideas such as liberty and equality had received at the hands of earlier philosophers."* Gerald Dworkin, *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press, 1988), Preface.

[34] Cf. FN 42. See also Section 13

[35] To quote, again, Gerald Dworkin: *"It would be unwise to assume that different authors are all referring to the same thing when they use the term "autonomy. […] It is apparent that, although not used just as a synonym for qualities that are usually approved of, "autonomy" is used in an exceedingly broad fashion." In* Dworkin, *The Theory and Practice of Autonomy*, p. 5.

[36] The content of this section explores some methodological obstacles to autonomy research and contextualizes the coming analysis. It is not a necessary prerequisite to understand the rest of the thesis, so the avid reader interested in the substantive content of this inquiry may want move to Section 2.1.3

[37] Naturally, all of the obstacles explained here might be understood to be of a *semantic in the wider sense* - nature; they all ultimately pertain to what is meant by the term autonomy.

[38] Andrews Reath, 'Ethical Autonomy', *Routledge Encyclopedia of Philosophy* (Taylor and Francis, 1998) <https://doi.org/10.4324/9780415249126-L007-1>.

[39] Compare for this the terms *heteronomous* (rule by others), *oudenonomous* (rule by nothing) and *cosmonomous* (rule by non-personal features of the world) as (mostly) coined in Andrew Sneddon, *Autonomy*, 1st edn (London/New York: Bloomsbury Academic, 2013), p.3f.

[40] Reath.

[41] Cf. Joel Feinberg, 'Autonomy', in *The Inner Citadel -Essays on Individual Autonomy*, ed. by John Christman (Oxford University Press), p.27f.

attachment points. In other words, different accounts of autonomy will map their theories on different real-life and/or metaphysical phenomenon. To illustrate the differences of mapping on real-life instances, one account might map its necessary prerequisites for autonomy on factual circumstances surrounding the individual in question (e.g. physical, cultural, political or societal constraints), as well as their internal mental states or capacities. Another account might only consider a hierarchical model of mental states of an individual to be of relevance to the assessment of their autonomy and discount all other potential factors. A third might forego all external factors but expand internal prerequisites (e.g. mental states) to require additional conditions of authenticity of the individual to attest their autonomy. To illustrate the differences of mapping on metaphysical instances (or lack thereof), one account might be requiring adherence to a certain framework of morality, e.g. non-compliance with a "golden-rule" would prevent an individual to be autonomous (with respect to the behaviour in question). Another account might require a certain continuity of an individual's (meta-physical) identity. It becomes clear, that these differences do not just amount to different assignation of shared vocabulary to a finite set of instances. One account of autonomy might fully encompass another but expand on it, but they might also describe a very different set of instances when defining autonomy. To recall, the first obstacle of semantics in the narrow sense, described above, can be partly overcome by the realization that most terminology is invariably connected enough to each other so that meaning might not actually be lost. The same is not true for this obstacle of dogma, in which autonomy and its ascribed meaning and framework can be laid out in such different ways that there is no overlap and consequentially no inter-systemic recourse.

The third obstacle to inquiry into autonomy is of *substantive* nature. Separate accounts of autonomy propagated by theorists can conform in their semantic structure and overlap (at least partly) in their scope, but nevertheless arrive at vastly different conclusions when applied to an instance. To illustrate, two accounts of autonomy might both consider only mental states (e.g. preferences) of an individual and discount factual constraints of external factors but differ in how mental state insertion by external factors should be assessed. Clearly, both frameworks can account for such a phenomenon, so their dogmatic scope/depth is principally aligned; they just arrive at different conclusions when applying their ruleset. Substantive obstacles may arise at every decision point, in which there is a factual (or metaphysical) phenomenon that an account of autonomy, by its own rules, must consider.

One then has arrived with the three aforementioned obstacles to inquiry into autonomy, bringing with them distinct challenges. When one aims to compare existing accounts of autonomy to derive a new theoretical framework, two types of problems arise. The information that triggers *semantic* and *dogmatic* obstacles, to the extent that they do not overlap, could very well be consolidated in a "super"-account of autonomy, but because of their different scope do not lend themselves well to be compiled, that is compared to each other. On the other hand, *substantive* issues prevent consolidation, as the underlying information exclude each other, but do allow compilation or, more precisely, comparison.

In addition, there is a final obstacle to investigating autonomy and it is of *epistemic* nature. An analysis of autonomy, and the systems that have been devised to be denominated by it, is necessarily constrained by the very nature of the object of inquiry. Autonomy as a philosophical concept, for now disregarding any questions of value, is arguably an exercise of stipulative definition. There is little epistemic recourse to assess the validity of one account of autonomy

over another, indeed it seems difficult to imagine, that an account of autonomy can be correct or valid in the meaning of it being true.[42]

These obstacles cannot be fully circumvented, but they can inform the structure of the following inquiry. The analysis of accounts of autonomy and the draft of a synthesized, new and pragmatic, account of autonomy can be guided and assessed by their usefulness pertaining to the investigation of individual autonomy affected by the Internet of Everything. Different accounts of autonomy will not have to be compared on their abstract merits alone but based on this query at hand. *Substantive* differences, where appropriate, will be highlighted to show the underlying dichotomies; similarly, *dogmatic* (or *semantic*) peculiarities can be explored without committing to their overarching framework. The structure of this chapter, as laid out in Sub-Section 2.1, will reflect this by (1) introducing the basic notion of a hierarchical account of autonomy as a reference point, (2) comparing in a cursory manner such prototypical hierarchical accounts to other accounts that have been suggested by theorists and (3) outline *substantive* differences that can undergird all types of accounts.

2.1.3 Precursors of Modern Autonomy Theory

The term autonomy is not used consistently by philosophers, especially when surveying theories over an extended period of time. In its earliest permutation, i.e. in Greek philosophy, autonomy did not necessarily reference the self-rule of the individual but denominated certain political entities and communities, such as towns or cities that had the power to rule themselves.[43] By and large, the notion of autonomy as pertaining to an individual was only devised much later. To give some sort of historical context and to distinguish the usage of the term autonomy in its earliest form from current understanding, this section aims to briefly highlight the account of Immanuel Kant who could be credited to paving the way for our more contemporary understanding of autonomy in philosophy. Insofar other theorists, who will be mentioned in the later sections, explore autonomy, they do so ultimately on the very basis that introduced a paradigm shift towards analysis of the individual.[44] A cursory understanding of their predecessor is therefore of some use. This is true even more given the enormous influence the Kantian theory has had on western legal systems and still retains a prominent position within the domain of philosophy as well. However, I will argue that the Kantian account is insufficient for the purposes of this chapter.

For Kant, one of the earliest and still supremely influential philosophers to focus on the concept of individual autonomy itself,[45] autonomy is (simplified) to act in accordance with reason, whereas reason means that the underlying rules that guide the action are suitable as a general

[42] The word true is to be understood here as correctly representing some sort of facts, in the tradition of correspondence theory (but without committing to isomorphism). Cf. David Armstrong, 'Truth', in *Belief, Truth and Knowledge* (Cambridge University Press, 1973), pp. 113–34 <https://doi.org/10.1017/CBO9780511570827.009>.

[43] With the notable exception of the lesser known Dio of Prusa, see John Cooper, 'Stoic Autonomy', in *Autonomy*, ed. by Ellen Frankel Paul, Fred Miller, and Jeffrey Paul (Cambridge: Cambridge University Press, 2003), pp. 1–29.

[44] Of course, this is a consecutive phenomenon. For example, Kant's conception of autonomy was itself influenced by Jean-Jacques Rousseau's understanding of moral freedom, see Frederick Neuhouser, 'Freedom, Dependence, and the General Will', *The Philosophical Review*, 102.3 (1993), 363–95 (p. 366) <https://doi.org/10.2307/2185902>. Nonetheless, the deliberate use of the term autonomy in a contemporary understanding can be understood as starting with Kant.

[45] Christen, p. 178; Sneddon, *Autonomy*, p. 4,42; Paul Guyer, 'Kant on the Theory and Practice of Autonomy', in *Kant's System of Nature and Freedom - Selected Essays*, ed. by Paul Guyer (Oxford: Oxford University Press, 2005), pp. 115–45 (p. 115f).

law.[46] An individual's will, if autonomous, will then display a freedom on two dimensions. Negative freedom describes the independence of external[47] influences. Positive freedom describes the ability of the will is positively free to act in accordance with its own law.[48] However, this Kantian "autonomy of the will" serves to create not only an abstract description of autonomy but as stipulates autonomy as a "supreme principle of morality". Put differently, an action that is immoral, hence irrational, is not autonomous. This seemingly results in a conflation of the characterizations of an action as morally permissible and as autonomous, and a one-to-one mapping of the concepts. [49] Interestingly, the Kantian conception of autonomy leads to the conclusion that (normative) constraints to an individual's behaviour are not seen as diminishing autonomy but as an expression of that individual's autonomy in the first place.[50]

Within the Kantian theory, specifically within the notion of positive and negative freedom, first attachment points to connect issues of autonomy and the Internet of Everything become visible. For example, alien influences compromise the negative freedom of a person; this seems relevant as the IoE can serve as a conducive forum for exactly the same influences, for example through targeted behavioural advertising. However the strict conception of autonomy limits its usefulness beyond such general statements. Every effective external influence on one's actions to the extent that one does not comply with one's own law (which naturally must fulfil the requirements of reason) renders that individuum non-autonomous. At the same time, an informational agent that influences an individual to take an action that may fulfil the above criteria does not trigger a conflict under this system. This is internally consistent but may seem unintuitive when applied to phenomenon that are within the scope of this thesis: Given that autonomy is by definition government of one individual about themselves, it seems valuable to consider that such interference with their decision process should at least be of some concern. Such coarse (i.e. binary) filter makes a more granulated analysis of and differentiation between different types of influence by IoE agents difficult. For the purpose of this thesis, an account of autonomy that allows for a finer filter when it comes to the purpose of analysing interferences with the decision-making processes of an individual by informational agents is useful. Consequently, the next section will explore accounts of autonomy which, building on this first individual-focused approach, ultimately allow for a more complete and useful analysis of the challenges to autonomy by agents of the Internet of Everything.

---

[46] *Das Prinzip der Autonomie ist also: nicht anders zu wählen, als so, daß die Maximen seiner Wahl in demselben Wollen zugleich als allgemeines Gesetz mit begriffen sein. (The principle of autonomy is thus: 'Not to choose otherwise than so that the maxims of one's choice are at the same time comprehended with it in the same volition as universal law).''* Immanuel Kant, 'Grundlegung Zur Metaphysik Der Sitten', in *Immanuel Kant: Werke in Zwölf Bänden. Band 7*, ed. by Wilhelm Weischedel (Frankfurt am Main: Suhrkamp, 1977), p. 74; for an english version compare Immanuel Kant, *Groundwork for the Metaphysics of Morals* (New Haven: Yale University Press, 2002), p. 58.

[47] External meaning external to reason, as one's own desires are also considered to be external.

[48] Compare for the distinction in depth: Nobel Ang, 'Positive Freedom as Exercise of Rational Ability: A Kantian Defense of Positive Liberty', *The Journal of Value Inquiry*, 48.1 (2014), 1–16 <https://doi.org/10.1007/s10790-013-9399-4>; Manfred Baum, 'Positive Und Negative Freiheit Bei Kant', *Jahrbuch Für Recht Und Ethik / Annual Review of Law and Ethics*, 16 (2008), 43–56 <http://www.jstor.org/stable/43579350>.

[49] Cf. FN 28

[50] As pointed out in Samuel Reis-Dennis, 'Understanding Autonomy: An Urgent Intervention', *Journal of Law and the Biosciences*, 7.1 (2020) <https://doi.org/10.1093/jlb/lsaa037>.

2.2 Hierarchical Autonomy

2.2.1 Autonomy as a Hierarchy of Mental States

To alleviate the overly formalistic and moralistic constraints of theories such as the one of Immanuel Kant and motivated by advances in social and medical science, modern theorists have built upon the notion of autonomy as a capacity or trait of an individual. Contemporary hierarchical[51] theories mostly understand autonomy as a result of reflection upon one's own thoughts or mental states. Simplified, to act autonomously is to act not upon a lower mental state or want or desire[52], but to reflect upon the same and act based on this higher-level mental state or want or desire.[53] For example, an individual may have the impulse or desire to buy and smoke cigarettes, but also have the want to stop smoking. The desire to smoke is a first-order mental state (i.e. a desire about a decision or action), the desire to not want to desire to smoke is a second-order mental state (i.e. a desire about a different mental state). One then can be seen to act autonomous if one foregoes one's first-order desire in favour of one's second-order desire. Similarly, acting against higher-level mental states on the basis of lower-level mental states, i.e. giving in to one's desire to smoke despite wanting to stop smoking, would not result in autonomous actions, as the incongruence between the different wants inhibit reflective self-governance. To be an autonomous person (which is also sometimes denotes as having global or deep autonomy) is then a result of acting autonomously with respect to oneself.[54]

However, there are intricate questions that are contentious. Many questions persist, such as if autonomy should be a result of behaviour conducted over a period of time[55], or if the analysis should be limited to a certain moment in time.[56] Another point of conflict is the question of the content-neutrality of autonomy. Certain behaviour might limit autonomy at a later time,[57] and it is unclear if the initial behaviour can still be considered autonomous.[58] Other, narrower, theories persist, partly focusing on certain mental states (e.g. thoughts based on reason), such

[51] The terminology stems from the fact that these theories all conceptualize a hierarchy of immediately applicable mental states, i.e. desires or wants or thoughts etc., and more abstract mental states, i.e. desires or wants or thoughts etc. about other, lower-order mental states. See James Stacey Taylor, 'Introduction', in *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy* (Cambridge: Cambridge University Press, 2005), pp. 1–30 (p. 1f).

[52] The term "desire" is often used as one of if not the mental state which serves as the focal point for autonomy theory. However this is often called out to be unduly narrow, and might also be rooted in imprecise language in some cases. Cf. John Christman and Joel Anderson, 'Introduction', in *Autonomy and the Challenges to Liberalism*, ed. by John Christman and Joel Anderson (Cambridge University Press, 2005), p. 20 <https://doi.org/10.1017/CBO9780511610325> FN 6.

[53] There are different possibilities of intensity to such a concept. As an example, the congruence between the higher and the lower mental state might be sufficient when the higher mental state is either (1) a specific desire, (2) explicitly reasoned, (3) merely not rejected after consideration, (4) consisting of a specific class of emotions, e.g. satisfaction etc. The more stringent the assessment framework is, and the higher the requirements of autonomy, the more likely it is that a majority of agents will act or be non-autonomous (at times). Cf. for this distinction Sneddon, *Autonomy*, pp. 20–21.

[54] Harry Frankfurt, 'Freedom of Will and Concept of a Person', The Journal of Philosophy, 68.1 (1971), 5–20; Gerald Dworkin, The Theory and Practice of Autonomy (Cambridge: Cambridge University Press, 1988) passim; Andrew Sneddon, Autonomy, 1st edn (London/New York: Bloomsbury Academic, 2013) p.11ff.

[55] John Christman, 'Autonomy and Personal History', Canadian Journal of Philosophy, 21.1 (1991), 1–24.

[56] Alfred Mele, 'History and Personal Autonomy', Canadian Journal of Philosophy, 23.2 (1991), 271–80.

[57] A common example for this behaviour is a person selling himself into slavery based on her own "volition".

[58] Natalie Stoljar, 'Autonomy and the Feminist Intuition', in *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*, ed. by Catriona MacKenzie and Natalie Stoljar. (New York: Oxford University Press, 2000), pp. 94–111; Paul Benson, 'Feminist Intuitions and the Normative Substance of Autonomy', in *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, ed. by James Taylor (Cambridge: Cambridge University Press, 2005), pp. 124–42.

as rationality, or solely focused on the relationship of a behaviour with the self.[59] This section will deal mostly with the hierarchical accounts of autonomy suggested by Harry Frankfurt, Gerald Dworkin and Andrew Sneddon to provide an overview of current theory. As will be self-evident, the following paragraphs do not claim to fully represent the extent of the scholars' theory but aim only to lay out the cornerstones of their theoretical account as they are relevant to the object of inquiry of this text.

2.2.2 Frankfurt's Theory of the Freedom of the Will

The concept of autonomy as a hierarchical concept was arguably first introduced by Harry Frankfurt in 1971 in his influential article named "Freedom of the Will and the Concept of a Person". Concerning himself not exclusively with the notion of autonomy (a term that is not named in the article at all), Frankfurt attempts to define the concept of personhood and differentiates a person from other "creatures" (human or not) by the structure of its will.[60]

Humans are deemed to display the characteristic of having second-order desires, i.e. wanting to have or not to have certain desires and motives. Contrasting this, there is the concept of a first order desire, available to animals as well as to humans, which is the desire to do or not to do something. [61] Frankfurt outlines that these are in a relationship of hierarchy, i.e. a second order desire is a desire about a first order desire, but that they do not necessarily have to be completely aligned. One example would be that an agent might have a second order desire to experience a want without wanting to fulfil that want itself. If these desires are aligned, it follows that it is the agent's second order desire to have a first order desire and for it to be effective. Effectiveness in this context describes that the desire matches the outcome or action taken by the agent; a want that is actionable (i.e. the reason for the agent's action) is called a will.

Based on this distinction, Frankfurt devised the term of second-order-volitions. A second-order volition is a second-order desire of which the agent wants a certain (first-order) desire to be its will, so to be the motivating desire that leads the agent to a certain action. It is those second-order volitions that then are defined as essential to personhood.[62]

Exercising freedom of will is then "securing the conformity of [the agent's will to its] second-order volitions", or in other words an agent enjoys freedom of will "if he is free to want [i.e. will] what he wants to want". An agent that displays a second-order volition, e.g. the desire that X (being a first-order desire), should be its will, whereas X is the desire to do A and not B, is not exercising free will if contrary to the second-order volition, a different first-order desire, Y, becomes the reason for the agent to do B and not A (and by that becomes the will of the agent with respect to this action.) [63] Crucially, Frankfurt recognizes the reoccurring problem of the *infinite regress* that plagues hierarchical accounts of autonomy. A second-order desire or volition could be compared or connected to a third-order desire or volition (i.e. the want to want to want) and so on. For Frankfurt, this issue is resolved by an agent identifying itself "decisively" with a first-order desire which can be seen as acting as an expression of sorts of all higher orders of desires or volitions.[64]

[59] Sneddon, 41f.
[60] Frankfurt, p. 6.
[61] Frankfurt, p. 7.
[62] Frankfurt, p. 10.
[63] Frankfurt, p. 15.
[64] Frankfurt, p. 16.

The notion of an agent deciding between different first-order desires he has at a given moment to with respect to which one to "promote" is focused on the inner processes of the agent. For Frankfurt, the fact that certain actions were not available to an agent does not mean that the agent choose the action taken less freely.[65]

Frankfurt also recognizes the effect of external factors on first-order desires. In his examples, he uses physiological addiction to drugs to illustrate cases in which an agent's will is not by itself free due to its addiction ensuring that the desire to take drugs will be effective. However, such externally induced wills can be internalized if they conform with the agent's second-order volitions.[66]

2.2.3 Dworkin's Theory of Autonomy

In 1970, briefly before the publication of Frankfurt's foray into the evolving field of autonomy studies, Gerald Dworkin laid the groundwork for a similar hierarchical model of autonomy (again without mentioning of the term) in his article "Acting Freely".[67]. In his later work, he utilized the concepts introduced by him and Frankfurt to specifically investigate autonomy, most importantly using the notion of first- and second-order mental states. In his later book "The Theory and Practice of Autonomy", Dworkin closely conforms to Frankfurt's theory. Initially, he distinguishes autonomy from liberty or freedom of an individual. In this context he defines liberty as "the ability of a person to do what he or she wants [and] to have (significant) options that are not closed or made less eligible by the actions of other agents".[68] To illustrate the difference between liberty and autonomy, he suggests that exposure to false information should be seen as impeding someone's autonomy, but that this would not be the case for the individual's liberty.

Dworkin ultimately defines autonomy briefly as "a second-order capacity of persons to reflect critically upon their first-order preferences, desires, wishes, and so forth and the capacity to accept or attempt to change these in light of higher-order preferences and values."[69]

Dworkin also touches upon an often brought up distinction between global and local autonomy[70], which requires a short elaboration. Local autonomy means that the object of analysis are singular actions by an agent. The act of doing X and not B can therefore itself have the characteristic of having been autonomous, the agent itself was potentially acting autonomously. To contrast this, global autonomy describes the capacity of an individual to *be* autonomous over an extended period of time and with respect to not just one of his actions. seeing his theory to describe autonomy.[71] Dworkin proposes the aforementioned hierarchical framework as an assessment of global autonomy. The question which is answered by analysis

---

[65] Frankfurt, p. 19.
[66] Frankfurt, pp. 19–20.
[67] Gerald Dworkin, 'Acting Freely', *Noûs*, 4.4 (1970), 367 <https://doi.org/10.2307/2214680>.
[68] Gerald Dworkin, 'The Nature of Autonomy', *Nordic Journal of Studies in Educational Policy*, 2015.2 (2015), 28479 (p. 11) <https://doi.org/10.3402/nstep.v1.28479>.
[69] Dworkin, *The Theory and Practice of Autonomy*, p. 20. It is worth noting that Dworkin, in his book, is less concerned with the delimitation of autonomy; indeed most of the text is connected to issues of morality and practical application. Part of his deliberations on the nature of autonomy, nearly coextensive to referenced text, was (re-)published in 2015 in Dworkin, 'The Nature of Autonomy'.
[70] Different terminology is used for these considerations. Local autonomy is also called episodic or shallow, global autonomy also non-episodic or deep.
[71] Dworkin, *The Theory and Practice of Autonomy*, p. 15f.

the second-order capacity of agents is therefore not if a certain action was conducted autonomously but if the agent itself is autonomous. This then also serves as his basis to avoid the problem of infinite regress of higher-order mental states. While acknowledging the principal problem, Dworkin poses that the infinite regress only affects questions of local autonomy, that is with the respect of individual actions. When analysing if an agent itself is autonomous, Dworkin sees no necessity to ascertain if there are higher mental states in congruence with the necessary second-order mental states.[72]

2.2.4 Sneddon's Theory of Autonomy

Maybe one of the more extensive unified treatments of the definition and conceptualization of autonomy comes from Andrew Sneddon's treatise, eponymously titled "Autonomy" first published in 2013, in which he surveys the existing literature and, based on this groundwork, proposes his own account of autonomy.[73] It remains similar to the influential accounts of Harry Frankfurt and Gerald Dworkin that preceded it. While his work has drawn some criticism, in part due to what is perceived as imprecise language, it remains one of the more recent and exhaustive theories available.[74]

Sneddon's theory accounts for both local and global autonomy, or in his parlance autonomy of choices and autonomy of people. The term "choice" can be seen as similar to an effective desire within Frankfurt's theory, i.e. it has to be a mental state that has been acted upon (including abstaining from action). Local autonomy describes the autonomy of certain actions themselves, while global autonomy describes the autonomy of a person as a whole. The assessment with respect to the autonomy behind a decision is a question of local autonomy, while the assessment of an individual's autonomy generally is a question of global autonomy.

Local autonomy encompasses at the least decisions, desires and actions of an agent.[75] For Sneddon, autonomy of choice is primarily achieved if a first-order choice is accepted by a second order choice and consequently acted upon, or when a first-order choice is rejected by a higher-order choice and consequently not acted upon.[76] Sneddon acknowledges the situation of competing mental states at the same level. An agent might have different mental-states over a first-order mental state, meaning that an array of choices might be legitimized as autonomous (to a lesser extent). A weaker form of autonomy is achieved if choices are not actually legitimized by a dedicated thought process but would be congruent with the respective higher-order mental state if they were to be deployed. In other words, an agent might make a choice without reviewing it against his higher-order mental states, but if it would do so, the higher-order mental state would be congruent with the choice nonetheless. [77]

Global autonomy is derived from (1) extended exercise of local autonomy[78], therefore connecting the two capacities, (2) self-knowledge and (3) self-shaping. Self-shaping is the

[72] Dworkin, 'The Nature of Autonomy', p. 14. The validity of this argument is left for the reader to assess.
[73] Sneddon, *Autonomy*.
[74] Compare for critical reviews of his work Nomy Arpaly, 'Autonomy by Andrew Sneddon', *Notre Dame Philosophical Reviews*, 2014; Andrew Jason Cohen, 'Autonomy, Written by Andrew Sneddon', *Journal of Moral Philosophy*, 13.6 (2016), 764–67 <https://doi.org/10.1163/17455243-01306007>.
[75] Sneddon, *Autonomy*, p. 18.
[76] Sneddon, *Autonomy*, p. 24. Note the imprecise terminology in which Sneddon jumps between the terms thoughts, desires and choices. This is most likely an editing error.
[77] Sneddon, *Autonomy*, pp. 24–25.
[78] Global autonomy is therefore explicitly historical for Sneddon, unlike the more dualistic view of local autonomy, where the authenticity conditions are partly historical (i.e. causal integration) and partly ahistorical (

process of making autonomous choices about oneself accounting for information about oneself. Self-knowledge is the information that is necessary to intentionally and successfully exercise self-shaping comprising mostly knowledge about oneself but also, to a lesser extent, more general knowledge. While Sneddon does not require explicit reasoning to allow local autonomy, he does require it for global autonomy due to its necessity for self-shaping.[79]

Sneddon accounts for quite a few common problems of hierarchical theory in his framework. He addresses the problem of a looming *infinite regress* by suggesting that (1) mental states are not only in a strictly vertical hierarchical theory, but also generally connected to other mental states (Sneddon speaks of a "web" instead of a "ladder") and (2) higher order mental states (but not first-order mental states) can be self (and "self")-referential and therefore be their own source of autonomy.[80] Sneddon also specifically mentions the issue of manipulation of an agent's autonomy by mental state insertion by laying out a set of conditions with respect to the "self" of the agent that are required to attest autonomy of choice.[81] If an agent's mental state is derived from an external source, the level of integration into the rest of the agent's decision system is deemed to be relevant. Non-integration leads to non-autonomous action, while sufficient integration can mean that even a mental state that is externally imposed can be classified as autonomous. Integration can be done both on a content-level and more deeply on a functional level. Content integration denominates an integration which extends existing mental states (e.g. preferences) to objects that are similar to objects already targeted by the mental states. For example, an agent might have a strong preference towards reading a wide range of news websites daily for a certain amount of time (e.g. to catch up on important stories from around the world) but has a strong adverse reaction towards even contemplating to visit one specific website (irrespective of if the agent would enjoy the content of said website), causing the agent to not engage with it at all. Content integration would then be the insertion of a preference with respect to that specific website, potentially counteracting her adverse reaction. Functional integration of a mental state goes beyond this and can show either characteristic of causal integration and attitudinal functional integration, both of which create a state of coherence. Causal integration requires a (historically[82]) causal connection between the self and the mental state (and other mental states) as a result of the self-producing and controlling said mental state. Continuing the example above, the preference for the previously shunned website is causally integrated, if it is, over a certain period of time, in a causal interplay with the self and other preferences. Attitudinal functional integration occurs if other mental states account for and support the inserted mental states (which is an ahistorical state). If using the aforementioned example, there is a (second- or third-order) mental state that seems to leave room or endorses the inserted preference, i.e. the preference towards reading a wide range of news generally. Because the (inserted) preference towards a specific website fits the overall system of existing preferences, the inserted mental is attitudinally aligned. Only with a mental state being integrated functionally can an inserted mental state be drawn upon to imbue an action with autonomy. However even if autonomy of a choice is not reached, a heteronomous choice might still be qualified as autonomy-friendly or autonomy-neutral instead of autonomy-undermining.[83]

[79] Sneddon, *Autonomy*, p. 48ff.
[80] Sneddon, p. 28ff. However Sneddon discounts the existence of inherently autonomous (first-order) mental states.
[81] I.e. authenticity conditions, see Section 2.3.3
[82] The term „historical" denominates the concept, that historical context is necessary for the assessment of the agent's autonomy. A split-second assessment of a system of mental states "frozen in time" without sufficient information on how the system came to be as it is now, is insufficient to assess causal integration according to Sneddon.
[83] Sneddon, *Autonomy*, p. 34f.

2.2.5 Preliminary Findings with Respect to the Relevance of Hierarchical Theory to IoE-Analysis

How do these concepts fare when applied to emerging issues of the IoE? We can understand hierarchical theory as allowing the study of decisions and actions by investigating decision and action process characteristics. The conception of a "ladder" or "web" of connected mental states is quite powerful in that it can account for an individual's preferences coming from different origins. Connected to this, the concept of mental state insertion is similarly valuable within the IoE-context. For example, an individual's desire or want may come from a serious and deep reflection or from being exposed to highly effective advertising; it may be formed by impulsive choices under time constraints or in a drug-induced intoxicated state. Comparing these situations, it seems reasonable to attest different levels of autonomy to the concerned individual. This translates to situations of interest to this inquiry as well. An informational agent may interact with an individual in a multitude of ways, directly or indirectly and all of these interactions can affect an individual's decision as well as her decision process. For example, an automated process of an online commerce platform might alter the listing order of its products or change which products are shown to an individual altogether. Such an interaction (or change of interaction) certainly has the potential to alter the decisions that the individual will make, but it becomes difficult to pinpoint exactly where the issue lies with the individual's autonomy when using sweeping and generalized notions of autonomy such as provided for in Section 2.1.3. A different example is the "Emotional Contagion"-experiment, carried by researchers at Facebook, in which the content display logic of the social media networked was tweaked to prioritize content of a certain type, e.g. content with negative expressions. In this case, researchers found that the users exposed to such a shift of content experienced a changed emotional mental state.[84] Naturally, such changed mental states could influence decisions made by the affected individuals. Both of these examples, due to their potential effect on an individual's decision-making and decision-making process can be seen as examples of constraints to autonomy, and intuitively, have been portrayed as problematic. However, coarse-grained accounts of autonomy may either not see the peculiarities of such situations as autonomy undermining at all or may consider them harmful but are unable to differentiate between them. A purely Kantian account of autonomy may very well call one or both of these constraints immoral and therefore autonomy-inhibiting, but the nuanced issue of how these examples differ between each other (and what implications this has when addressing them with legal or other normative measures)[85] is likely lost under the blanket filter of compliance with a hypothetical general law. In contrast, hierarchical theory can account for the phenomenon of mental state insertion and integration, thereby providing a useful tool for analysis for situations that are prone to arise in the IoE-context. For example, with respect to the two constraints just elaborated, hierarchical theory can recognize similarities and differences more granularly. While both situations deal with an informational agent that serves as an intermediary for information exposure, the second example deals with specific emotional mental state insertion, which may necessitate analysing their level of (functional) integration, when assessing its impact on human autonomy; while the first example deals mostly with the abstract informational environment of the agent meant to exploit cognitive biases (i.e. choice architecture) or the agent's sensitivity of search costs (e.g. time, effort).

[84] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, 'Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks', *Proceedings of the National Academy of Sciences*, 111.24 (2014), 8788–90 <https://doi.org/10.1073/pnas.1320040111>.

[85] See for this secondary goal of autonomy research Section 1.6.

However, hierarchical theory by itself is not uncontested, and, in its pure form, not sufficient to serve as a basis for this inquiry. Apart from its conceptual flaw, the always looming theoretical issue of the "infinite regress", pure hierarchical theory does not take into account external circumstances of the individual during the decision-making process.

2.3 Delimiting the Extent of Autonomy

Within scholarly discussion, certain lines of debate have become clear. As theorists propose their frameworks to define and explain autonomy, they invariably reach problem sets that require a commitment to one or another school of thought. Quite a lot but not all of these instances take the form of an "either-or"-decision point, i.e. a certain phenomenon is assigned relevance or it is not, while others allow a choice of which and how many of the potential options are to be incorporated into the respective theory. This section aims to shine light on some of the more prevalent points of contention in scholarly debate, namely the consideration of the content of an autonomous action, the place of certain characteristics that must be immanent to the agent with respect to his or her capabilities and self and the consideration of factual constraints. The structure of this text is owed to the prominent place given to hierarchical theory in the previous section and aims to present these issues as they are relevant to such accounts. Nevertheless, the choice to use a hierarchical account is in itself a decision point and open to criticism.[86] One must therefore keep in mind the implicit limitation of assessing the rest of autonomy theory and its set of problems from a decidedly non-neutral standpoint. This section will elaborate on mainly two extensions to pure hierarchical theory, namely the relevance of content of an autonomous action and the relevance of factual constraints on an autonomous individual and their relevance for analysing interference by informational agents.

2.3.1 Procedure or Substance

2.3.1.1 Introduction - Kantian Autonomy as a Narrow Account of Autonomy

Some autonomy theories postulate that some actions are so inherently antithetical to the principles of autonomy, that they themselves cannot be considered autonomous. This shall be shortly illustrated with Immanuel Kant's previously introduced concept of autonomy. As outlined above in Section 2.1.3, Kant's framework mandates the characterizations of an action as morally permissible and as autonomous, i.e. a one-to-one mapping of the concepts. This narrows the application of Kant's autonomy concept considerably. Actions which are not suitable to be expanded into universal law are not morally permissible, therefore not supported by reason and can hence not be conducted autonomously. While Kant's concept, by virtue of it being centred on reason, draws its morality assessment from process consideration as opposed to content consideration, this still creates an entanglement between concepts of morality and autonomy. When rejecting the underlying moral principles, one will likely also reject the concept of autonomy with which these are entangled.

2.3.1.2 Content Relevance

More broadly and with respect not only to *dogmatic* but also *substantive* obstacles to understanding autonomy, such considerations separate autonomy theories into two categories: content-neutral (or procedural) and (content-) substantive accounts. Procedural accounts of autonomy allow individuals, in theory, to conduct every conceivable action autonomously.

[86] Christman and Anderson, p. 6.

Under this, the content of an action is fully irrelevant when determining if the action is autonomous. Instead, procedural accounts solely analyse the mental process that precedes the action. As long as there a certain kind of critical reflection (and ignoring the content of the action or decision), the action is considered autonomous.[87] In short, every action may be taken in an autonomous way. In contrast, substantive accounts of autonomy place limits on the range of actions that the individual can take autonomously. Some actions are deemed to be inherently incompatible with the concept of autonomy, even when they are a product of a sufficient process of critical reflection.[88] Naturally, these categories are not definitive. More formally, procedural theories may be divided further into structural, historical and competency-based theories, while substantive theories may be divided by their purity into weak and strong categories.[89] A closer analysis of these is beyond the scope of this thesis.

The outlined Kantian system above is substantive in a certain way; it postulates that immoral actions cannot be autonomous in any case. A different connection, also focused on moral consideration, is made by Susan Wolf, who requires "normative competency", a concept which entails the differentiation between "right" and "wrong" by the autonomous subject as necessary prerequisite to autonomy.[90] However, substantive accounts need not be entangled with moral considerations, even if their rules likely follow them in spirit. Even if considered moral or morally neutral, irreversibly giving up future behavioural options could be deemed to be incompatible with autonomy, due to its inhibitive effect on the individual's (future) autonomous capacity.[91] In general, it seems completely valid to take decisions that constrain one's range of option at a later stage. Indeed most decisions will lead to the preclusion of some other decision paths, e.g. when traveling to one country, one may not travel to another in the same period of time.

2.3.1.3 Consequences of Narrow Application of Autonomy

The stricter assessment of substantive accounts of autonomy results in fewer actions or decisions to be classified as autonomous. This is irrelevant from a moral-neutral perspective, but when charging autonomy with value, this can be perceived as setting the prerequisite standard for something that ought to be achieved as prohibitively high.[92] On the other hand,

---

[87] Stoljar, p. 94.

[88] Andrew W. Schwartz, 'Autonomy and Oppression: Beyond the Substantive and Content-Neutral Debate', *The Journal of Value Inquiry*, 39.3–4 (2007), 443–57 (p. 447) <https://doi.org/10.1007/s10790-006-3124-5>; Sneddon, *Autonomy*, p. 67.

[89] Catriona Mackenzie and Natalie Stoljar, 'Introduction - Autonomy Refigured', in *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*, ed. by Catriona Mackenzie and Natalie Stoljar (Oxford: Oxford University Press, 2000), pp. 3–35 (p. 13f).

[90] Slightly different terminology is used by Wolf: „ The freedom needed for responsibility involves the freedom to see things aright – the freedom, if you will, to appreciate the True and the Good."Susan Wolf, 'Freedom within Reason', in *Personal Autonomy* (Cambridge University Press, 2005), pp. 258–74 (p. 273) <https://doi.org/10.1017/CBO9780511614194.012>. However the term "normative competence" has been assigned to her theory nonetheless, see among others Gary Watson, 'Two Faces of Responsibility', Philosophical Topics, 24.2 (1996), 227–48 (p. 228). Note that these sources do not always clearly separate terminology for autonomy, freedom, free will, etc.

[91] Sneddon, *Autonomy*, p. 62f.

[92] This has implications both on a philosophical level and beyond. Scholars have criticized overly narrow substantive accounts of autonomy on the basis that only few or no people or their actions would qualify as autonomous. Of course, this makes little difference if the attempt is solely to conceptualize autonomy. A definition attempt of an ephemeral concept, that is arguably not accessible by epistemic means is not more or less valuable because it includes or excludes certain phenomena. Issues arise when this concept is already used prior to its delimitation. A rule that calls for the protection of an individual's autonomy creates a benchmark on

there are intuitive arguments that can be brought against purely procedural theory. For example, legal systems generally allow an individual a wide range of sovereignty under the umbrella of "freedom of contract", but they tend to prohibit or limit certain transactions that strongly limit future exercise of the same sovereignty. Roughly translating between the legal domain and the philosophical domain, there seem to be situations in which the law ought to limit certain actions or protect certain agents from decisions that may diminish their agency in the future. For example, consumer protection law often limits the extent of obligations that can be imposed on the consumer by contract, even though the consumer generally has contractual autonomy. The same principle applies to the most illustrative example given by autonomy theorists: selling oneself into slavery. Indeed, most if not all legal frameworks will not allow giving up one's freedom (indefinitely), even if that is what the agent prefers. There are also instances of "softer" limits, in which actions that forego future exercise of autonomy are allowed after certain requirements are met. For example, while individuals have wide agency with respect to their sexual life, the potentially irreversible decision of obtaining a sterilization is often contingent on fulfilling specific (ideally) autonomy-supporting requirements (e.g. additional consultations ensuring informed consent). It seems plausible that a theory of autonomy that aims to aid legal efforts in safeguarding an individual's capacity of self-governance should follow the legal approach and take into account these precursory autonomy-relevant situations as well. Finally and relatedly, the lens of substantial autonomy, even if not definitively adopted, allows to highlight set of facts that are most likely considered to be problematic by a considerable group of people. In order to maintain the relevance of this investigation, despite the aesthetic appeal of pure procedural autonomy theory, this thesis argues that for the purposes of this thesis one must therefore take into account the notions of substantive autonomy as well.

2.3.2 Relevance of on Agent's Relations and Factual Constraints

2.3.2.1 Framework and Theory

As mentioned, the concept of pure structural / hierarchical autonomy is not universally accepted. Relational accounts of autonomy challenge some of the assumptions that are made by hierarchical theorists, and do not concur with the notion of pure hierarchical accounts that autonomy is determined solely by mental states, and by extension factors that are intrinsic to an individual.

This approach holds potential for understanding autonomy constraints caused by informational agents as it is more sensitive of relationships between different individuals and constraints outside of an individual's control. Put differently, these theories are valuable in the context of understanding the intersection of autonomy and the IoE because they build upon an understanding that certain circumstances that affect the actionable envelope of an individual should not be ignored. The underlying premises, such as the judgment that information deficiencies are undesirable, are derived intuitively, but they tend to follow an established

how autonomy should or should not be defined, as the call for protection already makes a value judgment. Similarly, communicating that autonomy is something to strive for will be less successful if autonomy is later to be defined in such a way that most individuals will not achieve it. As will be described later in this work, humans often benefit from perceiving themselves as acting with agency; consequentially autonomy is a core concept in applied ethics such as medicine ethics. It is understandable that defenders of such a framework that focuses on an individual's autonomy, however imprecise this may be sketched out, see little value in theoretical prerequisites that preclude most individuals from attaining this propagated value (under the respective framework). Such dissonance, where one disagrees with an account of autonomy based not on a specific logical flaw of the account but by a general discomfort may be called intuition (after Stoljars "feminist intuition"), but holds little persuasiveness by itself.

consensus that is prevalent in society. When analysing phenomenon that are considered to be problematic in the general discourse, such as targeted misinformation by informational agents, it seems relevant to delimit autonomy in a way that these issues can be accounted for. Relational autonomy theories may therefore hold useful tools for the construction of a pragmatic account of autonomy that will be undertaken later in this thesis.

In general, relational autonomy expands its scope to account for an individual's relation to the world and the people around them and can both be combined with a procedural or substantial view (including hierarchical varieties) but also be developed in distinct directions. The main concerns of relational autonomy are twofold:

- Consideration of individuals as "emotional, embodied, desiring, creative and feeling" (in addition to rational) agents and hence the impact of socialization and social relationships on their autonomy, and
- consideration of oppressive socialization and social relationships on the basis of them being obstacles to an (a) individual's formation of desires, beliefs and emotional attitudes, (b) development of competencies and capacities that are required for autonomy and (c) the ability to act on autonomous desires and hence act autonomously.[93]

There are varying theories that attempt to account for this set of facts. While some theorists lay out competing theories, some adapt frameworks of existing (structural) theories. A full survey of these theories is beyond the scope of this thesis; however some abstractions can be made. Generally, relational autonomy accounts introduce two factors that are not necessarily included in pure hierarchical theory.

Primarily, relational autonomy accounts consider the embedment of an individual within their network of social relationships and "shaped by a complex of intersecting social determinants, such as race, class, and ethnicity".[94] Causal relational accounts of autonomy refocus, reject or adapt the concept of internal mental processes to achieve autonomy with inclusion of constraints. Social relationships or the factual socio-historic environment of an agent are considered by some, such as Annette Baier and Diana Meyers, to be an important factor to consider when assessing one's autonomy, even if autonomy is constituted by a different process or set of facts.[95] Constitutively relational accounts go a step further and require the analysis of such extant conditions for autonomy, instead of just allowing their impact on an account of autonomy that might be derived separately. In other words, external factors such as social conditions are defining conditions of autonomy.[96] Some theorists such as Susan Brison require a (factual) sufficient amount of actions to be available to an agent for the resulting action to be classified as autonomous.[97] Others such as Sarah Buss just explore certain impediments to

[93] Mackenzie and Stoljar, p. 21f.
[94] Mackenzie and Stoljar, p. 8.
[95] Annette Baier, *Postures of the Mind: Essays on Mind and Morals* (Minneapolis: University of Minnesota Press, 1985), p.85; Diana Meyers, *Self, Society and Personal Choice* (New York: Columbia University Press), passim.
[96] John Christman, 'Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves', *Philosophical Studies*, 117.1/2 (2004), 142–64 (p. 147).
[97] Susan Brison, 'Relational Autonomy and Freedom of Expression', in *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*, ed. by Catriona Mackenzie and Natalie Stoljar (Oxford University Press, 2000), pp. 280–300 (p. 285).

autonomy, such as mental or physiological constraints, e.g. fatigue, pain or anxiety[98], but also more wide-ranging social conditions such as oppression based on gender or slavery as formulated by Diana Meyers.[99]

The second factor that is derived from relational theory is an affirmation of the relevance of factual constraints as a whole.[100] Insofar external factors that stem from the relationship of an individual with other individuals affect, constrain or otherwise are important for autonomy, external factors that are concerned with the relationship of an individual with the world around them as a whole can also be relevant.

2.3.2.2 Relevance for Analysis of Informational Agents

The above considerations serve as a very powerful attachment point for analysis of IoE-related issues.[101] As informational agents are virtual entities, humans only interact with them through some sort of medium of interface.[102] However, the interface itself is increasingly controlled and optimized by the same informational agent. Social media sites display content dynamically and in an individualized way, what kind of information is shown, in what order and in what prominence is all personalized by an algorithm that usually tries to maximize engagement of the user. Similarly, online commerce websites suggest certain types of purchasable goods over others by utilizing past purchase history and personal data of the user. In all of these cases the interface changes to serve a function of the informational agent such as engagement of the user, maximization of visiting time or maximized revenue through suggested purchases. This inherently impacts the result of user-process interaction in two different ways. Firstly, an individual is just more likely to follow the logic of the informational agent; or put frankly: No one keeps looking after page three of Google. On top of that, due to the highly optimized personalization of the interface, the individual is subjected to a choice architecture that is playing towards inherent human cognitive biases through so called "nudges", making compliance much more likely.[103] Secondly, even if the individual can withstand such nudges and other influences, they may still act in compliance as it is the rational thing to do within the imposed constraints. One may very well buy a product that has been prominently placed on the first page of an online retailer with full knowledge that this may not be the optimal choice instead of digging through the website's full catalogue and sacrificing time and convenience in the process. After all there is a personal cost, usually a time cost, to any selection process, and it is only rational to weigh the selection process cost against the projected benefit of the result.

Even more salient issues arise in cases of no direct interaction between an individual and an informational agent. In the previously introduced example of a credit scoring algorithm it is likely that the individual does not directly interact with an interface of the process. Instead,

---

[98] Sarah Buss, 'Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints', *Ethics*, 115.2 (2005), 195–235 (p. 215) <https://doi.org/10.1086/426304>.

[99] Diana Meyers passim; Marina Oshana, *Personal Autonomy in Society* (Ashgate: Ashgate Publishing, 2005), p. 2.

[100] See e.g. Catriona Mackenzie, 'Relational Autonomy, Normative Authority and Perfectionism', *Journal of Social Philosophy*, 39.4 (2008), 512–33 (passim) <https://doi.org/10.1111/j.1467-9833.2008.00440.x>.

[101] See as an example of strong advocacy in favour of a (substantial) relational autonomy lens into e-health devices John Owens and Alan Cribb, '"My Fitbit Thinks I Can Do Better!" Do Health Promoting Wearable Technologies Support Personal Autonomy?', *Philosophy & Technology*, 32.1 (2019), 23–38 <https://doi.org/10.1007/s13347-017-0266-2>.

[102] See for an in-depth discussion of this Section 8.2.1.

[103] Cf. for an analytical take on nudging, albeit from the perspective of rationality instead of autonomy Andreas Schmidt, 'Getting Real on Rationality—Behavioural Science, Nudging, and Public Policy', *Ethics*, 129.4 (2019), 511–43.g

third parties such as banks, transportation or hospitality companies will be provided with information about the individual through the informational agent complete with a proprietary analysis about such data. Of course, if the informational agent attests the individual a certain score and that score compels a transportation company to inhibit the individual from buying plane tickets, the individual is factually constrained. Such results seem to be highly relevant when considering their autonomy.

The problems that are visible when considering a set of facts under the lens of relational autonomy theory are therefore highly relevant to the existing discourse of risks and advantages of informational agents. Incorporating the more generalized version of relational autonomy, i.e. considering external constraints from other entities and circumstances, allows for a more precise understanding of what is intuitively perceived as challenges to an individual's autonomy.

2.3.3 Competency and Authenticity

Approaching autonomy from a slightly different viewpoint, John Christman and Joel Anderson, surveying existing literature, lay out two broad types of requirements that an account of autonomy might require from agents, namely competency conditions and authenticity conditions. Competency conditions lay out requirements with respect to capacities for rational thought, self-control, self-understanding, etc. and postulate that agents must be able to act on these without coercion. These conditions therefore spell out which capacities are necessary to be able to exercise the required control, e.g. explicit rational assessment as opposed to mere hypothetical (and passive) congruence. Authenticity conditions are required capacities to reflect, endorse or identify with one's mental states.[104] In the words of Andrew Sneddon, competency conditions make up the "control"-part, while authenticity conditions make up the "self"-part of "self-control.[105]

The accounts previously explored make statements with respect to these conditions. Dworkin's theory requires procedural independence, i.e. the lack of subversion of reflective and critical capacities of an agent, and the capacity to raise the question of whether one identifies with a mental state (here: desire), both competency conditions[106], while Sneddon rejects competency conditions in favour of a more "heterogenous" concept of autonomy.[107] Authenticity conditions capture very broadly the notion of congruence, identification or compliance of mental states with each other and are therefore part of all previously discussed accounts of autonomy.

2.4 Autonomy-Related Concepts

2.4.1 Freedom and Liberty

As mentioned in Section 2.1, this thesis uses the word "autonomy" to sketch out a more rigorous conception of intuitively used terminology that relates to self-government, but also freedom and liberty more generally. This thesis uses theorists that deal explicitly with (a philosophical view) on autonomy, but it also is informed by other normative regimes (e.g.

[104] Christman and Anderson, p. 3.
[105] Sneddon, *Autonomy*, pp. 25–26.
[106] Dworkin, *The Theory and Practice of Autonomy*, p. 15., although he has since revised his position on this slightly.
[107] Sneddon, *Autonomy*, p. 26.

fundamental rights or "freedoms"), as well as some empirical results[108] that do not conform clearly to one conception or the other. Nonetheless, given the title and scope of the present inquiry, and to complement the analysis herein, this section aims to give a quick overview of some related concepts.

Generally, the terms freedom and liberty are often used interchangeably (perhaps because many (European) languages also use the same terminology for both).[109] (This section will use the term freedom going forward.) Generally, freedom is usually conceived as a description of dual phenomena,[110] namely as consisting of positive and negative freedom.[111] Simplified, negative freedom encompasses the *absence* of interference or obstacles to action. Positive freedom instead describes the capacity or ability to take said action.[112] Some theorists suggest yet other "middle-ground"-approach[113] to freedom: according to a school of thought dubbed "republican liberty", freedom ought to be understood as the state in which non-interference is guaranteed (instead of just a mere state of non-interference).[114] This is sometimes also denominated as freedom as non-domination.[115] Conceptually, autonomy and freedom are closely interwoven. Particularities will thus depend on what kind of account of autonomy and what type of freedom is assumed as the theoretical basis of analysis.[116] The pragmatic account of autonomy that will be introduced later[117] can also be characterized through the lens of positive or negative freedom,[118] and is situated somewhat between the purely theoretical autonomy theories and conceptions of liberty and freedom.[119]

---

[108] See e.g. FN 10, 488, 589

[109] See for this (and the following section) Ian Carter, 'Positive and Negative Liberty', 2003. However, as pointed out by Carter there are some suggestions for a more clear distinction between the concepts of liberty and freedom, as outlined e.g. in Hanna Fenichel Pitkin, 'Are Freedom and Liberty Twins?', *Political Theory*, 16.4 (1988), 523–52; Bernard Williams, 'From Freedom to Liberty: The Construction of a Political Value', *Philosophy \& Public Affairs*, 30.1 (2001), 3–26.

[110] However, the degree to which these two are truly separate is contested. In particular, MacCallum has suggested that these describe the same concept, see e.g. Gerald C MacCallum, 'Negative and Positive Freedom', in *The Liberty Reader* (Routledge, 2017), pp. 100–122.

[111] Isaiah Berlin, 'Two Concepts of Liberty', in *The Liberty Reader* (Routledge, 2017), pp. 33–57.

[112] Another way of understanding this is to divide these two conceptions in external and internal factors as pointed out by Carter.

[113] However, this republican conception of freedom is conceptually closer to negative freedom than to positive freedom. The question that divides these two relies on how much weight is given to the probability and possibility of obstacles (or other agents) to interfere.

[114] As so often in questions of autonomy, freedom and related concepts, these conceptions tend to be closely connected to political philosophy. For some sources on republican liberty see e.g. Daniel Kapust, 'Skinner, Pettit and Livy: The Conflict of the Orders and the Ambiguity of Republican Liberty', *History of Political Thought*, 25.3 (2004), 377–401.

[115] See e.g. Ian Shapiro, 'On Non-Domination', *University of Toronto Law Journal*, 62.3 (2012), 293–336; M Victoria Costa, 'Neo-Republicanism, Freedom as Non-Domination, and Citizen Virtue', *Politics, Philosophy \& Economics*, 8.4 (2009), 401–19.

[116] See for some different approaches (not necessarily compatible with one another) e.g. Kenneth Baynes, 'Freedom as Autonomy', in *The Oxford Handbook of Continental Philosophy* (Oxford University Press, 2007), pp. 551–87 <https://doi.org/10.1093/oxfordhb/9780199234097.003.0017>; Joseph Raz, 'Freedom and Autonomy', in *The Morality of Freedom* (Oxford University PressOxford, 1988), pp. 400–430 <https://doi.org/10.1093/0198248075.003.0015>; Keith Lehrer, 'Freedom, Preference and Autonomy', *The Journal of Ethics*, 1.1 (1997), 3–25 <http://www.jstor.org/stable/25115533>. Note in particular the different language: While Raz suggests that autonomy "underlies" positive and negative freedom, Lehrer seems to suggest a typology of freedom that builds directly on a "hierarchical"-esque conception of autonomy. In general, freedom is seen as a "wider" concept than autonomy.

[117] See Section 5

[118] See Section 5.7

[119] Note that the proposed theory is not meant as a rigorous philosophical method, as explained in Section 5.2

2.4.2 Free Will and Determinism

Another concept that is closely related to autonomy and its exercise is the question of free will and determinism. Determinism describes the conjecture that events, actions and all other phenomena are completely determined by the existing (and previous) state of all relevant factors. This also entails human decision-making and actions. Free will in turn describes a state of control of an individual (e.g. over one's actions).[120] Differentiation of free will from autonomy is not easy; perhaps the easiest posture to take in this respect is to understand free will as the exercise of autonomy in internal deliberations, whereas autonomy generally also consists of the capacity to translate free will into action.

At first glance, determinism and (potential) lack of free will present problems for autonomy theories. This seems relevant in particular as theoretical approaches to describing and necessitating free will generally are in apparent (potential) conflict with some findings in the field of neuroscience (among others).[121] However, most accounts of autonomy do not necessarily rely on free will (or in extension of this argument, the non-deterministic nature of agents). Autonomy theories that are unaffected by the (non-)existence of free will, just as theories of free will that are unaffected by determinism are often referred to as belonging to a theory of "compatibilism".[122]

2.4.3 Agency

Finally, we consider the concept of agency. Generally, agency describes a capacity of entities (i.e. *agents*) with respect to taking actions, whereas autonomy describes self-government in a more holistic way.[123] The concept of agency is described in more detail later in this thesis.[124]

2.5 Conclusion

This section has dealt extensively with the term autonomy, its meaning and delimitation and, as a result, has answered research sub-question 1.1. As has become clear, there is far from a consensus when it comes to what it means to be autonomous or to act autonomously. This is an issue insofar that the use of the term autonomy in general discourse seems to be used on the basis of intuition and as a catch-all concern. Most generally, autonomy is self-government. To act autonomously is to act as an expression of oneself; to be autonomous is to embody that expression, or for some, to act autonomously over a certain period of time. Beyond this almost tautological expression, one can identify diverging branches of theories and research. Structuralist theory focuses merely on an individual's intrinsic mental structure and their respective motivation, which are most often put into a hierarchical dependency framework. To be autonomous, an individual's internal decision-making process must meet certain requirements that shield them from undue external influences. In contrast, relational theory considers social and factual factors around the individual to be constitutive or at least necessary

---

[120] See e.g. Christopher Franklin Timothy O'Connor, 'Free Will', *The Stanford Encyclopedia of Philosophy2*, 2022 <https://plato.stanford.edu/archives/sum2022/entries/freewill/>. However this may already be a definition too wide to be useful and too close to the concept of individual autonomy.

[121] See e.g. Kelly Burns and Antoine Bechara, 'Decision Making and Free Will: A Neuroscience Perspective', *Behavioural Sciences \& the Law*, 25.2 (2007), 263–80; Kerri Smith and others, 'Neuroscience vs Philosophy: Taking Aim at Free Will', *Nature*, 477.7362 (2011), 23–25.

[122] See e.g. Michael McKenna, 'Compatibilism', *The Stanford Encyclopedia of Philosophy*, 2004.

[123] Note however that some autonomy theorists indeed focus on the operationalization of decisions through actions to evaluate autonomy, as outlined above. This is also the reason why Sneddon denotes his granular autonomy theory as "autonomy of actions" (as opposed to e.g. autonomy of choices).

[124] See Section 9.1

to their autonomy, removing emphasis from pure mental processes. Further debate is held over if foregoing future autonomy is in itself antithetical to one's own autonomy in the present, with procedural autonomy theories denying and substantial autonomy theories affirming the notion. All of these lines of arguments inform how an individual is or is not constrained in their autonomy, and subsequently how one may identify and potentially counteract such constraints. Due to the wide-ranging scope of informational agent influence on individuals and the intuitive recoil against such influence that is visible in general and academic debate, this thesis advocates for a wide understanding of the term autonomy in order to accurately and relevantly identify and address issues of autonomy constraints. In order to cut through the competing theories of autonomy research and to develop a shared concept, a unified and simplified account of autonomy that reflects many of the concerns highlighted above is presented in Section 5.

3 The Link between Privacy and Autonomy

3.1 Introduction

This chapter aims to answer research sub-question 1.2, namely how the concepts of privacy and autonomy are connected to each other in current discourse. Analysis will proceed as follows. Section 3.2 shows the prerequisite nature of privacy to the practical exercise of autonomy, particularly in a domain interlaced with technology interaction. Section 3.3 gives an overview over different conceptualizations of privacy propagated by theorists in the current debate. Section 3.4 highlights that the scope of privacy is subject to shifts and briefly outlines this both with respect to legal developments and with respect to subjective privacy expectations at the level of the individual. Finally, Section 3.5 shows how these findings stand in connection with the remaining analysis of this thesis.

3.2 Privacy as a Prerequisite to Autonomy

Of paramount importance for understanding autonomy and its constraints is the concept of privacy. This holds true for at least two distinct reasons. First, some sort of privacy is often seen to serve as a prerequisite to autonomy,[125] or at least as an enabling state hereto.[126] Second, some sort of privacy is considered to hold intrinsic value and therefore to warrant protection by nearly global consensus.[127] Due to their entangled nature, principles that safeguard autonomy must do so informed by privacy concerns as well.[128]

In its most simplified form, *individual*[129] privacy is a sphere absent of interference with respect to certain aspects of an individual's life. A right to privacy is, based on this understanding, the right to exclude and protect oneself from such interferences. Put differently, the right to privacy is the "right to be left alone".[130] A harm to or interference with privacy is consequently an interference that does not respect this individual's sphere. In recent discourse, the term privacy is often used to deal with information that an individual would like to keep excluded from others or the general public, hence would like to keep it *private*. With new technologies that have ushered in the so-called *Information* Age, and the growth of both automated information gathering tools and information processing tools, this aspect of privacy has grown in relevance. At the same time, concerns are raised about privacy as something that is not limited to an individual but has significance on a collective dimension as well, i.e. in which the association

[125] See e.g. Stanley Benn, 'Privacy, Freedom, and Respect for Persons', in *Philosophical Dimensions of Privacy: An Anthology*, ed. by Ferdinand Shoeman (Cambridge: Cambridge University Press, 1984), p. 241ff.
[126] See e.g. Joseph Kupfer, 'Privacy, Autonomy, and Self-Concept', *American Philosophical Quarterly*, 24.1 (1987), 81–89 (p. 83).
[127] See the elaboration of privacy as a nearly globally accepted common principle as reflected in human rights law and similar below in Section 4.7.3.2.
[128] Indeed, there seems to be considerable overlap in concerns of privacy and autonomy. Holvast, surveying existing literature notes that the terms freedom, control and self-determination are used in almost all publications relating to privacy, from which the relevance to autonomy is self-evident; see Jan Holvast, 'History of Privacy', ed. by Vashek Matyáš and others (Berlin, Heidelberg: Springer Berlin Heidelberg, 2009), pp. 13–42 (p. 16).
[129] Nb. that this definition is not necessarily applicable to conceptions of collective privacy, i.e. that describe privacy as a state or right that is held by a group either as the group itself or by its members on behalf of the group. (Compare e.g. the differentiation between associative and corporate collective privacy in Ugo Pagallo, 'The Collective Dimensions of Privacy in the Information Era', in *ANNUARIO DI DIRITTO COMPARATO E DI STUDI LEGISLATIVI* (Edizioni Scientifiche Italiane, 2020), pp. 115–38.)
[130] See e.g. ECHR (1984) Malone v. The United Kingdom, [Concurring Opinion of Judge Pettiti]

of an individual to some sort of group is relevant or constitutive to a certain set of privacy concerns.[131]

Autonomy informs the consideration of privacy concerns.[132] Most broadly, an individual's ability to make decisions, i.e. to exercise their autonomy, is considered to be enveloped by a subset of their privacy-sphere, i.e. their decisional privacy.[133] Information-related privacy is an intuitive prerequisite to autonomy to the extent that interference with an individual's autonomy becomes more effective as more information about the individual is known and can be acted upon. Connecting the concepts of privacy and autonomy is then the notion that the right to privacy encompasses (partly) the concept of human autonomy and concerns itself specifically with the human being as autonomous subject. A measure that protects an individual's autonomy should usually come with a protection of an individual's privacy as well.

3.3 The Manyfold Meanings of Privacy

A clear delimitation and comparative survey of different privacy concepts is beyond the scope of this thesis. However, it is worthwhile to highlight shortly that beyond the basic tenets outlined above, privacy theories are manifold and many link closely to some concept auf autonomy.[134] For example, Nowak understands the scope of privacy as the exercise of individual autonomy, where its existence and field of actions do not touch upon the sphere of liberty of others.[135] Westin describes privacy as the self-determination of individuals and/or groups to determine for themselves when, how, and to what extent information about them is communicated to others.[136] Solove initially suggested to see privacy pragmatically as comprising of and with respect to various activities, customs, norms and traditions, i.e. practices that are derived from cultural and social background[137], and later outlined a framework to understand privacy based on different kind of impediments and intrusions that can be imposed on it; divided into the overarching categories of information collection, processing, dissemination and invasion.[138] Nissenbaum explains privacy as to be understood contextually, that is if information is collected and shared in line with the expectations of the

[131] See e.g. Pagallo; Mittelstadt and others.
[132] For example, the European Court of Human Rights has found that Article 8 of the ECHR (i.e. Right to Privacy) is based on the concept of "human autonomy"; see e.g. ECHR (2002) Pretty v. The United Kingdom § 61; ECHR (2002) Christine Goodwin v. The United Kingdom § 90; ECHR (2007) Evans v. The United Kingdom § 71.
[133] Bert-Jaap Koops and others, 'A Typology of Privacy ', *University of Pennsylvania Journal of International Law*, 38.2, 483–576 (p. 50ff) <https://heinonline.org/HOL/P?h=hein.journals/upjiel38&i=489>. Note the distinction between decisional and intellectual privacy (freedom of mental intrusions), in which the authors suggest that decisional privacy is the active exercise of intellectual privacy. This tracks well with the conception of autonomy as a product of congruent mental states.
[134] Such concepts are often difficult to compare, given their different dogmatic attachment points (e.g. descriptive of normative / prescriptive). Smith et al point out that privacy frameworks can differ widely based on the assumption of which meaning is assigned to the term privacy to begin with, and this is true both within a certain academic domain and beyond. Smith et al taxonomize privacy theories between value- and cognate – based frameworks; see H Jeff Smith, Tamara Dinev, and Heng Xu, 'Information Privacy Research: An Interdisciplinary Review', *MIS Quarterly*, 35.4 (2011), 989–1015 <https://doi.org/10.2307/41409970>.
[135] See Manfred Nowak, U.N. Covenant on Civil and Political Rights - CCPR Commentary, 2nd Editio (Kehl: N.P. Engel, Publisher, 2005) p.377f.
[136] See Alan F Westin, 'Privacy And Freedom', *Washington and Lee Law Review*, 1968.
[137] See Daniel Solove, 'Conceptualizing Privacy', *California Law Review*, 90.4 (2002), 1087–1155 (p. 1126ff).
[138]See Daniel Solove, 'A Taxonomy of Privacy', *University of Pennsylvania Law Review*, 154.3, 477–564 <https://heinonline.org/HOL/P?h=hein.journals/pnlr154&i=493>. Note in particular the section about decisional interference, that tracks closely with relational autonomy concerns raised in this thesis.

subject.[139] Floridi differentiates between physical, mental, decisional and informational privacy, understanding them as freedom from sensory, psychological, procedural and informational interferences and intrusions.[140] In a similar way, Rosenberg proposes understanding privacy as consisting of three aspects: territorial privacy, privacy of the person and privacy in the information context.[141]

Notably, privacy, or more precisely the minimum acceptable level of privacy for an individual or a group is neither static nor objective. Individuals may allow for different levels of privacy for different kinds of information about themselves at a certain point in time, but at the same time change their assessment as time progresses. On a larger scale, privacy norms are subject to change, both in response to a change in circumstances and emerging technologies, not limited to digital technologies,[142] but also based on a changed attitude with respect to specific types of information. Lastly, privacy expectations and concerns can be different in different cultural contexts; individuals of a certain country, region or cultural background might have different expectations than people from a different country, region or cultural background.[143]

3.4 Privacy in Flux – Shifts to Privacy and Privacy Perception

Shifts in privacy norms and their salience reflect the importance of privacy to society and are accessible to both doctrinal and empirical research. From a legal standpoint, one can assess the legal instruments and their implementation as they pertain to privacy, and how they have changed over time. From an empirical view, analysis of aggregate subjective privacy perception, desirability and scope allows for an approximation of the value and importance privacy holds in society. This section lays out relevant findings in both of these domains and then affirms their relevance for autonomy considerations.

3.4.1 Privacy in the Legal Domain

Legal considerations of privacy have become more extensive, strict and enforceable over time. It is noteworthy that the concept of privacy as a right seems to have been developed on an international stage first, before being enshrined into domestic legislation. National jurisdiction had hitherto protected only aspects of what is considered today as privacy, e.g. the protection of one's home, correspondence or the inviolability of the body. Contrary to the usual consensus-building by states on the basis common denominators, an explicit general right to privacy was indeed first discussed, albeit not necessarily as a conscious choice, as content of an instrument of international law, specifically the international bill of rights.[144] Today, privacy

[139] See Helen Nissenbaum, Privacy as Contextual Integrity, Washington and Lee Law Review. 79 101 (2004). Empiric research seems to confirm the importance of context with respect to privacy and privacy expectations, see e.g. Kirsten Martin and Katie Shilton, 'Putting Mobile Application Privacy in Context: An Empirical Study of User Privacy Expectations for Mobile Devices', *The Information Society*, 32.3 (2016), 200–216 <https://doi.org/10.1080/01972243.2016.1153012>.

[140] See Luciano Floridi, The Fourth Revolution: How the Infosphere Is Reshaping Human Reality (Oxford: Oxford University Press, 2014), p. 208.

[141] Richard Rosenberg, *The Social Impact of Computers.* (San Diego: Academic Press Inc, 1992), p. 197f.

[142] For example, the highly influential article by Warren and Brandeis was prompted by privacy violations of the then recently emerging yellow press, see Samuel Warren and Louis Brandeis, 'The Right to Privacy', *Harvard Law Review*, 4.5 (1890), 193–220.

[143] See e.g. for differences in privacy perception between the generally similar "western" nations Great Britain and the United States of America Neil Richards and Daniel Solove, 'Privacy's Other Path: Recovering the Law of Confidentiality', *Georgetown Law Journal*, 96.1 (2007), 123–82. This is again touched upon in Section 13.

[144] Oliver Diggelmann and Maria Nicole Cleis, 'How the Right to Privacy Became a Human Right', *Human Rights Law Review*, 2014 <https://doi.org/10.1093/hrlr/ngu014>.

per se is an almost globally agreed upon and explicitly protected fundamental right, as granted by e.g.

- Article 8 of the European Convention of Human Rights,
- Article 12 of the Universal Declaration of Human Rights,
- Article 17 of the International Covenant on Civil and Political Rights,
- Article 16 of the United Nations Convention on the Rights of the Child,
- Art. 14 of the International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families,
- Article 7 and 8 of the Charter of Fundamental Rights of the European Union; and
- Article 11 of the American Convention on Human Rights.

It is however absent from e.g. the African Charter on Human and People's Rights. The scope of protection of these regimes in connection to technological progress is evolving and not final.[145]

After its explicit establishment within the legal domain, privacy protection has split, with data-bound informational privacy[146] having emancipated into its own and separate domain that is addressed mostly by the field of data protection. Informational privacy can be described the protection of interferences with decision making.[147] Interferences with one's decision makings are seen to be facilitated by an increase in effective data protection,[148] therefore seemingly warranting stronger protection of such (digital) information. Early protective measures of informational privacy aimed to protect privacy *per se*[149] and not data by itself but merely as a means to aid the former. These thoughts are deeply enshrined in the corpus of human rights and fundamental rights. Later iterations, in particular in Europe[150] have shifted towards comprehensive protection of data that is not just included within the umbrella term of privacy.[151] At the same time, jurisprudence of the existing regimes has shifted to account for both privacy and data protection as somewhat separate rights. [152] The right to data protecting

[145] Ineta Ziemele, 'International Protection of the Right to Privacy', *Max Planck Encyclopedia of Public International Law*, 2009.

[146] Despite the similar terminology, informational privacy and the corresponding informational self-determination is but one question within the intersection of autonomy and technology. It is not congruent with the object of this thesis, but merely a part of it. This distinction is relevant, as informational self-determination (or informational autonomy) suffers from the same linguistic weaknesses as its non-modified base; that is semantic ambiguity.

[147] E.g. " interest in independence in making certain kinds of important decisions', according to the U.S. Supreme Court in Whalen v Roe 599-600, 1977;

[148] See e.g. the very influential decision by the German Constitutional Court (*Bundesverfassungsgericht*) with respect to Informational Self-Determination in VerfGE 65,1 - Volkszählung, 1983, para. 93.

[149] CA-PIPEDA has been recognized as providing adequate protection (with respect to commercial entities) by the European Union Commission via Decision 2002/2 EC on the 20th of December 2001, as available here: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32002D0002

[150] See for a description of the differences between the USA and Europe e.g. Jürgen Kühling and Manuel Klar, 'Privatheit Und Datenschutz in Der EU Und Den USA – Kollision Zweier Welten?', *Archiv Des Öffentlichen Rechts*, 141.2 (2016), 165 <https://doi.org/10.1628/000389116X14684978889181>; Paul Schwartz and Daniel Solove, 'Reconciling Personal Information in the United States and European Union', *California Law Review*, 102.4 (2014), 877–916.

[151] See for an elaboration of this distinction e.g. Walter Berka, 'Aktuelle Bedrohungen Des Grundrechts Auf Privatsphäre', *Österreichische Juristenzeitung*, 17 (2018), 755.

[152] See e.g. M. Tzanou, 'Data Protection as a Fundamental Right next to Privacy? "Reconstructing" a Not so New Right', *International Data Privacy Law*, 3.2 (2013), 88–99 <https://doi.org/10.1093/idpl/ipt004>; J. Kokott and C. Sobotta, 'The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the

may now be understood to extend beyond the protection of the right to respect of private life.[153] Indeed, the Court of Justice of the European Union has found that certain cases may pertain to issues of privacy without relation to data processing or vice versa.[154] However, such data protection measures are not as homogenized, and not as embedded in global international law. Nonetheless, at least in the European Union, the right to protection of personal data is a fundamental (and separate) right explicitly, virtue of Article 8 of the Charter of Fundamental Rights of the European Union (CFEU).

Where such explicit protection is not granted by an international rights catalogue, recourse is often had to the existing overarching privacy protection when it comes to modern technology. To this end, the European Court of Human Right has produced extensive jurisprudence based on Article 8 of the European Convention of Human Rights that deals with issues of technology. Within the framework of international legal regimes, and as a response to emerging technology, specific regional or domestic legislation has filled the gaps to protect individual and, in particular, informational privacy. This applies especially to personal data and their protection. Some of the most well-known (and generally considered to be wide-ranging) data protection safeguards in the "western world" at the time of writing are the European Union's General Data Protection Regulation (GDPR) and to a lesser extent the California Consumer Privacy Act (US-CCPA) and the Canadian Protection of Personal Information and Electronic Documents Act (CA-PIPEDA).

Interestingly, most recently, and helped through the adoption of the GDPR, proactive countermeasures to privacy intrusions such as privacy by design and privacy by default have emerged both as a sort of ethical imperative and normative requirement.[155]

3.4.2 Subjective Privacy Perception

There seems to be empirical evidence beyond mere legislation that suggests that privacy perception and expectation of society has changed. Commentaries about new or increasing threats to privacy, often with dramatic wording, are highly prevalent.[156] For example, Goldfarb and Tucker have, based on market data from the years 2001-2008, suggested that privacy concerns have increased both over their analysed timeframe and with age of the surveyed individuals, and that technological advances and the ability to automatically process

ECtHR', *International Data Privacy Law*, 3.4 (2013), 222–28 <https://doi.org/10.1093/idpl/ipt017>; Raphaël Gellert and Serge Gutwirth, 'The Legal Construction of Privacy and Data Protection', *Computer Law & Security Review*, 29.5 (2013), 522–30 <https://doi.org/https://doi.org/10.1016/j.clsr.2013.07.005>.

[153] Alexander Roßnagel and Christian Geminn, '"Privatheit" Und "Privatsphäre" Aus Der Perspektive Des Rechts - Ein Überblick', *Juristenzeitung*, 2015, 703.

[154] See pointed out with reference to the case of CJEU (2014) C-131/12, Google v. AEPD by Ugo Pagallo, in 'The Collective Dimensions of Privacy in the Information Era', in *ANNUARIO DI DIRITTO COMPARATO E DI STUDI LEGISLATIVI* (Edizioni Scientifiche Italiane, 2020), pp. 115–38 (p. 121).

[155] See Article 25 of the GDPR. It is noteworthy, that even soft regulation such as "best practices" can be effective in safeguarding privacy, as argued by Harry Surden: "*[…] privacy interests are protected not by positive legal prohibitions on behaviour, but by structural constraints which act as reliable substitutes for legal constraint.*"; see Harry Surden, 'Structural Rights in Privac', *SMU Law Review*, 60 (2007), 1605–29 (p. 1612).

[156] See e.g. Michael Froomkin, 'The Death of Privacy?', *Stanford Law Review*, 52.5 (2000), 1461 <https://doi.org/10.2307/1229519>; Clifford Fishman, 'Technology and the Internet: The Impending Destruction of Privacy by Betrayers, Grudgers, Snoops, Spammers, Corporations and the Media', *George Washington Law Review*, 72 (2004), 1503.

information pertaining to certain contexts has led to privacy concerns enveloping that context.[157]

Shifts in privacy perception may also concern the (perceived) actors and intruders of one's private sphere and the type of privacy intrusion that is considered the most prevalent or important. For example, Kasper notes that over the period of 1980 to 2003, U.S. American newspapers (being allegedly representative of the U.S. media ecosphere at the time of her writing) showed increased coverage of what she calls privacy invasions of which the "invadee" is unaware (and in contrast a decline of coverage of "known" privacy invasions), and that corporations (and government) have been increasingly seen as main culprits of such invasions.[158]

At the same time, there seems to be a meta-concern that the expected or desired level of privacy might become lower.[159] To this end, the divergence between a positive attitude towards strong privacy and engaging in privacy-compromising behaviour at the same time is by now well documented, and dubbed the "Privacy Paradox".[160] Nonetheless, while this raises important questions about the source of this incongruence, this does not diminish the clear propagation of privacy in current society *per se*. Indeed, privacy generally seems to continue to be valued, and intrusions into privacy are generally considered to be undesirable, even if acceptable trade-offs may exist from an individual's point of view. While unconscious biases or behavioural routines can undermine an individual's privacy-sensitive attitude, research suggests that when faced with appropriate information, individuals' place additional value on privacy. To this end, Egelman, Felt and Walter showed that consumers may be more likely to pay a premium for a smartphone application with enhanced privacy features, as opposed to the same application with a more intrusive privacy approach, but that this effect occurred mainly when consumers were presented these options in a way in which the privacy implications were easily comparable.[161] This seems to be in line with findings in the wider domain of behavioural economics, in which exact preferences are highly contextual, and choice architecture[162] can influence individual's actions even against some of their stated interests. Acquisti proposed in this context that individuals do not always have the capacity to act as economically rational agents concerning personal privacy.[163] Denying the importance and desirability of privacy to individuals based on behaviour that is not fully congruent with achieving the utmost privacy does not seem necessary in the light of findings like these. Indeed, differences between stated privacy preferences and taken actions might highlight certain type of privacy or autonomy

[157] Avi Goldfarb and Catherine Tucker, 'Shifts in Privacy Concerns', *American Economic Review*, 102.3 (2012), 349–53 <https://doi.org/10.1257/aer.102.3.349>.
[158] See Debbie V S Kasper, 'The Evolution (Or Devolution) of Privacy', *Sociological Forum*, 20.1 (2005), 69–92 <http://www.jstor.org/stable/4540882>.
[159] See e.g. Anita Allen, 'Coercing Privacy', *William & Mary Law Review*, 40.723–757 (1999), p. 730f.
[160] See for a systematic overview of available research both Spyros Kokolakis, 'Privacy Attitudes and Privacy Behaviour: A Review of Current Research on the Privacy Paradox Phenomenon', *Computers & Security*, 64 (2017), 122–34 <https://doi.org/10.1016/j.cose.2015.07.002> and Susanne Barth and Menno D.T. de Jong, 'The Privacy Paradox – Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behaviour – A Systematic Literature Review', *Telematics and Informatics*, 34.7 (2017), 1038–58 <https://doi.org/10.1016/j.tele.2017.04.013>.
[161] See Serge Egelman, A Felt, and David A Wagner, 'Choice Architecture and Smartphone Privacy: There's a Price for That', in *WEIS*, 2012.
[162] Specific choice architecture, i.e. "nudging" environments can themselves be classified as a privacy intrusion, and in any case as an autonomy constraint, as will be further elaborated in Section 6.4.2.
[163] Alessandro Acquisti, 'Privacy in Electronic Commerce and the Economics of Immediate Gratification', in *Proceedings of the 5th ACM Conference on Electronic Commerce - EC '04* (New York, New York, USA: ACM Press, 2004), p. 21 <https://doi.org/10.1145/988772.988777>.

constraints, that may need to be addressed, rather than invalidating the privacy preference in question. In any case, the Privacy Paradox and the change of privacy perception over time is an important reminder of the malleability of the concept of privacy and its perception both over time and between different groups.

3.5 Privacy Considerations as Relevant Factors for Conceptualizing Autonomy

Within the scope of this thesis, privacy and in particular data protection and their legal representations as well as privacy perception are important considerations for two reasons. Firstly, existing privacy and data protection frameworks, as they pertain to human autonomy, can be used to approximate the valuation of certain aspects of autonomy by society at large. Issues that have been enshrined in legislation or adjudicated by courts, in particular in final tribunals such as the European Court of Human Right or the European Court of Justice or national supreme courts, tend to highlight important and visible conflicts between certain practices and the existing ethical norms that undergird the applicable legal framework. Emerging technology, when creating new conflicts, highlights these issues as well as societal valuation of certain phenomenon, such as autonomy; once technology is threatening to constrain a certain feature of life, it cannot be taken for granted anymore. Insofar old legislation or treaties are re-interpreted to adapt their protective scope to account for these new potential intrusions, or new legislation is introduced for the same reason, this can be seen as an endorsement of the phenomena and values the legislation is meant to protect. When considering questions of autonomy in connection with emerging technology, it seems highly relevant to include the considerations of the judicial and legislative sector when staking out privacy and data protection as the prerequisite environment to enable autonomy.

Secondly, as will be highlighted later in this text,[164] informational agents rely on some sort of information, be it personal or more remote data. The collection, processing and enacting of this data tends to impede privacy and data protection preservation efforts. In reverse, data protection and privacy measures are an important lever to curb persuasive technology. Naturally, the same is true for empirical findings of a shift to privacy perception; societal concerns should be able to be represented in an autonomy framework for it to be relevant.

As a result, and answering research sub-question 1.2, the domains of privacy and data protection can serve both as a guidepost and as a tool to preserving individual autonomy and must be considered when sketching out a pragmatic account of autonomy in Section 5.

[164] See in particular Section 9.3.2.4

4 Legal Representations of Autonomy

4.1 Introduction

The concept of autonomy in some form or another permeates the legal domain. Indeed, the etymology of the word autonomy makes clear that it was first used in a legal context.[165] After the previous sections have surveyed the philosophical-theoretical and privacy-adjacent concepts of autonomy, this section will briefly reflect on how the notion of self-government is reflected generally in the field of law, and in particular in the substantive rules of international human rights law. Understanding the representation of the concept in the legal domain and outlining both its scope and the underlying valuations represent a necessary step before conceiving the pragmatic account of autonomy in Section 5, insofar that such pragmatic account should reflect the considerations of the legal domain to be useful in its application.[166]

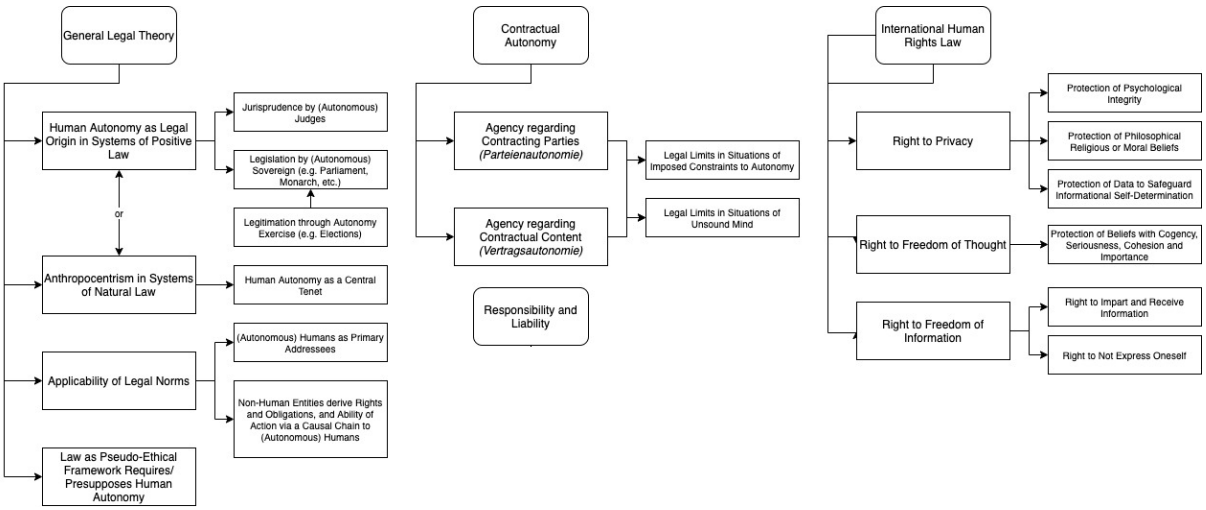

Figure 1 - Legal Representations of Autonomy

This section is structured as follows: Section 4.2 briefly addresses the issue that autonomy concepts do not cleanly transfer between different scientific domains. Section 4.3 introduces and contextualizes the legal debate about autonomy in an informational society by reference to the famous German court case regarding Informational Self-Determination and warns about the risks of adopting a myopic view of the same. Section 4.4 elaborates on the general presumption of human autonomy and agency that law in itself exhibits. Section 4.6 expands this notion by arguing that individual responsibility under civil and criminal law relies on assumptions on human autonomy as well. Section 4.5 analyses briefly the notion of contractual autonomy as one of the more explicit reflections of human autonomy in the legal domain. Finally, Section 4.7 conducts a deeper dive into autonomy and autonomy-adjacent values in the domain of international human rights with ample references to jurisprudence.

[165] See Section 2.1.
[166] See for a defence of this argument Section 1.6.

4.2 Issues of Interdisciplinary Transferability of Autonomy Concepts

The approach of comparing ideas originating from theoretical philosophical autonomy accounts with concepts from the legal domain is not without problems. This is part of a bigger methodological challenge that was outlined in Section 2.1.2 and will again be touched upon in Section 5.2; neither terminology nor concepts of different fields translate cleanly from one domain to the next. This section deals with the concept of autonomy in a broad, intuitive sense, and aims to find representations of parts of that concept in the legal sphere, but this must necessarily come with certain caveats. Insofar, for example, this thesis suggests that jurisprudence of the European Court of Human Rights reflects certain concepts of established autonomy theory as it does in Section 4.7, it must be clear that such comparison is only an approximation and is done for interdisciplinary comparison, not rigorous legal argumentation. A more detailed methodical justification is given at a later point,[167] where the findings of the first part of the thesis are applied to contextualize an account of autonomy that will be used for the rest of this inquiry.

4.3 Excursus - Beyond Informational Self-Determination

In 1983, in a landmark decision, the German Constitutional Court (*Bundesverfassungsgericht*) recognized the concept of informational self-determination as encompassed in the general personality rights of German Citizens, as imbued by the German Constitution. The decision noted that automated information gathering and processing allowed for an exceedingly complete profile of the personality and (consequentially) ever more effective means of influence.[168] In the wording of the court, there is no such thing of a "trivial datum" if it can be subjected to automated processing, as the significance of information relies not (only) on its intrinsic properties, but on external factors such as its (potential) use.[169] Consequently, the court recognized the right of an individual to (individually) determine use and disclosure of their personal data.[170] Naturally, this decision, affirming the arguments that German legal scholars had been bringing forward for a while,[171] had significant impact beyond the jurisdiction it was reached in.[172]

Informational self-determination is but one question within the intersection of autonomy and technology, but it is one of the first instances of legal authorities recognizing the increased leverage over human self-determination by the use of technology. As a result, the concept of informational self-determination enjoys wide-ranging recognition. It is however not fully congruent with the object of this inquiry, but merely a part of it. This distinction is relevant, as informational self-determination (or informational autonomy) suffers from the same linguistic weaknesses as its non-modified base; that is semantic ambiguity.[173] It is therefore necessary to separate the notion of informational self-determination or autonomy as enshrined in the case

[167] See Section 5.2.
[168] BVerfG, 1 BvR 209/83 (Volkszählung), 1983, para. 93.
[169] BVerfG, para. 98.
[170] BVerfG, para. 1.
[171] Cf in particular Wilhelm Steinmüller and others, *Grundfragen Des Datenschutz - Gutachten Im Auftrag Des Bundesministeriums Des Innern (Drucksache VI/3826)*, 1971, p. 93,96,120.
[172] With some scholars going as far as calling it an "*avant-garde* decision", see Antoinette Rouvroy and Yves Poullet, 'The Right to Informational Self-Determination and the Value of Self-Development: Reassessing the Importance of Privacy for Democracy', in *Reinventing Data Protection?*, ed. by Serge Gutwirth and others, 1st edn (Dordrecht: Springer, 2009), pp. 45–76 (p. 45).
[173] See Section 2.1.1

of the German Constitutional Court and widely adopted hereafter and the notion of autonomy of individuals within an informational technology sphere.

Autonomy is, as outlined in Section 2.1, a characteristic, state or capacity of an individual to be self-determined. This can be investigated with respect to specific domains; an individual can or cannot be autonomous with respect to some aspects of their existence and the system they are part of. Informational self-determination is, in a way, a type of autonomy that is twice delimited. It is realistically specific to a technological domain and the autonomy extends only with respect to the use of personal data. Said differently, an individual exercising their informational self-determination takes action with respect to the disclosure and use of their personal data. In contrast, this inquiry aims to investigates autonomy at a larger scale. The proposed framework for autonomy is conceptually more abstract as the concept of informational self-determination, and therefore encompasses it. At the same time, technology (e.g. automated data processing, subsequent profiling, etc.) serves not only as the *reason d'être* for the autonomy framework, but it is in active interplay with it. An individual's autonomy (of course including her capacity to exercise her informational self-determination) is not just recognized and protected because of technological capabilities of data processing operations, indeed it can be influenced by it as well.

In a way, the German constitutional court recognized these effects on an individual's autonomy via cursory remarks in its decision. The court (non-exhaustively) addressed areas of conflict that are also subject of this thesis, e.g. inhibitive behaviour of individuals, e.g. attempting to conform to a certain standard, because of their knowledge of automated data processing.[174] Of course, within the logic of the decision, this mostly serves to support the reasoning behind the value or need of informational self-determination, and is not explored as something that would influence the capacity of an individual exercising the same informational self-determination.

In summary, one must therefore be careful not to equate existing literature on informational self-determination (which is legion) to the inquiry into how an individual can be self-determined within a system that is home to agents imbued with automated information processing capabilities with all their consequences. However, we may take inspiration from this method all the same. Just as the German Constitutional Court found the right to informational self-determination within the scope of one of the broader fundamental rights of German Citizens in the *Grundgesetz*,[175] the legal protection of autonomy at large is governed first and foremost by rights guaranteed either by constitution or by means of international treaty, all connected to a select few human rights.

4.4 The General Legal Presumption of Autonomy

On the highest level of abstraction, law both presupposes and guards autonomy, i.e. the self-government of individuals as themselves and, as extension of this, as a collective. Law can be understood as a set of rules, both substantial and procedural, that structures aspects of human life by mapping theoretical normative concepts to real-life phenomena.

Most legal rules are an extension of human autonomy by virtue of their origin. Statutes and other written legislation are created by a legislative body that is meant to reflect the will of the sovereign, be that the general populace in case of a democracy or a more limited circle of

[174] BVerfG, para. 145.
[175] Specifically Article 2 para. 1 in connection with Article 1 para. 1 German *Grundgesetz*.

individuals. The makeup of the legislative body is usually directly linked to the exercise of human autonomy, e.g. in the form of elections. Jurisprudence and its status as a genuine source of law is similarly derived by the human nature of the deciding entity; insofar a non-human arbiter's decisions are considered to be legal sources themselves this must be reflected in appropriate legislation in the first place. Indeed, legal scholarship has widely denounced the idea that judicial bodies merely serve as legal calculators, instead imbuing them with the task of inserting real human choice into the process of legal problem solving.[176]

There are cases where certain rules of law are not considered to be mutable and therefore accessible to the agency of the sovereign and its legislative or jurisprudential bodies, such as is the case in the theories of natural law, or the international body of law denominated as *ius cogens*. However, the primacy of these norms is in themselves closely linked to the unique place that humans are thought to have within their domain; and consequently superior status of these rules can be traced to the core values of such human exceptionalism which include human agency and autonomy.

Legal rules pertain to a wide range of physical phenomenon, but a human link is always maintained. Human autonomy is reflected in the fact that law addresses rules to humans first and foremost. Insofar law addresses entities apart from a natural person, such as a collective of natural persons, legal entities (e.g. corporations), or certain property (e.g. estate in the sense of inheritance), there is generally a clear causal chain from such entities to a human individual, and most often these links display the importance of human autonomy as well. Legal entities such as companies, trusts, political organizations and the like are generally derived from an intentional act of human individuals. For example, an inheritance estate is assigned to and can only act through a human intermediary such as a notary public and is eventually transferred into the property of a human individual or collective.

Finally, one function of law is that it serves as outlining acceptable and unacceptable behaviour, i.e. law serving the function as a pseudo-ethical framework of sorts.[177] Responsibility under and with respect to compliance with this framework, as attributed under the law in many different varieties, seems to require and build upon the concept of human autonomy and agency. Most generally, the notion that law deals with human action and endorses some over others seems to be most consistent with the presupposition that humans can exercise autonomy by acting freely and in extension of their own self-governance.[178]

---

[176] See for a survey of this historical trend Regina Ogorek, *Richterkönig Oder Subsumtionsautomat?: Zur Justiztheorie Im 19. Jahrhundert* (Frankfurt am Main: Vittorio Klostermann, 2008), passim.

[177] Simplified, certain parts of law can take on the function of dividing up actions and/or their resulting states in the categories of legal and illegal (as well as unregulated or neutral, in which case the law can provide a default endorsement for one or the other). With the implicit assumption, strengthened by the general notion of law to be equipped with some sort of sanctions, individuals are "encouraged" to commit to the legally endorsed catalogue of actions and states. This in itself is arguably morally neutral, however, it creates a ethical-esque like framework in which there is a clear denomination of "right" and "wrong", not in a moral but in a practical-legal sense. Beyond this function of law, other rules or assignations may be merely descriptive, often to aid outlining subsequent distinction between legal and illegal phenomena.

[178] This mirrors the philosophical debate between moral responsibility and determinism. However, while many "hard determinist" theories don't allow for moral responsibility in the face of determinism, this does not necessary apply to questions of autonomy and agency. Indeed there is the question of human autonomy being possible in case of determinism that needs to be answered first, with the confirmatory view being called "compatibilist". See also Section 2.4.2. For a general and historical-philosophical analysis of the intertwining of legal systems generally and autonomy-conceptions specifically, see also the detailed analysis of Christian Bumke and Anne Röthel, 'Das Paradox Der Autonomy Und Seine Entfaltungen', in *Karsten Fischer*, ed. by Christian Bumke and Anne Röthel (Tübingen: Mohr Siebeck, 2017), pp. 411–35.

4.5 Responsibility and Autonomy in Civil and Criminal Law

Just as the notion of individuals as entities with autonomy (and agency) permeates the legal domain at large, we may find ample reflection in specific areas of the law. In brief, in many cases in which the law assigns responsibility (and/or liability), it does so as a consequence of the exercise of autonomous behaviour, and where autonomous behaviour may not have occurred, limits to responsibility are also often imposed.

Remaining with the concept of law as a framework that outlines acceptable and unacceptable behaviour, there are little domains in which this is more explicit than criminal law, i.e. the area of law that deals with the prevention and sanctioning of societally undesirable acts of sufficient severity. Criminal law and in particular the sanctions it imposes, i.e. any punishment, can be understood to have at least three distinctive purposes: First, punishment may be seen as a retributive element which, in addition to any potential obligations for restitutions, may help to make the injured counterpart of a criminal action "whole". Second, punishment (and in particularly the confinement of a suspected or convicted criminal) has the goal of potentially protecting the society around the individual from again suffering by their hand. Lastly, and most relevant for the purposes of the inquiry at hand is the purpose of prevention of such undesirable acts, as prospective actors are deterred by the likely consequences of their (criminal) actions.[179] Already, the notion of deterrence seems to suggest that the individuals to be deterred are considered to have some sort of reasoning and weighing capacity to compare the negative effects of punishment against the perceived benefit of the undesirable action. We may rephrase this in terms more closely connected to this thesis as individuals are presumed to have the capacity for self-government on the basis of sufficient reflection and consideration and are hence considered to be sufficiently autonomous. This reflects the previously made point that law presupposes individuals' ability to act freely in general terms. It is also important to note that criminal law imposes a gradient of normative assessment that tracks well with the grade of autonomy capacity we may reasonably ascribe to an individual.[180] Individuals that are, for whatever reason, in a state of diminished ability to self-govern (e.g. emotional affect or severe intoxication) are often ascribed reduced or no culpability, shielding them from certain punishment as a result.

A similar notion of responsibility (or liability) for the consequences of exercising autonomy can be found in the domain of civil law (in a wide sense) as well. Assigning tort or damages-related obligations[181] tends to follow the logic outlined just above.[182] Assignation is typically dependent on the individual displaying some sort of autonomy (including realistic alternative options) and particularly intent. But even for risk-based attribution, where intent is not needed,

[179] See e.g. Henry M Jr. Hart, 'The Aims of the Criminal Law Sentencing', *Law and Contemporary Problems*, 23.3 (1958), 401–41 <https://heinonline.org/HOL/P?h=hein.journals/lcp23&i=419>; Henry Weinhofen, 'The Purpose of Punishment ', *Tennessee Law Review*, 7.3, 145–76 <https://heinonline.org/HOL/P?h=hein.journals/tenn7&i=151>; and in particular Helmut Fuchs, *Strafrecht Allgemeiner Teil I*, 10th edn (Wien: Verlag Österreich, 2018), ch. 1.

[180] See e.g. the arguments about free will, autonomy (and a possible distinction between them) in the context of criminal law in Gerben Meynen, 'Autonomy, Criminal Responsibility, and Competence', *Journal of the American Academy of Psychiatry and the Law Online*, 39.2 (2011), 231–36 <https://jaapl.org/content/39/2/231>; Niklas Juth and Frank Lorentzon, 'The Concept of Free Will and Forensic Psychiatry', *International Journal of Law and Psychiatry*, 33.1 (2010), 1–6.

[181] Note for example that this area of law is literally called "law of obligations" (*Schuldrecht*) in the German language, further connecting its autonomy implications to issues of liability and responsibility.

[182] See also the discussion in Omri Ben-Shahar and Ariel Porat, 'Personalizing Negligence Law ', *New York University Law Review*, 91.3 (2016), [i]-688 OP-688 <https://heinonline.org/HOL/P?h=hein.journals/nylr91&i=650>.

attribution still often attaches to a previous exercise of autonomy from which the risk is now understood to be emanating. In other words a previous autonomous action that has incurred a situation in which damages arise outside of the control of the individual may still be attributed to the individual based on the first instance of autonomy exercise.[183] Ultimately, the interlacing of autonomy as a prerequisite of law generally is also clearly visible from a close viewpoint. Legal frameworks attach obligations not just based on innate status but, with respect to regimes of responsibility and liability, on the consequences of an individual's exercise of autonomy.

4.6 Contractual Autonomy and Freedom of Contract

One of the most explicit reflections of human autonomy in the legal domain and an extension of the concepts discussed in the previous section is the concept of contractual autonomy. Such principle is present and usually firmly established in both civil and common law jurisdictions. In essence, it describes the notion that individuals have agency and discretion with respect to the rules that should govern a contract or the conclusion of a contract *per se,* choice of the contracting partner and in particular the content of the contract.[184] Such discretion can be considered as an extension of their own general free agency. This becomes clear when considering when the principle of contractual autonomy is not accepted because of a perceived threat to the actual autonomy of one or more of the parties. For example, choices made under the scope of contractual autonomy may not be accepted by courts in the case that these individual deliberations have been poisoned by fraud, coercion, distress, or similar obstacles to effectuating one's free will. Similarly, in consideration of the difference in market power of the contracting parties, weaker contracting partners are often protected from disadvantageous agreements, e.g. by consumer protection law, or given the right to not be denied service (i.e. contract obligation in the narrow sense). This can be understood as a "levelling of the playing field", in which constraints on the stronger party enable the weaker party to gain an increase in realistic action choices.[185] Finally, consumer protection law also often recognises the limited resources consumers will realistically spend to pierce through complexities of a certain transaction; as a result consumer protection law often protects individuals agreeing but not understanding certain legal consequences, and hence shields them from a loss of autonomy incurred by giving uninformed acceptance.[186]

Similarly, the idea that autonomy is not immediately intrinsic to individuals but requires certain characteristics of the individual in question is reflected in contractual autonomy. Protection of minors and in particular children, as well as protection of people of unsound mind is often

[183] The exemption of this is risk-based attribution on the basis of consideration of benefit of legal status. For example, considering that an object or property represents value and utility for an individual, laws may attribute responsibility emanating from these risks as well (e.g. in the case of operation of inherently risky machines such as cars or planes). (In Austrian legal parlance, this is sometimes denoted as *Guter Tropfen, Böser Tropfen,* i.e. *good drop, bad drop,* signifying the responsibility that comes with the enjoyment of the respective benefits).

[184] Whereas the first and second notion are distinctively different, as pointed out with reference to the different German Terms *Parteienautonomie* and *Privatautonomie* or *Vertragsautonomie* by Patti, see Salvatore Patti, 'Contractual Autonomy and European Private Law', in *Rules and Principles in European Contract Law* (Intersentia), pp. 123–34 <https://doi.org/10.1017/9781780685434.007>.

[185] See e.g. with respect to choice-of-law situations (e.g. forum shopping) Hessel E Yntema, 'Autonomy in Choice of Law ', *American Journal of Comparative Law,* 1.4 (1952), 341–58 <https://heinonline.org/HOL/P?h=hein.journals/amcomp1&i=349>.

[186] See e.g. Margaret Jane Radin, 'The Deformation of Contract in the Information Society', *Oxford Journal of Legal Studies,* 37.3 (2017), 505–33; and in longer form Margaret Jane Radin, 'Boilerplate', in *Boilerplate* (Princeton University Press, 2012).

conducted by assigning them a limited form of contractual autonomy.[187] Contracts entered into by children may be void or will need action by a guardian or other adult person of sound mind. During lucid moments, adult individuals may regain their contractual competency under law, which mirrors their temporary access to the individual competencies of which individual autonomy arises.[188]

4.7 Freedom and Autonomy in International Human Rights Law

4.7.1 Introduction

After the rather cursory overview over autonomy in the legal domain, the following section aims to explore rules within established human rights regimes as they relate to the concept of autonomy and personal freedom, and here particularly in the context of information technology, in greater detail. The main reason for this is that they may be considered to serve as an internationally established consensus in many ways. Insofar the aim to protect individual autonomy, they serve as a strong goalpost against which any autonomy-defining framework may be measured.[189] The next paragraphs will explore the most important rules agreed upon between states and summarize their normative content and their protective scope to the extent that they are relevant for the overarching research objective. Due to the different nature of international law to the general corpus of law, some context is given where appropriate.

4.7.1.1 International Law as a Regulatory Corrective

Originally, international law was thought to be exhausted by the corpus of rules and treaties between states concerning the states themselves. In more recent time, an understanding has been reached that individuals themselves can be addressed by rights originating from international law and might even take actions themselves to ascertain these rights.[190]

The domain of human rights is a resulting sub-field of international law. Consequently, the characteristics of international law, namely its lack of effective enforcement, as well as a more ambiguous syntax apply. Certain "sources" of internationally recognized rules do not exert binding authority but might merely be guidelines or similar "soft law".[191] Additionally, rules of international law are usually not universal *per se* but apply only to those who have subjected themselves consciously, such as by means of signing a treaty or complying with unspoken rules with the understanding that these rules should have legal value.[192]

Taking this into account, the overlap between many of the human right frameworks become understandable. To illustrate, one can compare the instruments within the international bill of rights, namely the Universal Declaration of Human Rights to the International Covenant on Civil and Political Rights, whose Articles 12 and 17 respectively are mostly identical. A key difference between these two instruments from a legal perspective is the non-binding nature of

[187] See e.g. Margaret Isabel Hall, 'Mental Capacity in the (Civil) Law: Capacity, Autonomy, and Vulnerability', *McGill Law Journal*, 58.1 (2013), 61–94 <https://doi.org/10.7202/1013386ar>.
[188] This mirrors the notion of autonomy competency requirements elaborated on in Section 2.3.3.
[189] See in particular the elaboration in Section 1.5. The mentioned comparison will serve as a justification for the pragmatic account of autonomy in Section 5.
[190] Malcolm N. Shaw, *International Law, Sixth Edition*, *International Law, Sixth Edition*, 2008, p.45f.
[191] Daniel Thürer, 'Soft Law', *Max Planck Encyclopedia of Public International Law*, 2009.
[192] See Article 38 para 1, point a and b of the Statute of the International Court of Justice; Tullio Treves, 'Customary International Law', *Max Planck Encyclopedia of Public International Law*, 2006.

the declaration, while the covenant is a binding source of law, with a dedicated entity, the Human Rights Committee, tasked with observation of compliance and enforcement.

4.7.1.2 Excursus - Interpretation of Instruments of International Law

Instruments of international law, such as discussed here are often brief compared to domestic legislation, using sweeping or undefined terminology. Interpretation of treaties follows the rules codified in the Vienna Convention on the Law of Treaties, specifically Articles 31-33, taking into account ordinary meaning of the words used, the object purpose of the treaty and subsequent agreement and practice of states. Recourse can be had to the supporting documents from the drafting stage, the so called "*travaux preparatoires"*, if no meaning can be established. The rules codified within the Vienna Convention on the Law of Treaties connected to interpretation are widely assumed to have reached the rank of international custom and as a result are applicable globally.[193] Treaties can include rules about their interpretation, which take precedence.

4.7.1.3 Excursus - Normative Gradient of Rules of Human Rights Instruments

The rules of human rights instruments lay out rights individuals may exercise or may not be deprived of. This lays out boundaries of a set of values that are intended to be protected by the rule. However, this protection must be balanced against interests of states or the rights of third parties. For this reason most rules allow for limitations of their protection in accordance with a certain standard, by the use of limitation or claw back-clauses. These can be rule-specific or apply in general for all rules of a human rights instrument.[194]

When analysing this as outlined in Annex 4.7.1.3 a gradient is evident. Indeed, there are situations in which the rule itself states (implicitly) that a certain value is worth protecting, all the while not granting protection due to other more exigent factors and allowing interferences. A value might be protected in one situation with a certain set of factors, while in a different situation with a different set of factors that trigger the limitation clause of the rule protection might not be granted. From this it does not follow however that the value itself is considered differently.

When deductively abstracting the boundaries of ethical principles (i.e. personal freedom and autonomy) from legal regimes, it is therefore necessary to primarily consider the value itself, not the resulting protection triggered by an interplay of the protective imperative and the limiting factor of the human rights rule. Conversely, when examining the effectiveness of normative regimes, it is primarily the resulting protection that is of interest.[195]

To illustrate this, we can consider Article 8 of the European Convention on Human Rights which stipulates that everyone has the right to respect for a set of values, and that there shall be no interference by a public authority with the exercise of this right for some enumerated

---

[193] Cf. Mark E. Villiger, *Commentary on the 1969 Vienna Convention on the Law of Treaties*, *Commentary on the 1969 Vienna Convention on the Law of Treaties*, 2008 <https://doi.org/10.1163/ej.9789004168046.i-1058>. p. 439-440, ECHR (1975) Golder v. United Kingdom § 29 (nb. that this was before the Vienna Convention on the Law of Treaties entered into force, indicating the court's opinion that these rules have entered into international custom).

[194] This is the case with the Universal Declaration of Human Rights, in which Article 29 para. 2 creates a blanket limitation for all rights "conferred" by the declaration.

[195] While not fully encompassed by the scope of this thesis, Section 13 will trace the increasingly confident protection of individual autonomy (and particularly decisional privacy) in European legislation. As for its effectiveness, the time of writing may still be to early to arrive at a fair assessment.

reasons. Here, the set of values intended to be protected are private and family life, home and correspondence. These are therefore the main object of inquiry in this chapter.

When addressing effectiveness, the full rule must be assessed. In this case, the limitation clause lays out that interference such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others. It follows that before protection of Article 8 applies, a situation has to pass two filters: (1) an action or inaction of a state must rise to the level of an "interference, and (2) the interference is not justified by law and necessity in the interest of a catalogue of factors central to a democratic society (national security, public safety, economic well-being of the country, etc.). Even actions or inaction that rise to the level of "interference" might not be in violation of Article 8 if they are justified. In practice, this results usually in having to weigh if the reason of justification is sufficient compared to the necessity and intensity of the interference.[196]

4.7.2 Human Rights Instruments of Interest

The theme of human rights has inspired many instruments of international law. These regimes differ by geographical applicability (global or regional) and their binding nature. Some of these regimes have instated enforcement bodies or tribunals that can exercise control over the parties to a treaty, while others have not. In addition, complicated interplay exists between international obligations and domestic legislation. This thesis will rely mostly on the European Convention on Human Rights (ECHR), with occasional reference to the Charter of Fundamental Rights of the European Union (CFEU). The Convention is particularly useful, due to its extensive amount of case law by the European Court of Human Rights (ECtHR). In addition, the "living instrument"-approach, i.e. the expansionary method of interpretation of the convention that is propagated by the ECtHR, has led to the application of the ECHR to situations where new technologies have emerged as a potential conflict factor.

An overview over the major instruments of International Human Rights Law that are somewhat congruent with the instruments discussed here can be found in Annex 4.7.2 but are not all discussed in greater detail in the following sections.

4.7.3 Subsumption of Protection of Autonomy under Existing Doctrine

4.7.3.1 The Trifurcated Protection of Autonomy in Existing Human Rights Law

Protection of autonomy and personal freedom against undue interference by the type of highly sophisticated automated technology is a relatively new field of inquiry; it follows that concrete and explicit legal countermeasures may be lacking. For example, there is no extensive accepted human rights catalogue that accounts specifically for human rights in the Information Age, digital advertising or computer-assisted choice architecture (i.e. "nudging"). However, the arising conflicts can be located within the protective scopes of existing human right frameworks; data-driven targeting of individuals might collide mainly with

(1) the right to privacy (and as an extension, data protection),
(2) the right to freedom of thought, conscience, religion and belief, and

[196] These considerations may inform the ethical viewpoints one adopts when conducting an ethical analysis as outlined in Section 11.4

(3) the right to freedom of opinion, expression and information.

4.7.3.2 The Right to Privacy

Connecting the concepts of privacy and autonomy is the notion that the right to privacy encompasses (partly) the concept of human autonomy and concerns itself specifically with the human being as autonomous subject. The concept of privacy, the link between privacy and autonomy, as well as its legal representation were outlined above in Section 3 and in particular Section 3.4.1.

As noted above, the right to privacy is *inter alia* codified in Article 8 of the European Convention of Human Rights, where it is written:

> *1. Everyone has the right to respect for his private and family life, his home and his correspondence. […]*

As described Article 8 is wider than mere privacy protection, indeed an individual's private life, family life, home and correspondence are protected. While all of these can potentially be affected by agents of the IoE, the strongest topical is found in the protection of private life and correspondence as case law deals with exposure to new technologies primarily in these areas. Surveying the existing case law, connections to the concept of autonomy become visible. (A more extensive overview of Article 8 jurisprudence with respect to emerging technology not directly applicable to autonomy is given in Annex 4.7.3.2)

With respect to reflections of autonomy theory in the right to privacy as enshrined in Article 8 of the ECHR, the ECtHR has found that the terminus "private life" is wide-ranging, not accessible to exhaustive definition, and therefore generally also covers the psychological integrity of a person[197], and that protection may extend to a person's inner life as well, e.g. philosophical, religious or moral beliefs, emotional life. [198] Of course, this also extends to mental health as an expression of an individual's moral integrity. [199] Additionally, personal development falls within the protected scope. This aligns well with autonomy as discussed in the theoretical elaborations above. Psychological integrity and its change processes track closely within the dictus of this thesis and can be understood as a systemic adaptation of internal mental states to be aligned with each other, thereby reflecting ideas of hierarchical autonomy theory.

Factual autonomy constraints, a theme of relational autonomy theory (see Section 2.3.2), on the basis of data processing is also recognized within the case law. For example, the court has dealt with cases in which medical information was at risk to be used when determining an individual's job prospects.[200] Very explicitly, the ECHR found that "the protection of personal data is of fundamental importance to a person's enjoyment of her right to respect for private and family life as guaranteed by Article 8 of the Convention".[201] Consequently, domestic law is to ensure that personal data must be efficiently protected from misuse and abuse,[202] and therefore instate appropriate safeguards to prevent use of personal data inconsistent with

[197] Cf. ECHR (2019), Nicolae Virgiliu Tănase v. Romania [GC], § 128
[198] Cf. ECHR Bensaid v. the United Kingdom, § 47.
[199] Cf. ECHR Bensaid v. the United Kingdom, § 47.
[200] Passim ECHR (2017) Surikov v. Ukraine
[201] Cf. ECHR Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland [GC], § 133
[202] Cf. ECHR (2009) Gardel v. France, § 62

Article 8.[203] This connects closely to the concept of informational self-determination outlined as a subset of autonomy outlined in Section 4.3.[204] By enforcing the absence of potential constraints, such as is the case here by strict data processing regulation, autonomy under a relational framework is strengthened. Very specifically, the ECHR Guide on Article 8 of the Convection also considers identity and autonomy as one of the protected categories of privacy.[205] This holds true beyond the ECHR. For example, privacy under the similarly worded Article 17 of the International Covenant on Civil and Political Rights (ICCPR) is also considered to cover, among other things, autonomy (encompassing in particular "self-realization through actions not interfering with the liberties of others and including bodily autonomy").[206]

4.7.3.3 The Right to Freedom of Thought, Religion and Belief

The right to freedom of thought, religion and belief describes the entitlement of holding and changing beliefs, opinions, considerations or viewpoints. Article 9 of the ECHR outlines this right as follows:

> *1. Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief and freedom, either alone or in community with others and in public or private, to manifest his religion or belief, in worship, teaching practice and observance. […]*

The most commonly adjudicated part of Article 9 is its protection of religion, which is not directly relevant for purposes of this inquiry, however it also deals more generally with thoughts, beliefs, views and convictions. In practice, separating those concepts from religion is difficult, as the very idea of defining religion might lead to undue withdrawal of religious protection from certain individuals, which may explain the expansive scope of the wording of Article 9.

Relevant to the domain of autonomy, the ECtHR has found that an individual's conviction (or thought or view or belief) is protected once it attains a certain level of cogency, seriousness, cohesion and importance.[207] Among such convictions are not merely religious views but also philosophical or deeply personal ones such as pacifism,[208] veganism,[209] medical philosophy with respect to alternative medicine, [210] or the actively reflected absence of religion, such as secularism[211], and atheism,[212] as well as convictions with respect to conscientious objections.[213] Notably, the court has named the rights reflected in Article 9, including freedom of thought

[203] Cf. ECHR (1997) Z v. Finland, § 95
[204] Similar protection of informational self-determination is also granted under Article 8 of the CFR, see Alexander Roßnagel, 'Kein „Verbotsprinzip" Und Kein „Verbot Mit Erlaubnisvorbehalt" Im Datenschutzrecht', *Neue Juristische Wochenschrift: NJW*, 1 (2019), 1.
[205] Cf. ECHR (2019), Guide on Article 8 of the Convention – Right to respect for private and family life, p.20ff
[206] Nowak, p. 385ff; Sarah Joseph, Jenny Schultz, and Melissa Castan, 'The International Covenant on Civil and Political Rights: Cases, Materials, and Commentaries', *The American Journal of International Law*, 2004, p. 558f <https://doi.org/10.2307/3216728>.
[207] Cf. ECHR (2013) Eweida and Others v. The United Kingdom §81 with further references
[208] Cf. ECHR (1978) Arrowsmith v. the United Kingdom (Commission Report), §69
[209] Cf. ECHR (1993) W. v. the United Kingdom (Commission decision)
[210] Cf. ECHR (1998) Nyyssönen v. Finland (Commission decision)
[211] Cf. ECHR (2011) Lautsi and Other v. Italy [GC] § 58
[212] Cf. ECHR (1986) Angeleni v. Sweden (Commission decision); (1994) Union des Athées v. France (Commission report) § 79
[213]Cf. just ECHR (2020), Factsheet – Conscientious objection, available at https://www.echr.coe.int/Documents/FS_Conscientious_objection_ENG.pdf

and conscience, a "precious asset".[214] This mirrors the idea of sound mental state integration, as brought forward by Sneddon (see Section 2.2.4).

The right codified in Article 9 provides protection for both holding a religion or belief and manifesting it. This phrase omits the terms thought and conscience, however, the belief in question must not be religious. Protection of internally maintaining a conviction is absolute and unconditional, that is no coercive steps may be taken to change the conviction by a state. Protection of the right to manifesting one's conviction is less pronounced; as it (and only it) is subject to additional limitations (e.g. as set out in Article 9 § 2).[215]

Again jurisprudence also directly links issues freedom of thought and belief to autonomy. Indeed the court explicitly notes the connection as follows:

"The freedom to accept or refuse specific medical treatment, or to select an alternative form of treatment, is vital to the principles of self-determination and personal autonomy. A competent adult patient is free to decide, for instance, whether or not to undergo surgery or treatment or, by the same token, to have a blood transfusion. However, for this freedom to be meaningful, patients must have the right to make choices that accord with their own views and values, regardless of how irrational, unwise or imprudent such choices may appear to others."[216]

Article 9 also aims to protect the mental integrity of individuals in certain cases, as is visible in the protection of individuals against religious indoctrination by a state.[217]

An additional survey of jurisprudence relating to Article 9 can be found in Annex 4.7.3.3

4.7.3.4 The Right to Freedom of Opinion, Expression and Information

The right to freedom of opinion, expression and information is often expressed as "freedom of speech" (aided by the title of Article 10 that only mentions freedom of expression), but it does entail a more extensive scope than just protection from sanctions of the state for expressing opinions. Outlining this principle, Article 10 of the ECHR reads:

- *Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.*

Freedom of expression applies primarily to the expression of information and ideas as a function of one's self-fulfilment. [218] However, crucially Article 10 of the ECHR does not only cover the imparting of information but also the right of the public to receive it[219], as well as the choice not to express oneself, i.e. abstain from an expression.[220] Protection is given not only to

[214] Cf. ECHR (1993) Kokkinaki v. Greece § 31
[215] Cf. ECHR (2007) Ivanova v. Bulgaria § 79
[216] ECHR (2010), Jehovah's Witnesses of Moscow and Others v. Russia, 36
[217] Cf. ECHR (1986), Angeleni v. Sweden (Commission Decision), (1996) C.J. J.J. and E.J. v. Polan (Commission Decision)
[218] Cf. ECHR (2011), Palomo Sanchez and Others v. Spain [GC], §53
[219] Cf. ECHR (1991), Observer and Guardian v. the United Kingdom, §59
[220] Cf. ECHR (1994), Stohal v. Austria (Commission Decision)

the content of expression but also to the means of dissemination.[221] Very explicitly (despite missing from the title of the article) the right to freedom to hold opinions is mentioned in the article as well. According to the court, Internet, in light of its accessibility and its capacity to store and communicate vast amounts of information, plays an important role in enhancing the public's access to news and facilitating the dissemination of information generally.[222] However, this is also seen as a potential risk, as the risk of harm posed by content and communications on the Internet to the exercise and enjoyment of human rights and freedoms, particularly the right to respect for private life is higher than the one posed by traditional press.[223]

There is little jurisprudence that explicitly connects the concepts of autonomy and rights under Article 10 of the ECHR. However, aside from the most prominent feature of said right that is the freedom of expression, the right to hold opinions connects very well with general concepts of autonomy. Protection of held opinions can be seen as a prerequisite of a right of informational self-determination, itself a part of autonomy (see Section4.3); if information processing should be limited because it could lead to external actors effectively influencing opinions, those opinions must be protected to be begin with. (Similar argumentation is available with respect to the protective scope of the rights codified in Article 9 of the ECHR.) But a connection to an individual's autonomy is also visible on a more abstract level. The protection of a held opinion can be easily understood as the protection of some assortment of mental states, or alternatively as a self-expression of identity and its self-realization, both of which connects to contemporary autonomy theory.

4.8 Analysis and Conclusion

As outlined in Section 4.2, it is extraordinarily difficult to transfer concepts between different scientific domains. This holds true for the analysis of the value and recognition of autonomy concepts between philosophy and law as well. However, with awareness of this limitations, this section has argued that there is indeed a common thread between legal and philosophical notions of autonomy. Careful analysis of jurisprudence and literature can reveal that international human rights regimes may rely on very similar principles and assign importance to similar factors of an individual's self-determination and self-realization.

This section has argued that some concept of autonomy permeates the legal domain; from being a prerequisite in the most abstract structure of law as well as being reflected in specific singular rules. The very notion of law as a regulatory framework that is armed with sanctions and the corresponding concept of personal responsibility in civil and criminal law seems based on the concept that its subjects are self-governed entities. Similarly, contractual autonomy can be understood as an expression of an individual's autonomy as an actor within the legal sphere. Lastly, the connection seems to hold in the domain of international law as well, with seems especially important due to its function as an international consensus. Comparing the rights guaranteed by international regimes with respect to privacy, freedom of thought and freedom of opinion, it should have become clear that their protection is compatible with the theoretical-philosophical frameworks that were discussed in Section 2.1. For example, Article 8 of the ECHR guarantees protection of one's psychological integrity, personal beliefs and personal development, as well as one's physical and moral integrity sphere of relations to and between

[221] Cf. ECHR (2012), Ahmet Yildirim v. Turkey § 50
[222] Cf. ECHR (2012), Ahmet Yildirim v. Turkey § 54
[223] Cf. ECHR (2011), Editorial Board of Pravoye Delo and Shtekel v. Ukraine § 63

other people (and oneself). Article 9 of the ECHR guarantees protection of religious and non-religious convictions including more philosophical, and manifestations of such beliefs, even if they are irrational, setting a strong threshold against paternalism in favour of one's autonomy. Again, mental integrity is protected as evidenced by the prohibition of religious state indoctrination. Even more explicit is Article 10 of the ECHR: "the holding of opinions" is protected under the regime. The same themes run through other human rights regimes as well. For example, the ICCPR's and its co-developed Universal Declaration of Human Right's right to privacy, also encompasses autonomy. Article 8 CFEU also extends to informational self-determination. As is becoming evident, these valuations are not just done in favour of a strictly hierarchical view of autonomy either, although the comparisons are often most salient within such a framework. The enforcing bodies such as the European Court of Human Rights do not just protect mental integrity of an individual, there also seems to be quite explicit consideration of factual constraints as they are recognized in relational accounts of autonomy. In particular with respect to emerging technology, there is real concern about potential implications of what may happen once an entity has accrued a sufficient amount of data. In a way, and as the literature to informational self-determination shows, data collection is not necessarily seen as harmful *per se* but needs curtailing due to the enabling effect on third parties to impose autonomy constraints. Crucially, despite a still persisting lack of case-law, the enforcing bodies have recognized the potential impact of informational agents with respect to human autonomy, sometimes even explicitly, in connection with the emerging technology of the IoE.

Finding these reflections of both autonomy and potential conflicts in light of technological change in the legal domain serves two purposes. Firstly, it grounds the previously discussed philosophical theories and gives them a sort of empirical relevance. As an extension, issues that are highlighted by such theories are given salience if they are recognized by the legal domain directly or indirectly. Secondly, due to the origin and function of law we can derive importance of certain elements of the concept of autonomy that are considered widely relevant. These concepts, that connect closely to the (intuitive) main points of philosophical theories should therefore be reflected in the pragmatic account of autonomy that will be attempted in Section 5.

5 A Pragmatic Account of Autonomy

5.1 Introduction

Within this section, a framework to understand autonomy constraints, denoted as the pragmatic account of autonomy, is introduced. This framework serves as the answer to Research Question 1 and Research Sub-Question 1.4, which asked how autonomy can be conceptualized to be relevant and salient for investigating emerging issues of the IoE and what the main obstacles to conceptualize autonomy are generally and in the context of information technology. The section is structured as follows: Section 5.2 elaborates the use-case and methodology of this theoretical framework while Section 5.3 defends the methodology and submits three major axes of justification on which the theory ought to be evaluated: utility, legal plausibility and philosophical plausibility. Section 5.4 then outlines the structure of the theoretical framework in depth, identifying autonomy constraint as constraints on three aspects of an individual's autonomy, namely their intrinsic, relational, and informational autonomy. Section 5.5 and 5.6 reflect further on this structure and the decision to focus mostly on autonomy constraints. Section 5.7 briefly touches upon the concept of positive and negative freedoms under this account and how they are reflected in this framework. The chapter concludes in Section 5.8 The pragmatic account of autonomy will be used for the rest of the thesis when discussing autonomy constraints.

The remainder of this chapter is dedicated to fleshing out this framework and providing methodological justification for the approach taken.

5.2 Use-Case and Methodology of the Pragmatic Account of Autonomy

I have argued before that elements of conflicting autonomy theories are highly salient with respect to understanding interference in an individual's capacity to self-governance by informational agents. However, as the above sections have highlighted, there is no consensus about how to conceptualize autonomy. A strictly formalistic account of autonomy may benefit from intellectual elegance. However, as the feminist-relational critique of structural autonomy theories has highlighted, overly purist autonomy accounts run the risk of excluding situations from their scope that are intuitively considered to be an autonomy issue. Similarly, legal frameworks evidently provide protection under the umbrella of guaranteeing autonomy to a wider extent than just the protection of intrinsic mental states, as will be shown below. The practical relevance of an account of autonomy instead increases with its ability to track issues that become salient both (a) from intuitive concerns of individuals, and (b) from societal concerns at large. As outlined in the introduction, autonomy is seen as being under threat by many contemporary commentators and scholars. Utilizing a framework of autonomy that excludes the phenomena that are so publicly decried does not yield the most amount of relevant and useful knowledge.

Therefore, in order to analyse the effects of emerging technology within the domain of the IoE, that is to map the issues the emerging IoE could impose on an account of autonomy, the aspects of the previously discussed theories of autonomy need to be extended and combined. This section aims to do that by introducing what will here be denominated as the pragmatic account of autonomy. The goal of this is not to satisfy all logical, normative, empirical or other criteria that would make it a fully consistent philosophical theory. Instead, this account is meant to

serve as a toolset to denominate risks to autonomy as they are caused by the emerging IoE.[224] It is the issues that are brought up by a legion of contemporary authors, seemingly on the basis of intuition instead of a dedicated analysis of human autonomy, that this pragmatic account aims to enable access to as well as to create a basis for subsequent analysis in the later parts of this thesis.

5.3 Methodical Justification and Objections to the Pragmatic Account of Autonomy

The model of autonomy outlined in the second half of this chapter, i.e. the pragmatic account of autonomy, is justified by three different means:

- The particulars of my proposed account are *useful* to analyse and address emerging situations that are caused by informational agents; and
- The particulars of my proposed account are *commonly derived* from established philosophical theory, and/or
- The particulars of my proposed account are made *plausible* by common legal concepts, in particular widely accepted case-law of the human rights domain.

I will argue the relevance of the proposed model with respect to emerging issues connected to the IoE by illustrating it with (non-exhaustive) examples of common practices of informational agents, so widespread that they require no additional reference. A logical-philosophical plausibility check will be done mostly by reference to previously elaborated theory, i.e. with the general notions of hierarchical and relational autonomy theory. [225] A jurisprudential plausibility check will be conducted by reference to selected legal sources, and in particular to relevant decisions of the European Court of Human Rights.[226]

While the first two justifications are somewhat self-explanatory, the third requires further elaboration.[227] To this end, I also recognize the following two objections to such an approach and will try to address them before presenting the framework: Firstly, (case-) law as presented in the decisions outlined below does not explicitly define autonomy and can therefore not be

[224] Compare the discussion of the advantages of such an open-ended approach in Nissenbaum and Regulating at the End of Privacy, p. 221. Just like Nissenbaum with respect to her subject of privacy, the author considers the task of (exhaustively) defining autonomy via a complete theory a "likely hopeless ambition". Cf. H. NISSENBAUM, Privacy in Context (Stanford University Press, 2009) p.189.

[225] However, the validity of the pragmatic account of autonomy is not limited to instances where it reflects existing theory. Instead, such references are made to disclose openly the sources of inspiration as well as highlight the perceived need by scholars to conceptualize phenomenon on the respective autonomy dimension described.

[226] International law, and more specifically human rights regimes are very well equipped for this task, as they represent a wide consensus not just between individuals but also between states. With that said, arising conflicts can be located within the protective scopes of existing human right frameworks; data-driven targeting of individuals might collide mainly with (1) the right to privacy and more specifically data protection, (2) the right to freedom of thought, conscience, religion and belief, or (3) the right to freedom of opinion, expression and information. For this inquiry, focus will be had on the European Convention of Human Rights, namely Art. 8, 9 and 10, due to the abundance of case law available. The following paragraphs are dedicated to give a specific overview over case law that is relevant to potential infringements to autonomy in the meaning as previously explored. Given the great overlap between different instruments of international law, these findings can likely be extrapolated to other treaties as well.

[227] For highlighting the need of this section, I am indebted to an anonymous reviewer affiliated with the 2020 ACCA conference.

used to justify any approach to human autonomy, and secondly terminology cannot be transferred between the legal and philosophical domain.[228]

Most importantly, the aim is not to *derive* or *endorse* specific autonomy concepts from legal sources but to find autonomy concepts that spring from theory (or intuition) reflected and protected within the legal domain. On a high level of abstraction, this is not an issue. The concept of autonomy in some form or another permeates the legal domain. Indeed, the etymology of the word autonomy makes clear that it was first used in a legal context. In some way or another, any autonomy model can arguably draw from a rich context within the legal domain to justify its assumptions and delimitations.

In this thesis, mainly case law from the domain of international human rights is used due to its special status as a (semi-)global consensus on rights of an individual.[229] These draw importance and (at least empirically) legitimation from their widespread adoption and acceptance through parties to the regimes and their application by international tribunals, governments or domestic courts.

At the same time, reliance on case-law based on such human rights regimes for the purposes of this thesis has disadvantages as well. Like many instruments of international law, treaties such as the European Convention of Human Rights use sweeping, often undefined language, requiring specific methods of interpretation.[230] This can make it difficult to pinpoint the exact content and scope of a given provision that ought to protect a human right that may be relevant to an individual's autonomy. At the same time, there is tension when crossing concepts from the legal sphere and the analytical-philosophical sphere. After all, any case-law within the human rights domain deals with legal problems using legal instrumentation to come to a legal outcome. Case-law, and its underlying treaty provisions, its terminology and concepts can therefore not always be translated directly into a theoretical model such as is propagated here. Some thoughts on why this does not defeat the attempt undertaken here are given at the end of this thesis.

To this end, making the pragmatic account of autonomy *plausible* via reference to case-law cannot mean to find exact corroboration of specific theoretical concepts taken from the domain of philosophy. Indeed, the provisions codifying human rights and respective case law may be applicable to a variety of autonomy concepts, that may be incompatible with each other. It follows that the limitations of such a plausibility check must be kept in mind: legal sources generally and case-law in particular will usually not endorse a certain autonomy theory over another. However, these sources can help to visualize that some (meta-)physical phenomenon (e.g. agent behaviour, constraints, etc.) that are considered of note by autonomy theorists are considered to be relevant, and in many cases worthy of protection from a legal perspective as well. Ultimately, this is less an exercise of finding concrete representations of specific theories of one domain in the other, but more of finding general patterns that persist between both domains.

---

[228] For these, I am grateful for the suggestion of an anonymous reviewer affiliated with the 2020 ACCA conference.

[229] Naturally, this space is not uniform. The collective corpus of values undergirding what is currently understood as encompassed by human rights have inspired many instruments of international law. These regimes differ by geographical applicability (global or regional) and their binding nature, but also to a lesser extent in their content and scope. For example, the African Charter on Human and People's Rights does not contain a provision protecting an individual's privacy, while being very congruent with other human rights regimes otherwise.

[230] As codified in Art. 31-33 of the Vienna Convention on the Law of Treaties.

This is further complicated by the fact that in the legal domain there is a distinction between granted protection and the intention behind the respective protective scope: e.g. in the domain of human rights, rights of an individual will be balanced against interests of states or the rights of third parties. For this reason many such rules allow for limitations of their protection in accordance with a certain standard, by the use of limitation-, derogation- or claw back-clauses.[231] However, even when protection is not extended due to a value conflict, this does not imply protection was not intended to apply to a situation more generally. Put differently, there are situations in which the rule itself states (implicitly) that a certain value is worth protecting, all the while not granting protection due to other more exigent factors. A value might be protected in one situation with a certain set of factors, while in a different situation with a different set of factors that trigger the limitation clause of the rule, protection might not be granted. From this it does not follow that the value itself is considered differently. Put bluntly: even if an interference with an individual's autonomy (or privacy, etc.) is justified and in compliance with the law via an exemption or "claw-back"-clause (e.g. due to other parties' interest), this has no bearing on the notion that the individual's autonomy (or privacy, etc.) generally was considered to be worthy of protection. Case-law *in concreto* usually highlights some sort of conflict between different values of different parties, but the recognition of said conflict already presupposes that the value is of importance in abstract.[232]

Insofar case-law is used to make plausible the pragmatic account of autonomy propagated here, it serves the latter purpose, not the former. For example, when an international tribunal finds that the right to privacy extends to psychological integrity of an individual, this does not entail the court taking a stance towards whether the inherent psychological state of an individual is part of their autonomy, but merely that their psychological integrity is (to an extent) worthy of protection.[233] Of course, theorists may connect the psychological state of a person with the concept of individual autonomy explicitly, and simultaneously consider it worthy of protection.[234] Similarly, there seems to be some intuitive concern among commentators and the like that intrusions of an individual's psychological integrity are to be considered harmful to their autonomy. The pragmatic account of autonomy then "closes" the circle and stipulates that within the model psychological integrity is relevant for an individual's autonomy, and that the relevance of such inclusion (and as a result of the model) is given both by these theoretical and legal concerns over the underlying (meta-)physical phenomenon.

This thesis will rely mostly on the European Convention on Human Rights (ECHR). The Convention is particularly useful, due to its extensive amount of case law by the European Court of Human Rights (ECtHR). In addition, the "living instrument"-approach, i.e. the

[231] C.f. for a in-depth differentiation between such provisions R. Higgins, 'Derogations under Human Rights Treaties', *British Yearbook of International Law*, 48.1 (1977), 281–319 <https://doi.org/10.1093/bybil/48.1.281>.. (Indeed the term claw-back was first used by this author in this context).

[232] As put forth concisely in the statement "the operation of law depends on the existence of extralegal values", c.f. Yehezkel Dror, 'Values and the Law', *The Antioch Review*, 17.4 (1957), 440 (p. 453f) <https://doi.org/10.2307/4610000>.

[233] Of course, the court may take such a stance, as done by the German *Bundesverfassungsgericht* in its landmark decision on informational self-determination, where certain intrusions where considered to be harmful to an individual's self-determination capability (which serves as an obvious sub-set of autonomy, i.e. the exercise of autonomy over their personal data), see BVerfG, paras 93f, 146f.

[234] Indeed, this need for protection stems not just from a theory in question in particular; e.g. a theory that considers mental state congruence relevant for individual autonomy does not necessarily imply that these mental states ought to be protected. For this, another assumption must be taken, which propagates the value (and worthiness of protection) of human autonomy in general.

expansionary method of interpretation of the convention that is propagated by the ECtHR, has led to the application of the ECHR to situations where new technologies have emerged as a potential conflict factor.[235] But while this inquiry draws mainly upon a specific set of human rights case-law, this is not the only venue that is open to this approach. As mentioned before, some concept of autonomy permeates law in many ways. Maybe most explicitly, contractual autonomy (both with respect to deciding on the content of a contract and also with respect to deciding on the contracting party) reflects the concept of decisional autonomy rather explicitly.[236] But this point may also be made at a more general level: Legal notions of responsibility (and liability as a consequence) seem to presuppose some idea of individual autonomy as well. In situations where theorists (or the pragmatic model) would assess that one's autonomy is diminished, legal systems tend to limit responsibility as well (e.g. exigent circumstances, self-defence, duress). An even more abstract reflection of autonomy may be found in the primary function and form of law as a regulative instrument itself, e.g. Hans Kelsen famously argued that law presupposes free will.[237] Of course, taking such sources as venues of justification for the pragmatic account leads to many of the same challenges as taking recourse to the domain of human rights law. A careful separation of concepts of the legal and philosophical domain is necessary in any case, and the same limits to transferability between these spheres apply. In any case, a full survey of these concepts is beyond the scope of this thesis and subject to further exploration. However, the existence and validity of these arguments drawing from additional legal sources to find commonality with the main ideas undergirding autonomy theory is compatible with the pragmatic account as a whole. Ultimately, further reflection of what is intuitively or explicitly considered to be relevant for human autonomy by legal professionals, commentators, theorists or laymen only serves to strengthen the need for an open-ended system such as is suggested here.

To summarize, one can circle back to the main objections presented back at the beginning. It is true that there is little indication that any source of law has had the goal of defining autonomy in the first place. However this is no obstacle to analysing the results of such normative safeguards, and its corresponding court decisions and, by tracing them, outlining an approximate picture of what we perceive to be worth protecting. Cross-referencing this with the widest possible interpretation of self-governance gives us a useful map of autonomy as we deem it relevant in the current times. In addition, given the highly technical nature of philosophical inquiry into autonomy, it is nearly certain that any legal authority deciding within this wide scope of what ought to be considered as autonomy-relevant, and perhaps even using the term autonomy does not consciously adapt its writing to any particular account of autonomy. In this respect we may see them analogously to the commentators highlighted in Section 1, which intuitively locate an issue that connects with autonomy. But again, no precise intention is needed, as the very objective of the approach of this thesis is to map such intuitive autonomy concerns, be it expressed in popular or legal language onto a model of autonomy. (It is not the aim of this model to derive philosophical concepts from case-law, but to show that case-law considers certain (meta-)physical phenomenon to be worthy of protection, that are also considered relevant from a theoretical point of view.) In addition, most of the dimensions

[235] See for a detailed discussion of this approach George Letsas, 'The ECHR as a Living Instrument: Its Meaning and Legitimacy', *Constituting Europe: The European Court of Human Rights in a National, European and Global Context*, 2 (2013), 106.

[236] Whereas the first and second notion are distinctively different, as pointed out with reference to the different German Terms Parteienautonomie and Privatautonomie or Vertragsautonomie by Patti, see Salvatore Patti, 'Contractual Autonomy and European Private Law', in Rules and Principles in European Contract Law (Intersentia), pp. 123–34.

[237] C.f. Hans Kelsen, 'Causality and Imputation', *Ethics*, LXI.1 (1950), passim. N.b., that, while not of issue here, in theoretical discourse free will and autonomy are not necessarily fully congruent

of the pragmatic account can be derived both from philosophical and legal frameworks. Hence, while the objections are valid insofar that they highlight important challenges to the general transfer of knowledge between the legal and philosophical domains, they do not inhibit the purpose of goal of the pragmatic account of autonomy within the limits that have been outlined above.

5.4 Structure

A couple of limitations, caveats and general comments must precede the conceptualization attempted here. Within this chapter, I do not aim to define autonomy exhaustively. Instead I propose to divide the not definitively defined corpus of what it means to be self-governed as a bundle of dimensions of autonomy, whereas these dimensions function as a viewpoint of sort that are sensitive to different constraints. Autonomy as a whole is inhibited if there is an inhibition within one of these dimensions. To anticipate the following text, we may consider someone to be under an autonomy constrain if they are subject to misinformation provided by an informational agent e.g. with respect to their eligibility to vote. Similarly, we may consider someone to be constrained in their autonomy if they are assigned a social credit score (unbeknownst to them) by an algorithm and because of this (perhaps accurate) score, they are denied a financing loan to purchase a house. It is one thing to denote this generally as an autonomy constraint, but there is clearly a difference between these two cases. *In concreto*, the pragmatic account displays that both of these situations constrain the individual on different dimensions, here on the dimension of their informational and their relational autonomy.

Previous theory has distinguished between the autonomy of actions and the autonomy of individuals. The pragmatic account is action-centric, i.e. mostly concerned with so-called local autonomy. Of interest is thus primarily the assessment of an agent's individual decisions, and, only as a function of these collective data points, could one determine the state of an agent as generally autonomous or not. Considering a situation under the pragmatic account hence means considering a certain action or interaction or a certain choice.

The pragmatic account is not meant to be understood as inherently binary. Previous analysis has shown that there is a trade-off between overly sensitive and overly blunt assessment-strategies to classify situations that raise intuitive concerns. A constraint (on a certain dimension) of the pragmatic account is recognized as displaying a certain intensity. Clearly, not every contact with misinformation purveyed by an algorithm warrants denoting an individual as not autonomous. Similarly, contact with misinformation that is not very effective may have limited impact on any decision made or action taken, even if the mere exposure has lingering effect. It makes sense then to understand autonomy on a "sliding scale" and displaying a gradient of autonomy. This means that under this approach, a decision or action is not taken either fully autonomously or not. Instead, one may consider a decision or action to fall somewhere on a spectrum in between a theoretical maximum (full autonomy) and minimum (no autonomy). This allows a more effective differentiation (or comparison) between constraints which are similar in nature but of different intensity.

The pragmatic account introduces three major dimensions on which individual autonomy can be constrained. Importantly, the dimensions overlap and can be used to assess each analysed instance on dogmatically different levels within the same model. Put differently, any situation can be assessed from different analytical perspectives at the same time, and any phenomenon that affects an individual's autonomy can do so in manifold ways. An action by an informational agent towards an individual (human) agent (e.g. the automatic action of shadow-

banning[238] that individual) can have implications of different nature (e.g. transparency, actual impossibility of interaction) for their autonomy at the same time. In terms of the pragmatic model of autonomy, this means that an individual is constrained on different dimensions of their autonomy simultaneously.

Because of this overlap, there is no hard limit on the number of dimensions one may find useful to analyse with respect to a certain research objective; the approach propagated here is deliberately open-ended. This section merely outlines the three most important dimensions as they pertain to autonomy as potentially constrained by IoE agents in the opinion of the author. These main pillars of these models are the intrinsic dimension, the relational dimension and the transparency dimension on which this section will focus.

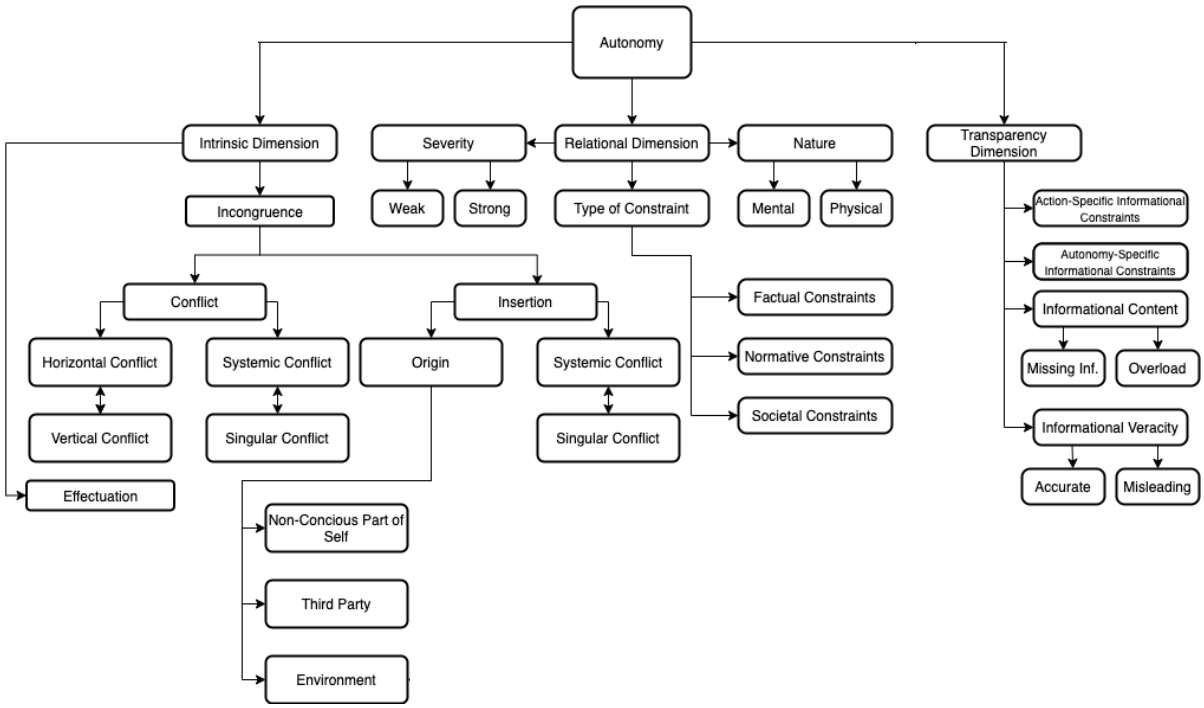

Figure 2 - Selected Dimension of the Pragmatic Account of Autonomy

5.4.1 Intrinsic Autonomy

Autonomy on its intrinsic dimension is concerned with internal endorsement of mental states by other mental states and their effectuation.[239] An individual displays intrinsic autonomy if their wants, preferences, or desires are in a state of coherence to each other and the actions the individual takes is brought about by these coherent wants, preferences, or desires. Consider the following example:

Example 1.1 *A has contemplated buying a tent for a long time. After some adequate analysis and research narrowing down his selection, he ends up purchasing one tent that fits his needs.*

[238] Shadow-banning denominates the action of restricting access of other users of a platform to communication material (e.g. posts, comments, etc.) put out by the "shadow-banned" individual without their knowledge. This leads to diminished interaction between the individual and the wider platform community.

[239] The term effectuation is here used to describe the realisation of a mental state; e.g. the effectuation of a desire to own a product may be to buy that product. This dimension assumes a structural-hierarchical system of mental states as suggested by most contemporary autonomy theorists.

From an intrinsic autonomy standpoint, the above (as presented) is unproblematic. The decision was informed and based on consideration, the purchase in line with his preference. As a result, there is apparent congruence between his preference and his action.
In contrast, the intrinsic dimension of an individual's autonomy is inhibited by mental state incongruence or the lack of effectuation of certain mental states. Such incongruence may be caused by

- conflicting (already integrated) mental states.
- mental state insertion without sufficient integration, where a certain desire may be introduced to an individual and is acted upon without fully aligning with the individual's other mental states.

This text will consider both of these in turn. As the pragmatic account of autonomy is action-centric, it considers the autonomy of (realized) actions. Mental state conflicts result in diminished intrinsic autonomy if the conflict is realized by taking a specific action (i.e. acting on behalf of a conflicted mental state).

5.4.1.1 Conflicting Mental States

Adapting the initial example, we can describe a conflict of mental states.

Example 1.2 *A would like to buy a tent immediately because they want to increase their time spent outdoors camping as soon as possible. At the same time, A wants to live frugally and stick to a budget they made earlier, which does not allow this expenditure.*

In the above situation, the preference to purchase and to not spend money are in conflict. Purchasing means that the action is incongruent with one of their preferences, not purchasing means it is incongruent with the other. In situations of true mental state-conflict, autonomy of all potential actions is diminished.[240] As a result, creating these situations decreases intrinsic autonomy.

Example 1.3 *A wants to live frugally and stick to their budget. As a result of exposure to videos on a social-media platform that praise the experience of spending time outdoor camping,*[241] *they develop the wish to purchase a tent immediately.*

As is apparent from this example, the introduction of further information (perhaps playing on some deeper preference for hedonism or social status) can create new conflicts. In this case, we can attribute the mental state incongruence (and the resulting diminished autonomy) to an interaction with information technology. To solve this conflict and to validate any actions made

[240] In real-life situations, because of the complex system of desires and preferences of individuals, "full" intrinsic autonomy is unlikely.
[241] See for some potential impact of social media content that aims to connect with specific (more abstract) preferences such as a preference of novelty, acquisition of cultural knowledge, entertainment, prestige or escape, and for some discussion of the effectiveness of such content e.g. Christina Katsikari and others, 'Push and Pull Travel Motivation: Segmentation of the Greek Market for Social Media Marketing in Tourism', *Sustainability*, 12.11 (2020), 4770; Orhan Icoz and others, 'Social Media and Consumer Buying Decisions in Tourism: The Case of Turkey', 2018; Wei-Tien Hung and Guo-Bao Liou, 'The Influence of Social Media on Outdoor Recreation Participation', 2022.

on the basis of these preferences, the individual can affirm one of the mental states, i.e. recognize its relatively increased importance over the conflicting mental state. Consider the following addition to Example 1.3

Example 1.3 (cont.) *After careful consideration A decides that their preference to live frugally is derived mainly from a higher-level preference for financial safety, and that their preference to purchase a tent and spending more times camping outdoors (while having seemingly from online videos) is mainly connected to his preference for having a close connection to nature. A realizes that when comparing these two preferences, the close connection to nature is more important (and more in-line with their other identity attributes) to them, so they purchase the tent.*

As shown, upon further inspection, mental state-conflicts may be defused by discovering (or assigning) a relative ranking between each other.[242] If such relationship exists, intrinsic autonomy is not necessarily constrained.

5.4.1.2 Mental State Insertion

Mental state insertion can constrain individual autonomy, if the inserted mental state is in conflict with established mental states. There are multiple ways from which mental states can originate.

5.4.1.2.1 Insertion by Foreign Entity with Attributable Agency

The easiest insertion of a mental state to conceptualize (and the most relevant for this thesis) is the insertion by another individual, group of individuals, informational agent or similar entity that can be attributed some sort of agency. [243]

Example 2.1 *After seeing an online advertising for a car that was selected and shown to him by an algorithm that selected the advertisement based on a profile of him, B feels a strong want to own his own vehicle. B purchases a car as a result. Prior to seeing the online advertising, B did not have this preference.*

Within the IoE, oftentimes there is some sort of motivation or intent behind this attempted mental state insertion. For example, (online) advertising in general can be understood as attempted mental state insertion, that is the attempt of creating a first- or higher order mental state to induce a certain action. Naturally, not all advertising has to be autonomy-undermining.

Example 3.1 *C is planning to quit smoking because of preference to live a healthy life unaffected by nicotine addiction. After seeing an advertisement on a social media platform for it, they purchase an (effective) nicotine patch. Prior to the advertisement, they would not have considered this purchase.*

In this example an individual that is (through a second-order mental state) planning to quit smoking is not undermined by personalized and targeted advertising for products or strategies

[242] Acting without this information is arguably a sign of diminished autonomy. If A purchases a tent on a whim instead of acting upon their conscious decision fuelled by reflective considerations, it seems unreasonable to assign this decision heightened autonomy, even if they would have come to the same decision in either case.

[243] For the peculiarities of informational agents versus non-agent entities see Sub-Section 9.2. In the parlance of Sneddon, this would be a case of *oudenonomy* and *cosmonomy*, see footnote 39.

that can help him override the (first-order) desire to smoke. In this case the individual already desires to have no want to smoke, an IoE-process enabling them and circumventing the conflicting first-order desire of wanting to smoke (e.g. due to addiction) is ultimately supportive of an individual's autonomy. Conversely, the opposite can also occur.

Example 3.2 *C has a preference to live a healthy life unaffected by nicotine addiction. After seeing an advertising on a social media platform, showing individuals consuming nicotine products as particularly desirable, C starts purchasing and consuming the product despite his overarching preference.*

In the counterexample, the action of consuming nicotine products is clearly in conflict with his higher-order preferences. If due to the exposure to persuasive information, the individual now goes against their higher-order preference in order of a lower-order preference (at first e.g. connected to social status, later due to cravings or peer pressure), the mental state insertion has created a conflict and diminished the individual's intrinsic autonomy. In both cases one can locate the origin of the mental state insertion as coming from a third party

5.4.1.2.2 Insertion by External Factors and Environment

Mental state insertion can also originate from environmental factors that cannot be (subjectively) ascribed agency (as opposed to other individuals or certain IoE-processes). Consider the following example.

Example 2.2 *After watching live-streams of traffic webcams online, B feels a strong want to own his own vehicle. B purchases a car as a result. Prior to seeing this footage, B has not had this preference.*

As opposed to example 2.1, there is no intent behind the insertion and no information provision by an entity that can be described as an agent (or as autonomous) between the provision of the information that inserts the want for a certain product in the individual. Instead it is the mere features of environmental factors (here observed through a digital lens) that inserts the mental state. Within the digital domain, environmental or external factors can encompass both the actual (physical) environment, and the digital representation of it.

5.4.1.2.3 Self-Insertion

It is not implausible that a novel "inserted" mental state originates from the individual themselves. This may be the case for "dormant" or "emerging" preferences or hidden conflicts of preferences, upon their activation the newly introduced mental state is immediately in conflict with established preferences.

5.4.1.3 Conflict Orientation

Conflicting mental states can be divided between with respect to their underlying relationship. In particular, a conflict can be a

- horizontal conflict, or
- vertical conflict.

Mental states will typically display some sort of dependence to each other. If these dependencies take the shape of a hierarchical relationship, (i.e. one mental state encompasses a preference about holding or acting upon another mental state), the hierarchically higher

mental state is of higher order. Higher-order mental states do not need to exhaust themselves by strictly meta-preferential content, i.e. being solely about the endorsement of another lower-order mental state. The endorsement of lower-order mental states can be implicit.

5.4.1.3.1 Vertical Preference Conflict

The archetypical type of conflicting mental states is a conflict in which one is subsumed in scope and consequently in conflict with the other mental state. These types of conflicts are vertical conflicts as they span different levels of a conceptual hierarchy of the mental states in question.

Example 3.3 *C quit smoking due to conscious preference to abstain from addictive substances for health reasons. Nonetheless, C still craves cigarettes when they are reminded of them. A contextual recommender algorithm embedded in a mapping and navigation service suggests, based on previous data, nearby tobacco shops. Based on this prompt, C gives in to his cravings and consumes cigarettes.*

In this example, there is a self-imposed conflict between a higher-order mental state, i.e. the preference to abstain from addictive substances, and a (self-induced) first-order mental state, i.e. the primary craving for nicotine. These preferences are in a hierarchy: the preference to abstain from addictive substances implicitly encompasses the preference to not smoke. The craving for nicotine is a preference to smoke in this very moment. The conflict between these mental states is vertical, as they are not on the same order.

5.4.1.3.2 Horizontal Preference Conflict

The necessary counterexamples to vertical conflicts are conflicts in which incongruence can be traced to two mental states that (ostensibly) hold a similar hierarchical rank, i.e. that do not encompass or subsume each other. This may be the case because both of these mental states are equally interconnected, or because they are somewhat unrelated to each other. In this case, the conflict is horizontal, as its origins does not lie in the dependency between the relevant mental states.

Example 3.4 *C is an occasional smoker. Upon reflection, he finds that smoking satisfies his preference for short-term hedonism but conflicts with his preference to abstain from addictive substances for health reasons. C finds that these two preferences are of similar importance to them. A contextual recommender algorithm embedded in a mapping and navigation service suggests,*[244] *based on previous data, nearby tobacco shops C acts on their hedonism-fuelled preference and consumes cigarettes.*

The above describes a vertical conflict of preferences. While the consummation of nicotine is endorsed by one higher-order preference it is discouraged by another higher-order preference of similar importance. When acting upon one, the action is conflicted with the other. Actions

[244] So called Location-Based Services (LBS) will also be discussed in more detail later in this thesis as well as in of the Annexes.

taken in conflict are still considered to be of diminished intrinsic autonomy under this account, even when conforming with another higher-level mental state.[245]

5.4.1.4 Conflict Depth

Conflicts can also be analysed based on their status within the hypothetical mental hierarchy and their relation to other mental states, resulting in at least two variants:

- singular conflict, in which specific mental states are in conflict with one another or
- systemic conflicts, in which a considerable set of mental states are in conflict.

5.4.1.4.1 Singular Conflict

The conceptual starting point to understand conflict depth is the (arguably purely hypothetical) singular conflict, in which the minimum number of mental states are in conflict with each other.[246]

Example 1.4 *A would like to buy a tent immediately because they want to increase their time spent outdoors camping as soon as possible. At the same time, A wants to live frugally and stick to a budget they made earlier, which does not allow this expenditure. Apart from the preference to spend time outdoors and to live frugally, A holds no preferences that are touched on by the potential purchase (or non-purchase) of the tent in question.*

In this (contrived) example, the conflict as observed is incurred as a result of two conflicting mental states, the minimum number of conflicting preferences. Singular conflicts denote the lowest level of intensity of a given conflict imposed by an autonomy-constraint.

5.4.1.4.2 Systemic Conflict

In most if not all cases, the complex interdependencies of human mental states will result in conflicts that affect multiple mental states of an individual.

Example 1.5 *A wants to live frugally and stick to a predetermined budget they made earlier. A also does not want to buy goods shipped in from overseas, goods that contain excess amount of plastics and chemicals and goods that were manufactured under ethically problematic working conditions such as by using child or slave labour. After being exposed to highly effective advertising, he impulse-buys a tent that was made overseas in questionable working conditions and is predominantly plastic despite his deeply held beliefs.*

As more mental states are affected by a conflict, the intensity of the autonomy constraint can be considered to be increased. The more held preferences a prompted action violates, the more impactful for the affected individual the deviation from existing self-imposed preferences is.

[245] This is also sensible when considering the situation from a non-action-centric perspective. If an individual has a divided mind, i.e. has competing and contradicting preferences *of same importance*, his capacity of self-government seems to be questionable. In many cases, this may be due to opacity about the true value of the preferences in conflict; upon further reflection a hierarchical relationship may be discovered.

[246] The reason why this is hypothetical even at this level of abstraction is that the complex interdependencies of an individual's preferences will hardly allow for conflicts which only affects a pair of preferences.

5.4.1.5 Relevance

The above subdivision of intrinsic autonomy constraints is not merely academic. The differentiation between origins of autonomy-constraining mental state insertions is salient to determine ethical responsibility. Clearly, self-insertion of a conflicted mental state puts less if not none responsibility on other actors even it an autonomy constraint is incurred, but the same is not true for autonomy constraints imposed by third parties. Similarly, vertical and horizontal mental state conflicts arrive with different implications; vertical conflicts will likely hint at existing conflict that is triggered by the insertion of a lower-order mental state while horizontal conflicts do not require existing conflict. Both verticality and conflict depth are useful to describe the intensity of an intrinsic autonomy constraint and ethical harm it imposes. Effecting action in systemic conflict is clearly a more intense breach than the breach of a few or only one higher-order preference. Effecting behaviour that is in conflict with preferences that are relatively higher than others is a more intense autonomy constraint as relatively higher-order mental states represent convictions of increasing sincereness and strength. A higher intensity of a given autonomy constraint warrants closer scrutiny when it comes to its justification. Later sections will delve into the implications of this more deeply.

5.4.1.6 Methodical Justification

The intrinsic dimension of autonomy represents many of the main ideas of hierarchical autonomy theory. It is therefore plausible from a philosophical standpoint. The consequential imperative to protect individuals in their exercise of their intrinsic autonomy, that is that the intrinsic integrity of agents is worth protecting, seems to be covered from a legal standpoint as well. With respect to Article 8 of the ECHR, the ECtHR has found that the terminus "private life" is wide-ranging, not accessible to exhaustive definition, and therefore generally also covers the psychological integrity of a person[247], and that protection may extend to a person's inner life as well, e.g. philosophical, religious or moral beliefs, emotional life. [248] Additionally, personal development falls within the protected scope.[249] In cases concerning the rights under Article 9 the court has found that an individual's conviction (or thought or view or belief) is protected once it attains a certain level of cogency, seriousness, cohesion and importance.[250] Among such convictions are not merely religious views but also philosophical or deeply personal ones such as pacifism,[251] veganism,[252] medical philosophy with respect to alternative medicine, [253] or the actively reflected absence of religion, such as secularism[254], and atheism,[255] as well as convictions with respect to conscientious objections.[256] Notably, the court has named the rights reflected in Article 9, i.e. freedom of thought, conscience and religion, a "precious asset".[257] These findings are valuable for the purpose of this chapter; the intrinsic integrity of agents that the pragmatic model (based on hierarchical theory) defines as a relevant dimension of autonomy, seems to fall generally within the protective scope of human rights law. As warned in Section 5.3, legal concepts cannot be directly translated into the theoretical domain

[247] Cf. ECHR (2019), Nicolae Virgiliu Tănase v. Romania [GC], § 128
[248] Cf. ECHR Bensaid v. the United Kingdom, § 47.
[249] Cf. ECHR Bensaid v. the United Kingdom, § 47.
[250] Cf. ECHR (2013) Eweida and Others v. The United Kingdom §81 with further references
[251] Cf. ECHR (1978) Arrowsmith v. the United Kingdom (Commission Report), §69
[252] Cf. ECHR (1993) W. v. the United Kingdom (Commission decision)
[253] Cf. ECHR (1998) Nyyssönen v. Finland (Commission decision)
[254] Cf. ECHR (2011) Lautsi and Other v. Italy [GC] § 58
[255] Cf. ECHR (1986) Angeleni v. Sweden (Commission decision); (1994) Union des Athées v. France (Commission report) § 79
[256] Cf. just ECHR (2020), Factsheet – Conscientious objection, available at https://www.echr.coe.int/Documents/FS_Conscientious_objection_ENG.pdf
[257] Cf. ECHR (1993) Kokkinaki v. Greece § 31

of the pragmatic account, nevertheless there is sufficient congruence with respect to the underlying values and principles. To illustrate, the protection of personal development under the right to privacy covers the phenomenon of a systemic change of mental states over a period of time, a relevant phenomenon under the pragmatic account. Similarly, an interference with a deeply held belief pertaining to one's medical philosophy can trigger legal protection; such interference may also be classified as an intrinsic constraint causing conflict under the pragmatic model. More generally, one can observe that this case-law tracks well with the concept of sound mental state integration, as brought forward by Sneddon.[258] The intrinsic dimension of autonomy is therefore arguably sufficiently reflected in the legal domain to enable the above-mentioned plausibility-check.

5.4.2 Relational Autonomy

Autonomy on its relational dimension is concerned with an individual's relation to other entities, the environment and actual phenomena. An individual's relational autonomy is therefore constrained by obstacles related to these domains. Explicitly, constraints can be at least of these three types:

- factual constraints,
- normative constraints, or
- societal constraints.

5.4.2.1 Factual Constraints

Factual constraints are obstacles to an individual's autonomy at large (i.e. the potential for effectuation of one's mental states) that are introduced by some sort of manifested circumstances that arise out of the embedding of the constrained individual within their environment.[259]

Example 2.3 *After careful consideration, B decides to purchase a car. When applying for financing, a credit-scoring algorithm finds B to be not credit-worthy. As a result, B cannot purchase the car.*

Example 2.4 *B has purchased the car but falls behind on his loan-payments. B makes the decision to go on a road trip, but an algorithm recognizes the loan default and their attempt to cross state-lines, identifies this as potential fraud and disables the car remotely. B cannot drive the car anymore.*

In both examples, the individual in question is faced with circumstances that serve as obstacles to the exercise of an autonomous choice that were imposed by some sort of technology. They

[258] Of course, this does not entail, that case-law generally endorses this specific theory or hierarchical theory in general.

[259] As a result of this definition, relational factual constraint serve as a sort of catch-all for obstacles that do not cleanly fit into the other obstacle categories. On the other hand, they overlap with other autonomy constraints. For example, an information-scarce environment is a factual autonomy constraint but the lack of information is also an informational autonomy constraint. In general we may think of factual autonomy constraints as something that manifests itself somewhat physically (or in a digital sphere-equivalent) and perhaps may circumscribe it as "environmental" autonomy constraints in the common use of the word. The reason this term is not used here primarily is because the term "environmental" is already used in the context of intrinsic (or mental) autonomy constraints.

are *matter-of-fact* as they break the chain of causality between the individual's decision to act, and the effectuation of the result of that decision.

In the context of information technology specifically, factual constraints also describe the factors imposed by a digital environment, which is particularly important if an individual is required or strongly encouraged to exercise their autonomy towards their preferences within a digital environment. This autonomy-compression phenomenon and some of its implications are outlined in more detail in Section 11.3.1.2.

5.4.2.2 Normative Constraints

Normative constraints are obstacles that are introduced by normative, that is somewhat coercive and authoritative frameworks, the most important of which are legal frameworks.

Example 2.5 *B has purchased a car through a financing loan. The car is equipped with geospatial capabilities and transfers location data to the manufacturer, who provides this information to the financier. The laws in their country don't allow car owners with outstanding debt to leave the country without permission of the financier under punishment of a monetary fine. B cannot leave the country without incurring a fine.*

The implied coercive element of a normative constraint can often lead to an entangled factual constraint (e.g. fines, or imprisonment as a result of non-compliance with a legal framework).

Example 2.5 (cont. var. a) *B attempts to leave the country. When noticing the breach of the geofenced area, an algorithm within the car's software records all relevant data and alerts the authorities.*

Most relevant for this thesis, normative constraints can be operationalized by including the prescriptive limits laid out by the normative constraint into the technological framework in question.

Example 2.5 (cont. var. b) *B attempts to leave the country. When noticing the breach of the geofenced area, an algorithm within the car's software records all relevant data, alerts the authorities and prevents the car from driving any further.*

In this case, the operationalizing of the normative constraint in question has resulted in a factual constraint similar to Example 2.4. The legal constraint prohibiting the operation of the vehicle has been translated into a physical impossibility of operating the vehicle for the individual in question.

5.4.2.3 Societal Constraints

Societal Constraints are obstacles that are introduced by a suggestive, that is principally non-coercive, framework, such as other's established moral frameworks, religious frameworks, rites or traditions.

Example 3.5 *C is a smoker. They live together with other family members. C's family adheres to religious code that prohibits smoking, so C only smokes in secret. Suspecting foul play, a family member of C has installed smoke detectors.*

When alone, C wants to smoke but abstains because of his concern about repercussions of being found out.

The above example illustrates how religious norms can create a relational-societal autonomy constraint. In this case, technology is used to enable the relational-societal autonomy constraint, and the autonomy constraint is incurred in a situation in which the individual inhibits themselves in expectation of repercussions.[260] Pre-emptive inhibition is a typical response to looming relational-normative or relational-societal autonomy constraints; indeed certain strategies of technology-assisted behavioural steering rely explicitly on knowledge and transparency of repercussions.[261] But these obstacles can be imposed more directly as well.

Example 4.1 *D is a fitness enthusiast and values living a healthy fitness-oriented lifestyle. D also has a strong preference for eating bananas and a strong dislike of eating coconuts due to ethical concerns about their sourcing. D is constantly exposed to algorithmically prioritized information on social media highlighting the alleged need of conforming to a certain body type and respective metrics such as weight. The algorithm exposes her to a new trend of "clean eating", discouraging and vilifying the consumption of most berries including bananas but strongly encouraging coconut consumption. Worried about fitness goals and wanting to attain a similar body as the social media testimonials, D conforms to these new dietary rules despite her preferences.*

While the above clearly presents some problems (the ethicality of subjecting people to questionable health advice by using body image as a lever is well discussed by now), imposing constraints on relational-societal autonomy is not necessarily unjustified. (Naturally, the same holds true for relational-normative constraints and autonomy constraints in general).

Example 4.2 *D is outgoing and highly social, having a strong preference to meet friends in person regularly. During a global pandemic outbreak of a respiratory disease, D is constantly exposed to algorithmically prioritized information on social media highlighting the scorn of peers who denounce risky behaviour such as leaving home unnecessarily. As a result, D cuts back on personal meetings.*

In the last three examples, technology is not truly the origin of the substance of the autonomy constraint but serves more as a transfer-and-amplification infrastructure. This is intuitively reasonable; in most cases societal (just as normative) constraints origin from a source native to the domain of society. But this again is not necessary, in particular if the relevant information the individual is exposed to is not genuine.

Example 4.2 (cont. var. a) *Unbeknownst to D, the social media algorithm displaying, and ranking content is using the profiles, pictures and hence social clout of D's friends to transmit the negative sentiment, despite these friends not having denounced risky behaviour themselves.*

[260] These set of circumstances is further discussed in Section 13.

[261] See e.g. Section 6.1. Within the realm of persuasive technology, these strategies are grouped under the domain of "surveillance, see e.g. Brian Jeffrey Fogg, 'Persuasive Technology', *Ubiquity*, 2002.December (2002), 2 <https://doi.org/10.1145/764008.763957>.

If the need to conform to societal expectations is subverted by introducing a set of "false" expectations, the origin of the constraint's substance lies with the technological process in question. As information intermediation between individuals and the digital domain is conducted through a finite interface,[262] synthetic information may be added by virtue of the intermediating technology.

Example 4.2 (cont. var. b) *The social media algorithm presents D with aggregated information about the contact limitations D's friends are undertaking. The graphical representation of this factual aggregated information leads D to overestimates the importance of maintaining contact restrictions to D's friends. As a result, D holds themselves to the unnecessarily strict standard believed to be in line with the friends' expectations, foregoing more personal meetings than necessary.*

The potential of intermediation technology to result in synthetic information, such as erroneously inferred or approximated information, means that autonomy constraints through such interactions will typically be attributable to some extent to the technology in question.

5.4.2.4 Repellency / Permissiveness

A sensible distinction when comparing different relational constraints is their repellency, i.e. their allowance for the affected individual to overcome the constraint. If the constrained individual cannot overcome the constraint, the constraint displays absolute repellency and is fully impermissive.

Example 2.3 (rep.) *B decides to purchase a car. When applying for financing, a credit-scoring algorithm finds B to be not credit-worthy. As a result, B cannot purchase the car.*

Example 5.1 *E is a regular social media user. Due to prolonged exposure, E has formed a type of social-media addiction. During a holiday trip, E pledges to not access social media for the duration of a few days. Due to the way the addiction-like condition is affecting E, he is unable to adhere to this pledge.*

In most cases, constraints will not be absolutely repellent. In other cases, repellency is relative to the scope of considered action potential. For example, while Example 2.3 is absolutely repellent when considering the potential for a car purchase through the financier in question as it is described, the existence of a potential second financial institution willing to extend a loan would mean that the target action can still be reached.

Example 2.5 (cont. var. c) *B attempts to leave the country. When noticing the breach of the geofenced area, an algorithm within the car's software records all relevant information and displays a warning that B may be in violation of their contract. The software displays a warning that the car will not be able to be launched again, but for the duration of the current drive, nor further actions are taken.*

[262] This means that information will be typically limited, ranked or otherwise presented in a way that has the potential to play into cognitive biases of human information processing capabilities (e.g. through graphic representation), see for a more exhaustive explanation of this Section 10.3.

In this variation, the relevant technology still imposes a factual constraint, but with respect to the action of leaving the country, it is only a weak obstacle as it can be overcome by not turning the car engine off. Most normative and societal constraints are permissive; their enforcement represents (new) factual constraints that may be less permissive.

Choice architecture (or *nudges*) are arguably the most famous examples of permissive potential autonomy constraints. While the environmental factors or processes can be (often easily) overcome, individuals may apparently[263] be steered towards certain outcomes with statistic significance. A very present case of such nudges in information technology are dark patterns and default settings.[264]

5.4.2.5 Internality

Another way of categorizing constraints is to divide between

- physical constraints, and
- mental constraints.

This distinction is well entangled with the other axes of consideration already suggested. In particular, most factual constraints will have at least some physical element to them, while the pressure exerted by societal and normative constraints is primarily mental. There is some overlap; for example the case of social media addiction derived from the (perhaps inherently) addictive design of the technology in question in Example 5.1 is a factual mental constraint.

5.4.2.6 Parallel Properties of Distinctive Characteristics

The distinction by kind, by permissiveness and by internality are just three schemes of classification that can be applied to a given situation at the same time. For example, a pronounced case of addiction to a social media platform that renders an individual unable to attend to their other interests or needs is a strong constraint and it is also a mental constraint.[265] An individual may be "shadow-banned", i.e. unable to interact with other individuals without their knowledge, on an internet forum due to automated flagging by a moderating informational agent. In this case he faces a strong factual constraint.[266] An individual trying to purchase an airplane ticket in the cheapest way possible who has to click through myriads of misleadingly construed websites of a discount airliner that are intended to "trick" them into incurring additional costs faces a weak constraint that is both mental and factual.[267] An algorithm influencing a judge to deny bail of a defendant imposes constraints that are strong, factual (the restriction to the defendant's movement as a result of this decision are physically manifested, e.g. jail), and normative, in particular of legal nature (the judge's decision comes attached with

[263] See the emerging research that argues against such an effect as evidenced in Maximilian Maier and others, 'No Evidence for Nudging after Adjusting for Publication Bias', *Proceedings of the National Academy of Sciences*, 119.31 (2022) <https://doi.org/10.1073/pnas.2200300119>.
[264] See Section 6.4.3
[265] This assessment varies based on the severity of the addiction. See for this emergent situation Yubo Hou and others, 'Social Media Addiction: Its Impact, Mediation, and Intervention', *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 13.1 (2019) <https://doi.org/10.5817/CP2019-1-4>.
[266] Cf. Callie Middlebrook, 'The Grey Area: Instagram, Shadowbanning, and the Erasure of Marginalized Communities', *SSRN Electronic Journal*, 2020 <https://doi.org/10.2139/ssrn.3539721>.
[267] Cf. for this topic Adrian Palmer and Stephanie Boissy, 'The Effects of Airline Price Presentations on Buyers' Choice', *Journal of Vacation Marketing*, 15.1 (2009), 39–52 <https://doi.org/10.1177/1356766708098170>.

legal consequences that imply coercive measures).[268] Peer pressure on an individual to cut down social contacts during a pandemic incurred by a social media algorithm that highlights health warnings and statements denouncing risky behaviour imposes a weak mental and decidedly societal constraint.

5.4.2.7 Source of Constraint

Just as above outlined in Section 5.4.1.2, constraints may be imposed by other entities, environmental factor or the individual themselves. This distinction is not necessary to categorize autonomy here but becomes relevant when considering the ethical implications as is done in Section 11.

5.4.2.8 Relevance

The three main axes of observation are all useful to understand the invasiveness and ethical challenges that come with technology imposing relational constraints. Permissiveness of a relational constraint scales well with the need for ethical justification: clearly fully coercive measures require a stronger justification than effective persuasion. The distinction between mental and physical constraints is interesting as well. Physical constraints are intuitively more problematic than mental constraints up to a point, after which mental constraints are often considered (much) more problematic.[269] The recognition of mental constraints also ties relational and intrinsic autonomy together: many intrinsic autonomy constraints can also be understood under the lens of relational autonomy. The tripartite division of relational constraints' usefulness is immediately apparent. Recognizing normative and societal aspects of autonomy constraints acknowledges the importance of the socio-cultural embedding of individuals,[270] and again highlights that autonomy constraints must not be completely impermissive. Factual constraints represent the most direct form of constraining an individual's autonomy. Because of the increasing prevalence of and reliance on information technology, and the fact that individuals interact in increasingly digital spaces, technology has increasing potential to impose factual constrain, as alternatives to certain actions become sparse.[271]

5.4.2.9 Methodological Justification

The relational dimension of autonomy incorporates ideas of feminist or relational autonomy as outlined above. It is therefore plausible from a philosophical standpoint. From a jurisprudential perspective, much speaks for the ideas of this approach as well; many aspects of human behaviour and relevant constraints to the same that are relevant under the relational dimension are relevant from a legal standpoint also. There are two different approaches to this autonomy dimension: one focuses on the general importance of an individual's relationship to other individual's and their environment, the other focuses on constraints on an individual. With regards to the first approach, the European Court of Human Rights has repeatedly focused on the importance of individual's relation to each other. For example, privacy under Article 8 of

---

[268] Cf. for this highly debated practice Julian Adler, Sarah Picard, and Caitlin Flood, 'Arguing the Algorithm: Pretrial Risk Assessment and the Zealous Defender Jed D. Melnick Symposium: Innovations in Justice: Experiments in Restorative Justice', *Cardozo Journal of Conflict Resolution*, 21.3, 581–96 <https://heinonline.org/HOL/P?h=hein.journals/cardcore21&i=601>.

[269] Consider the ethical debates surrounding the ethicality of subliminal messaging or the explicit exclusion of coercive mental persuasion from the domain of persuasive technology, see e.g. Section 6.

[270] This is further explored in Section […]

[271] For example, if all financiers require a credit score calculated by a specific algorithm to take out a loan, the financing transaction has *enveloped* the credit scoring algorithm, and alternative means of financing may become more unlikely. Compare for the concept of enveloping also L Floridi, 'Enveloping the World: The Constraining Success of Smart Technologies', *CEPE 2011: Crossing Boundaries*, 2011, 111.

the European Convention on Human Rights is found to include the right for each individual to approach others in order to establish and develop relationships with them and the outside world, so including potentially also professional and commercial activities.[272] Indeed, even interactions between people even when conducted in a public context may be covered.[273] Establishment of social, professional or other relationships with other individuals, which pertains to the first approach of the relational autonomy dimension, are hence recognized and generally encompassed by the protective scope, even if in specific situations such protection may not be granted due to other conflicting values.

With respect to the constraints, the situation is less clear cut. To this end this thesis presents a weaker and a stronger argument. On the weak side, it is argued here that one can derive the concern for constraining an individual's autonomy by informational agents by observing the general importance of safeguarding data to avoid such constraints. Very explicitly, the ECHR found that "the protection of personal data is of fundamental importance to a person's enjoyment of her right to respect for private and family life as guaranteed by Article 8 of the Convention".[274] Consequently, domestic law is to ensure that personal data must be efficiently protected from misuse and abuse,[275] and therefore instate appropriate safeguards to prevent use of personal data inconsistent with Article 8.[276] This has heightened priority when automatic processing is in place.[277] However, data protection as a protective measure to benefit autonomy is, at best, an indirect measure. It is plausible that data protection, while supporting what the pragmatic account considers to be an individual's autonomy, is intended to protect a different set of values first and foremost. A stronger argument comes from the jurisprudence to Article 9 of the European Convention of Human Rights in which the right to manifest one's (also non-religious) belief is repeatedly affirmed. The extension of protection of Article 9 § 1, with reference to cogency, seriousness, cohesion and importance, is limited. For example, protection is not granted with respect to just any thought. Instead the court applies a filter of intensity, in which remedies can only be had if it imposes a sufficiently intense constraint; consequently; a lack of "inconvenience" ("*un désagrément suffisant"*) as a result of the respective constraint might preclude protection.[278] This aligns with the gradient of diminished autonomy that was proposed by dividing constraints into more or less severe obstacles to autonomy. One can therefore argue that the relational dimension of autonomy is sufficiently reflected in jurisprudence for the purposes of this thesis.

5.4.3 Informational or Transparency-Related Autonomy

Autonomy on its transparency dimension is concerned with information and its disclosure. Obstacles on the transparency dimension can be considered in three ways, which groups them into:

- autonomy-specific informational constraints,
- action-specific informational constraints, and
- context-specific informational constraints.

[272] Cf. ECHR Botta v. Italy, §32, Fernandez Martinez v. Spain [GC §110, Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland [GC], § 130)]
[273] Cf. ECHR (2012) Von Hannover v. Germany (No. 2), [GC], §95
[274] Cf. ECHR Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland [GC], § 133
[275] Cf. ECHR (2009) Gardel v. France, § 62
[276] Cf. ECHR (1997) Z v. Finland, § 95
[277] Cf. ECHR, (2009) Gardel v. France, § 62
[278] Cf. ECHR (1999) Viel v. France (Decision)

All of these constraints deal with insufficient information accessible for the constrained individual. The first is imposed by insufficient information about the fact that autonomy is constrained *per se*, the second by insufficient information about potential actions the individual could exercise their autonomy by.

5.4.3.1 Autonomy-Specific Informational Constraints

Autonomy-specific informational constraints are a lack of information about inhibitions to other dimensions of one's autonomy. In other words, autonomy-specific informational constraints impose a lack of reactive potential on an individual as they are unaware of their current situation. Consider the following examples.

Example 2.6 *After careful consideration, B decides to purchase a car. When applying for financing, a credit-scoring algorithm finds B to be not credit-worthy due to his political beliefs evidenced by online interaction. As a result, B cannot purchase the car.*

Example 2.7 *An automatic lead-generating algorithm is analysing B for their eligibility for car financing. B has no plans to buy a car and has not inquired about a financing loan with anyone. The algorithm creates a credit-score and finds B to be not credit-worthy due to his political beliefs evidenced by online interaction. As a result, B could purchase the car.*

In both examples, the individual faces a relational-factual constraint about their capacity to purchase a car by utilizing financing loans. However, in Example 2.7, the individual has no knowledge about this whatsoever. Should they decide to purchase a car at a later stage, the credit-score that was assigned to them surfaces as an autonomy-constraint that has already persisted. Such situations describe autonomy-specific informational constraints.

5.4.3.2 Action-Specific Informational Constraints

Action-specific informational constraints concern information on the basis of which autonomous actions would be possible. If an individual is not aware of actions, they are able to take, or their awareness is made difficult, then they are constrained on their this aspect of their autonomy.

Example 4.3 *D is outgoing and highly social. One of D's wearables tracks his location and informs D's friends when he is out and travelling through his city. Because of an ongoing pandemic, D prefers some of his friends not knowing how often he goes out. Unbeknownst to D, there is an option of hiding this information from a select circle of friends, hidden behind confusing user interfaces.*

Example 3.6 *C is a smoker trying to quit. After a promising treatment option becomes available at his location, a consortium of tobacco-companies buys up digital ad-space discrediting the new treatment. This leads to mistrust into the treatment and diminished visibility.*

Example 3.6 (cont. var. a) *When C hears about the new treatment, they don't consider it as a valid option due to the consortium's advertisement that has left a big impact on them.*

Example 3.6 (cont. var. b) *C never hears about the new treatment as the consortium's advertisement blots out any information about it.*

Not having a sound overview over potential actions at one's disposal is an action-specific informational constraint. This informational obstacle can be present due to intentional design, non-intentional environmental factors or due to the particularities of the individual. (Clearly, no individual has constant overview over their complete action-potential; to this end (like in other aspects) autonomy is always *somehow* constrained.)

5.4.3.3 Context-Specific Informational Constraints

Context-specific informational constraints which describe the informational obstacles to an individual's autonomy that are neither autonomy-specific nor action-specific as outlined above. These constraints occur when contextual information about an autonomy-specific constraint is not sufficiently accessible.

Example 2.6 (cont. var. a.) *Upon B's question about why the loan was denied, the financier refuses to elaborate, denoting it as proprietary information.*

Example 2.6 (cont. var. b.) *Upon B's question about why the loan was denied, the financier (wrongly) references insufficient income levels.*

In Example 2.6, the individual in question is the subject of a relational-factual constraint on his capacity to receive financing and purchasing a car. In this case, the individual knows that there is some autonomy constraint levelled on them. But the peculiarities of the constraint, i.e. the fact that they were denied the loan due to political beliefs is not information that they are privy to. This lack of information by itself qualifies as context-specific.

5.4.3.4 Information Veracity

When considering informational constraints, the obstacles imposed on the individual parsing the information can have different qualities related to its informational veracity. Broadly when we consider informational veracity, we can distinguish between an informational autonomy constraint utilizing *false* information, or an informational autonomy constraint using *misleading* information.

Example 1.6 *After some reflection, A decides to buy a tent that ought to have certain features, such vestibules on both sides, the option for a cold sink and durable outer fabric. A customer-analysis algorithm has profiled A on the basis of A's internet search history to the extent that these preferences are now reflected in their profile and creates an advertising playing to A's preferences.*

Example 1.6 (cont. var. a) *On the basis of A's profile, a seller advertises a tent to A by emphasizing the desired features. In reality (and unbeknownst to A), the advertised tent does not have the features, making the advertisement untrue. As a result of this advertisement A buys the tent.*

Example 1.6 (cont. var. b) *On the basis of A's profile, a seller advertises a tent to A by suggesting (but not stating) the desired features. In reality (and unbeknownst to A), the*

advertised tent does not have the features, making the advertisement untrue. As a result of this advertisement A buys the tent.

In the first example variant, the individual is exposed to and acts on the basis of untrue information, while in the second example the information is wrongly inferred by the individual as a result of misleading representation. Clearly both impede the ability of the individual to realize their preference to buy a certain product, so both ought to be recognized as autonomy constraints.

5.4.3.5 Information Prevalence

Informational autonomy constraints can be imposed even if when not utilizing misinformation or misleading information. Along the axis of information prevalence, both the lack of information and an exceeding amount (i.e. information *overload*)[279] can lead to autonomy constraints as well.

Example 4.4 *D prefers some of his friends not knowing how often he goes out, but one of D's wearables tracks his location and informs D's friends when he is out and travelling.*

Example 4.4 (cont. var. a) *While there is an option to disable this feature, the documentation of the device and its software is barebone and does not mention the feature. As a result, D does not know about this option and does not disable the feature.*

While the individual in this case was not exposed to false or misleading information in a pure sense, the lack of information about the individual's action potential imposes a constraint in itself. One cannot take actions if one does not know the action can be taken in most cases.

Example 4.4 (cont. var. b) *While there is an option to disable this feature, the documentation is unordered, of excessive length and cannot be parsed via electronic search. As a result, D does not find any information about this option during his limited search and does not disable the feature.*

In this variant, information is accessible theoretically, but it is drowned out in other (irrelevant) information. If the search cost is too high, the information becomes inaccessible for most purposes; if the search cost is made high intentionally, then this intent extends onto the respective informational autonomy constraint.

5.4.3.6 Source of Constraint

Just as above outlined in Section 5.4.1.2 and 5.4.2.7, constraints may be imposed by other entities, environmental factor or the individual themselves.[280] Again, this distinction is not necessary to categorize autonomy here, but becomes relevant when considering the ethical implications as is done in Section 11.

---

[279] See for this phenomenon e.g. David Bawden and Lyn Robinson, 'Information Overload: An Introduction', in *Oxford Research Encyclopedia of Politics* (Oxford University Press, 2020) <https://doi.org/10.1093/acrefore/9780190228637.013.1360>.

[280] An interesting case of self-constraining actions with respect to informational autonomy is the phenomenon of self-deception, see e.g. Ann E Tenbrunsel and David M Messick, 'Ethical Fading: The Role of Self-Deception in Unethical Behaviour', *Social Justice Research*, 17.2 (2004), 223–36.

5.4.3.7 Relevance

Autonomy-specific informational constraints represent the problems arising out of individuals not having access to or understanding of their data that is being accumulated and operationalized. Within the information that is attributable to them in some scenarios, existing and future profiles can be used to impose autonomy constraints on them without knowledge. Action-specific informational autonomy constraints capture a wide range of technological phenomena, one of the most important ones being adversarial user design (i.e. Dark Patterns).[281]

5.4.3.8 Methodological Justification

Most philosophical theories of autonomy account for information deficits. There is usually a general requirement of knowledge and information processing capabilities by virtue of an individual's authenticity requirements to have the capacity of autonomy.[282] However, knowledge about the option at one's disposal is characterized explicitly by theorists as well, so for example by the foundational and widely influential account of Gerald Dworkin.[283] That transparency benefitting the individual as a safeguard towards their rights is valued from a legal standpoint hardly requires explanation. . Within the scope of human rights regimes, we can find reflections of that e.g. as a tacit requirement of Article 6 of the European Convention of Human Rights, protecting the right to a fair trial. This is considered to include the right to silence (i.e. to not self-incriminate).[284] To ensure this protection is adequate, individuals must be informed of this right (and with that of their potential actions at their disposal), a notion that has garnered widespread acceptance.[285] Generally however, the transparency dimension is made plausible better by recourse to other legal instruments beyond the human rights domain.[286] For example, in consumer law, information with respect to a consumer's right to withdrawal is often an obligation imposed on the counterparty. This principle can be found across the European Union as a result of legal harmonization, e.g. as a consequence of Article 6 of the E-Commerce Directive. Similarly, Article 13 of the GDPR[287] imposes obligations on data controllers to ensure that data subjects are informed about their rights to access, rectification or erasure (among others).[288] All of above examples highlight, that in order for an individual to adequately choose between options at his disposal, i.e. behaviour that one may see as an exercise of his decisional autonomy, legislators have repeatedly seen the need to ensure that the individual has adequate information with respect to these potential options. Put abstractly, there seems to be an understanding of the (trivial) fact, that at least in certain

[281] See Sub-Section 6.4.3

[282] E.g. Sneddon, *Autonomy*, p. 27f.

[283] Gerald Dworkin, 'Autonomy and Behaviour Control', *The Hastings Center Report*, 6.1 (1976), 23 (p. 14) <https://doi.org/10.2307/3560358>.

[284] Cf. ECHR (1996) Murray v. UK § 45

[285] See e.g. the Miranda Rights in the U.S. and the Reding Rights in the EU.

[286] Indeed, human rights instruments typically do not have an explicit "human right to transparency", as is relevant for the autonomy dimension of transparency as discussed here. Instead, transparency in the context of human rights is most often used with respect to transparency in the sense of a "right to know" with respect to information flows more generally or as a control mechanism vis-à-vis public institutions; see for this e.g. the wording of Article 19 of the Universal Declaration of Human Rights.

[287] While not fully congruent with the European Convention of Human Rights in its application on states it here stands as another widely adopted consensus that has been welded into a normative instrument.

[288] Indeed, the GDPR explicitly mandates compliance with a transparency principle in Article 5. Similarly, Article 7 of the GDPR in connection with Recital 32 lays out that consent must be given via an (inter alia) "informed" indication of an agent's agreement.

contexts, the choice between certain actions is hindered by a lack of information. The transparency dimension of this account aims to reflect the very same notion.

5.5 Autonomy and Autonomy Constraints

The pragmatic account and its dimension outlined above provide for understanding of the exercise of autonomy of an individual and the concept of autonomy in general. However, its relevance and salience has been highlighted mostly using the contraposition of obstacles or constraints to autonomy. This is intentional. While the framework is valid without context and in the abstract, it is the need to differentiate between different autonomy constraints that warrants use of the pragmatic account of autonomy in the first place.

Therefore, a different way of understanding autonomy dimensions is of seeing them as classification of different constraints.[289] By definition, each obstacle for or interference with someone's autonomy constrains the respective individual in a certain way; and at least on one specific autonomy dimension. Informational agents (as may other individuals or just general environmental factors) can induce such constraints and therefore limit autonomy. We can therefore assign certain actions or operations with an impact on someone's autonomy.

Between an entity's constraining action and the incurred autonomy constraint of an individual, there lie additional factors of a causal chain. These intermediary factors describe, among other things the type, technical characteristics, and venue of the constraint. Within this inquiry, such factors are called vectors of influence and will be explored further in a later section.

5.6 Interim Analysis

Comparing the rights guaranteed by international regimes with respect to privacy, freedom of thought and freedom of opinion with the ideas of theoretical-philosophical frameworks it becomes evident that the approach is plausible and compatible with contemporary efforts to conceptualize autonomy. More importantly the view is endorsed here that the pragmatic multi-dimensional approach to autonomy, in particular the intrinsic, relational and transparency dimensions, displays a clear utility towards some of the purer autonomy theories with respect to addressing emerging issues connected to the IoE.

5.7 Positive and Negative Freedom or Liberty under the Pragmatic Account of Autonomy

As a by-product of its structure, the pragmatic account also registers constraints on both positive and negative autonomy. The peculiarities of this are worth exploring. To recall Section 2.1.3, both the freedom to choose and the freedom from outside influence are often considered

[289] Similar observations have been made about the strategies to approach the concept of privacy, see Solove, 'A Taxonomy of Privacy'. Compare his justification: "*In devising a taxonomy, there are many different ways to go about carving up the landscape. I focus on the activities that invade privacy. The purpose of the taxonomy is to assist the legal system in grappling with the concept of privacy. Since the goal of the law is to have privacy protections that best prevent and redress particular problems, we need to first understand the problems in order to evaluate the effectiveness of the protections. Therefore, my focus is on activities that create problems. I aim to show that these activities differ significantly yet share many commonalities.*" The same logic applies to the present endeavour. See also for a different take but the same logic Kasper.

relevant.[290] Assignment of autonomy constraints to the sphere of positive or negative freedom (or autonomy or liberty) is not always clear cut. On first glance, the pragmatic account, due to its focus on autonomy constraints seems to be most closely aligned with matters of negative freedom, i.e. the absence of outside influences. We may recall that Berlin considered negative liberty as "the degree to which no man or body of men interferes with my activity".[291] Beyond the most obvious extrapolation from "man or body of men" to artificial entities such as agents of the IoE, the question then relies on what ought to be considered interference.[292] Under the pragmatic account, interference may come in many forms. For example, on the intrinsic dimension, it is sensible to count mental state insertion like highly effective advertising as interference. Similarly, the relational dimension encompasses some of the more original constraints that may be considered coercive such as legal constraints (and their respective enforcement), or factual constraints (e.g. the actual enforcement).

But here, connections to positive autonomy start to become clearer. Consider the relational factual constraint of an individual being denied a financing loan for a purchase by an algorithm. The individual is clearly interfered with in his activity to purchase whatever they were planning to purchase. But the purchase in question represents a potential action the individual could not have taken on their own means anyway (as they lacked the resources to do so). Perhaps it would be better then to understand these particular constraints from the viewpoint of positive autonomy: The granting of financing would have empowered the individual to exercise their autonomy in new ways, so the denying of financing still creates a disparity of options in comparison. However, the loss of (potential) autonomy was not incurred by actual coercion but by not embedding the individual with factual circumstances which would have bolstered their capacity for self-government.

Ultimately, these types of distinctions can be drawn for all three of the major dimensions of the pragmatic account.[293] Consider the following table for an overview of examples.

Table 1- Examples of Positive and Negative Autonomy Constraints

	Negative Autonomy Constraints (select.)	**Positive Autonomy Constraints (select.)**
Intrinsic Autonomy	highly effective advertising (incongruent mental state insertion)	lack of assistance by an anti-addiction algorithm with respect to mental load (lack of congruent mental state insertion)
Relational Autonomy	algorithmic assignation of scoring value vehicle-based algorithm determining "fit-to-drive"-status insufficient reflection of legal requirements for data access requests	lack of privacy with respect to online search data (relational-societal constraint)
Informational Autonomy	misinformation manipulative user design	lack of ancillary information relevant for the user lack of relevant information structure in case of information overload

[290] The distinction between positive and negative freedom is usually had with this exact terminology: "freedom", or perhaps "liberty". But some scholars have expanded this to the domain of autonomy as well, see e.g. Krzysztof Nawratek, 'Top-Down Revolutions. Negative and Positive Autonomy', in *Total Urban Mobilisation* (Singapore: Springer Singapore, 2019), pp. 37–46 <https://doi.org/10.1007/978-981-13-1093-5_5>. As this text uses freedom, liberty and autonomy somewhat interchangeably, it makes sense to extent the dichotomy of negative and positive freedom to capacity for self-government (i.e. autonomy) as well.

[291] Berlin. See for an application to this specifically on data-driven technology e.g.

[292] Berlin himself considers coercion as the necessary factor that ought to be present.

[293] Nb that for most purposes the relational aspect of one's autonomy is very much the absence of constraints, aligning it closer to negative conceptions of autonomy (or freedom), while the other aspects do have constituting factors by themselves, making them closer related to positive conceptions.

Both the negative and positive conceptions of autonomy are encompassed by the pragmatic account. The implications of this scope depend on the use and purpose of analysis. For example, using the pragmatic account of autonomy to classify all the autonomy constraints an individual may face as a result of a social scoring algorithm (e.g. lack of information about the scoring, denial of products and services, etc.) may be overly sensitive when adopting a liberal viewpoint of concern only related to issues of negative autonomy. Within this text, both constraints on negative and positive autonomy are considered relevant.

5.8 Interim Conclusion

The overarching research question guiding the first part of this thesis was the following: How can autonomy be conceptualized to be relevant and salient for investigating emerging issues of the IoE. In addition and in complementing and inversing this question, research sub-question 1.4 has asked, what the main obstacles to conceptualize autonomy are. Both of these have been addressed by this Section. This text argues that the main focus on conceptualizing autonomy and freedom ought to be to conceptualize its potential constraints. Subsequently, we ought to understand constraints as adhering to (one of) three types of imposition: intrinsic, relational and informational.

As has by now hopefully been evident, the pragmatic model allows for a useful analysis compared both to an intuitive method as deployed by academic and popular commentators, the methods suggested by either hierarchical, relational or other accounts of autonomy, or by relying on deduction from relevant legal safeguards, such as Article 8, 9 and 10 of the ECHR alone. The pragmatic mode of autonomy allows to take intuitive concerns and connect them to one or many specific autonomy dimensions, thereby allowing for greater transparency of any arguments made in favour or against the concern. In many of its dimensions the relevant "attachment points" of theory to the real world, i.e. specific (meta-)physical phenomenon that are relevant under a certain autonomy dimension (e.g. psychological integrity as the state of mental state congruence) are often within the scope of legal protection (e.g. the right to privacy extending to psychological integrity generally), which was shown via a cursory reference to selected legal sources.

In Section 1.2, I have outlined what I called the secondary goal of autonomy research, that is to undergird discussion and remedies of issues with autonomy in, among others, the legal domains. The author hopes that this added precision might be a step towards solving the issues of identifying risk to an individual's autonomy, analysing and comparing such risk to other potentially harmful situations and ultimately transfer these findings into more effective legal safeguards against intrusions into human autonomy by informational agents. To this end, the author believes that a more granular mapping of situations that arise in the interaction between humans and informational agents that is made possible by the above approach, is fully relevant for the legal domain as well. The author therefore sees this model as a contribution to the overall discourse of autonomy studies, theoretically and applied alike.

This framework comes not without limitations in the current form. The outlined approach is barebone and should be expanded and fleshed out by contextual need. Further analysis of the legal domain to find autonomy-relevant protections, in particular in the fields of legal responsibility, seems warranted. Further research, as an application of this model, is also especially needed with respect to mapping out the autonomy related issues of specific contextual domains of the human-informational agent interactions. The outlined model here serves as a context-agnostic toolset. Its application towards specific environments that come

with their own conflicts of values will require further research but seems promising. Challenges introduced by the peculiarities of location-based services – algorithms may turn out to require a yet undescribed different dimensional lens to adequately assess their risks; this may be entirely different to the risks of an algorithm in the domain of the Internet of Health or Internet of Money. Also, as is later outlined in Section 11.2.1, additional prerequisites are needed to use the pragmatic account of autonomy for ethical analysis; however this section also shows that the account is a strong tool for such purpose. Finally, the system is somewhat domain-agnostic. Naturally, these constraints can be imposed in environments that do not belong to the IoE. But the focus of this thesis does lie on the interactions of entities within the IoE and individual autonomy. To complement this generalized[294] account of autonomy, it is therefore valuable to consider the peculiarities of the IoE more closely and to see where these connect to the autonomy-relevant considerations outlined in this section. Whereas the first part of this thesis has outlined the concept of autonomy both generally and how we ought to apply it in light of the research questions, the following sections of the thesis aim to contextualize. Insofar agents of the IoE are capable of imposing autonomy-constraints (and the many examples provided in this text already suggest that they are), understanding autonomy constraints includes understanding the domain from which these constraints emanate from. Of interest must then be how the domain of information technology (and changes to it) are conducive to its agents constraining individual autonomy. What characteristics of the information domain and the direction it is shifting to increases the risk to individual autonomy? What characteristics of the agents of the IoE are relevant for the risk to individual autonomy? The next part engages more deeply with the impact of technology on the main research question of this thesis.

Part II: Influence

Individual autonomy can be exercised and constrained in any domain accessible to humans. The way this exercise can occur, and the ways in which this exercise can be constrained is dependent on the domain in which it is situated. In Part I, this text has analysed the concept of individual autonomy in general. From existing general ethical and legal theory but also input from technology-related case law, this thesis has derived the pragmatic account of autonomy. During this process, the text has already exemplified many autonomy constraints by reference to matters of information technology. However, this cursory analysis, while sufficient at the time, deserves a more comprehensive treatment.

Part II investigates the flow of information that can display autonomy-relevant characteristics, a phenomenon below introduced as the *informational pipeline*, and the relevant phenomena that affect their potential impact, which are defined within this text as *vectors of influence.* The analysis of these vectors of influence is divided in three parts, generally following the structure of said *informational pipeline*: domain characteristics, agent characteristics, and information receival characteristics. The domain of information technology (i.e. the IoE), the paradigm shifts that have occurred (as relevant for individual autonomy), and the characteristics of entities along this domain are highlighted first. Second, analysis will investigate the most important agent characteristics. To this end, the text also provides a more formal definition of the concept of informational agents (or agents of the IoE), a term that has already been used in this text. Finally, analysis will focus on the phenomena attributable to the information recipient and their environment.

[294] The word "generalized" here having the meaning of generalized in its application, and not in the way it was derived, as the examples and sources underlying the reasoning laid out in this section and previous ones where indeed closely connected to the domain of information technology.

The main argument of this section is this: If technology has the potential to impose autonomy constraints, and the technology is increasingly more powerful, more widespread in its use and its domain permeates more of the lived realities of individuals on which constrains may be imposed, the risk for autonomy constraints also rises.

6 Excursus - Persuasive Technology

Within this inquiry, the autonomy-relevance of interconnected information as present in the IoE is analysed. This does not mean that technology (even information technology) that does not fall within this scope (e.g. because it is not connected in any sense) is not autonomy-relevant. Previous scholarship has engaged with the implications and permutations of technology generally which is specifically intended to or at risk of affecting behaviour under the label of persuasive technology. This research is a valuable starting point for the inquiry at hand as it deals with many similar issues and concerns. To this end, this section aims to answer research sub-question 2.2, which asked how the existing field of research on persuasive technology connects to understanding the exercise and maintaining of autonomy constraints in presence of IoE-systems.

Within this space, persuasive technology is understood to be the set of all technology that (intentionally) imparts persuasive qualities and social influence on information with the consequence of modifying people's attitudes, behaviours or beliefs, and as a result, their actions.[295] It is considered to represents the application of from behavioural and cognitive science, the domain of human-computer interaction and design studies in the design of information systems. From this it becomes clear that the set of persuasive technology is overlapping with and of high relevance for the inquiry at hand. It is however not fully congruent for the following reasons: First, persuasive technology is usually considered to require human intent, which is too narrow for the inquiry at hand, and second persuasive technology consists of all levels of technology with certain requirements, irrespective of their attribution to the Internet per se or specifically to the IoE, which is too encompassing for the inquiry at hand. Nonetheless, many of the findings hold value for the analysis of the exercise of autonomy within the IoE, so the following section will attempt to shortly survey existing research in this space, with the aim of contextualizing the analysis of a more constrained set of technology to follow.

This analysis will proceed as follows: Section 6.1 provides the original definition of persuasive technology, that continues to strongly influence this field of research, and outlines the main strategies understood the be a part of this domain. Section 6.2 investigates closer the qualities, technology ought to have to be considered persuasive. Section 6.3 highlights the most important ethical concerns that attach to persuasive technology. Section 6.4 provides information about distinguishing persuasive technology from related concepts such as nudging or dark patterns. Finally, Section 6.5 builds a bridge from this survey to the main objective of this thesis, highlighting the parallels between research into persuasive technology and autonomy-constraining technology that is of interest in the present inquiry.

6.1 Original Meaning

The term persuasive technology was coined originally by the social scientist Brian Jeffrey Fogg, who used it to describe "any interactive computer system designed to change people's

[295] Compare for this also the definitions in Maurits Clemens Kaptein and others, 'Persuasion in Ambient Intelligence', *Journal of Ambient Intelligence and Humanized Computing*, 1.1 (2010), 43–56 <https://doi.org/10.1007/s12652-009-0005-3>. and D. O'Keefe, *Persuasion: Theory and Research*, 2nd edn (Beverly Hills: Sage Publications, 2002).

attitudes and behaviours.[296] This change in behaviour (like all behaviour) is thought of being a product of the three factors *motivation*, *ability* and *triggers*. In other words, persuasive technology must ensure that the subject to be persuaded is sufficiently motivated, has the ability to perform a certain behaviour and is triggered to this end.[297] Fogg later supplemented this concept with an underlying matrix of behaviour types, which require different types of triggers (and hence persuasion measures).[298] Fogg identified the main strategies when using technology to impart persuasion as reduction, tunnelling, tailoring, suggestion, self-monitoring, surveillance, and conditioning: Reduction strategies decreases complexity and hence friction in actions individuals take towards a certain goal, decreasing their cost (e.g. their mental load or time investment), and thus improving the cost/benefit ratio of the entire procedure in question. Tunnelling strategies aim to decrease uncertainty, and hence achieve simplification, by providing guidance meant to persuade individuals towards a specific type of interaction.[299] Tailoring strategies encompass efforts to personalize products or services, or the environment in which they are provided.[300] Suggestion strategies aim for time-sensitive interaction when the information recipient is thought to be particularly receptive for persuasion.[301] Self-monitoring strategies mean to persuade by presenting salient information about an individual directly to the same individual, decreasing the cost an individual may experience when tracking the information of interest.[302] Surveillance strategies on the other hand aim to facilitate persuasion by presenting salient information about an individual to a third party, while still ensuring that the observed individual is aware of the surveillance and acts in accordance with that knowledge.[303] Lastly, conditioning strategies utilize instrumental conditioning, i.e.

[296] See e.g. Brian Jeffrey Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do* (San Francisco: Morgan Kaufmann Publishers, 2003), p. 1. Relatedly, he also coined the term „captology" to describe the study of computers as persuasive technologies.

[297] Brian Jeffrey Fogg, 'A Behaviour Model for Persuasive Design', in *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09* (New York, New York, USA: ACM Press, 2009), p. 1 <https://doi.org/10.1145/1541948.1541999>.

[298] Note that Fogg originally conceived 35 different behaviour types but later simplified this down to 15, see Brian Jeffrey Fogg, 'The Behaviour Grid', in *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09* (New York, New York, USA: ACM Press, 2009), p. 1 <https://doi.org/10.1145/1541948.1542001>.

[299] FOGG, *supra* note 2 at 70f. An example of tunnelling are registration processes on websites or multi-step information intake procedure during set-up for a financial planning software.

[300] *Id.* at 73f. Tailoring examples include suggested items for purchase on websites based on previous information or personalized newsletters. See for further information also Arie Dijkstra, 'Personalization/Computer-Tailoring in Persuasive Technology: Tailoring Ingredients Target Psychological Processes.', in *PPT@ PERSUASIVE*, 2016, pp. 6–12.

[301] Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do*, p. 78f. Suggestion examples include (for implicit suggestion) speed radar devices that display the speed of a vehicle for the driver to see to regulate their driving, or notifications triggered by entry into a specific geographic area. See for this also Adrienne Andrew, Gaetano Borriello & James Fogarty, *Toward a Systematic Understanding of Suggestion Tactics in Persuasive Technologies* 259–270 (2007), http://link.springer.com/10.1007/978-3-540-77006-0_32.

[302] Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do*, p. 84f. Many health-related applications allow self-monitoring with the stated intent that more control and knowledge can lead to positive outcomes, see for these applications specifically John Matthews and others, 'Persuasive Technology in Mobile Applications Promoting Physical Activity: A Systematic Review', *Journal of Medical Systems*, 40.3 (2016), 72 <https://doi.org/10.1007/s10916-015-0425-x>.

[303] Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do*, p. 88f. Examples for surveillance are overt employee monitoring systems, traffic speed traps (if announced by signs or similarly known by drivers) or electronic ankle monitors for criminal offenders, see for this e.g. Robert S. Gable and Ralph Kirkland Gable†, 'Remaking the Electronic Tracking of Offenders into a "Persuasive Technology"', *Journal of Technology in Human Services*, 34.1 (2016), 13–31 <https://doi.org/10.1080/15228835.2016.1138839>. However, surveillance strategies, while ensuring compliance, often do not bring with them the more subtle change of changing the subject's viewpoints undergirding the prevented non-compliance.

reinforcement learning through positive and/or negative consequences for an individual's behaviour.[304] This first conception of persuasive technology has proven to be very influential and is widely referenced, including in most sources relied upon in this text.

6.2 Qualifications

6.2.1 Technology

Traditionally, the term persuasive technology is used primarily to denote user-facing (and as such *interactive*) devices and processes. However, this does not exhaust the scope of the above definition, but merely highlights that some sort of interacting interface is needed at some point to translate the persuasion from within the technology onto the individual to be persuaded. In cases of persuasive technology being a part of the domain of information technology, individuals cannot directly interact with the digital domain but need an intermediating device to send and receive information.[305] This creates the previously mentioned "informational pipeline" of sorts, a chain of devices linked to each other spanning from the source of information to the individual receiver.[306] This informational pipeline as a whole, as well as its individual parts may qualify as persuasive technology, i.e. as technology that (intentionally) imparts persuasive qualities to information with the consequence of modifying people's attitudes, behaviours or beliefs, and as a result, their actions. To illustrate, consider the case of a computer application to be used on a mobile phone intended to help its user to stop smoking.[307] Using the above posture, the software-front end is clearly persuasive technology, as is the device (in this context) on which it is used.[308] But the same holds true for some of the less visible algorithms and their manifestations as well. The application may rely on a complex machine learning method, e.g. for profiling reasons to deliver persuasive information more effectively, and that method may be deployed server-side (i.e. not on the device considered). If this method is sufficiently causal for the persuasion incurred,[309] it is clearly also persuasive technology, even though it does not directly interact with the subject. We may hence differentiate manifestations of persuasive technology both by their interaction-distance from the final subject to be persuaded and by their physical manifestation.[310]

---

[304] *Id.* at 93f. Many language learning applications use conditioning through in-application rewards (such as application-specific currency) obtained by regular (e.g. daily) use. Similarly, negative conditioning is also possible through the use of *punishing* instead of *rewarding* reinforcements, e.g. by appealing to negative emotions such as guilt or by threatening negative consequences if an individual is not compliant.

[305] See Section 8.2.1.

[306] Compare for this concept also the concepts of *horizontal* and *vertical axes* in Richard E. Petty, Duane T. Wegener, and Leandre R. Fabrigar, 'Attitudes and Attitude Change', *Annual Review of Psychology*, 48.1 (1997), 609–47 <https://doi.org/10.1146/annurev.psych.48.1.609>. as picked up e.g. in Kaptein and others.

[307] See for this e.g. Cosima Rughiniș, Răzvan Rughiniș, and Ștefania Matei, 'A Touching App Voice Thinking about Ethics of Persuasive Technology through an Analysis of Mobile Smoking-Cessation Apps', *Ethics and Information Technology*, 17.4 (2015), 295–309 <https://doi.org/10.1007/s10676-016-9385-1>.

[308] Of course, the device itself may (only) be considered Persuasive Technology to the extent it is used for applications like this, with the device otherwise not being an (active) exemplar. At the same time, the device may qualify as persuasive due to multiple overlapping applications that use it as an interface, e.g. for other health application. See for an overview over these applications typically bound to mobile phones Predrag Klasnja and Wanda Pratt, 'Healthcare in the Pocket: Mapping the Space of Mobile-Phone Health Interventions', *Journal of Biomedical Informatics*, 45.1 (2012), 184–98 <https://doi.org/10.1016/j.jbi.2011.08.017>.

[309] See e.g. Maryam Abo-Tabik, Yael Benn, and Nicholas Costen, 'Are Machine Learning Methods the Future for Smoking Cessation Apps?', *Sensors*, 21.13 (2021), 4254 <https://doi.org/10.3390/s21134254>.

[310] This mirrors the narrow and wide concepts of vectors of influence outlined in Section 7.2.

6.2.2 Persuasion

Persuasion is the act of inducing a result (e.g. a specific action or also just a mental state pertaining to that action) through appeal to a feature of an individual.[311] Much of this relies upon the type of responses individuals display when interacting with machines, in particular the similarities between responding to other humans and computers in which humans tend to exhibit *mindless* social responses.[312] For this purpose, some consider the use of the term *voice* to describe the specific information imparted by the technology and to denote that this interaction is usually conducted on the basis of certain social language cues.[313] If the persuasive element is less concerned with conveying a specific datum of information explicitly but instead takes a more structural form in its persuasion, we may also consider it an example of *choice architecture*, able to *nudge* its interactors.[314] In any case, persuasive technology tends to play of certain mental peculiarities and cognitive responses of humans, e.g. their heuristics, aversions or proneness to mental shortcuts. [315]

Naturally, there is a gradient between different shades of persuasion. Most unproblematically is "rational persuasion", i.e. an explicit appeal to reasons with acceptable degree of plausibility.[316] On the other end of the spectrum, teetering on the verge of persuasion and manipulation (and intuitively ethically more dubious) is a subversive appeal to subconscious features of the information recipient. To distinguish different types of persuasion, we can taxonomize persuasive technology along three dimensions[317]: transparency of persuasion, assertiveness and allocation of interest.

First, persuasive technology can impart persuasive quality on information with or without the knowledge of the information recipient. A subject that is aware of being exposed to persuasive technology can engage or disengage more consciously with the technology, bolstering their autonomy. The inverse of this holds true as well: Being unaware of influence measures levelled against oneself, one's autonomy is limited due to the lack of information about potential actions to take. The phenomenon of technology that effects an individual's mental state or consequent actions by enveloping them in an environment in which the influence is not perceived

[311] See e.g. Daniel O'Keefe, 'Persuasion', in *The Handbook of Communication Skills*, ed. by Owen Hargie, 4th edn (New York: Routledge, 2018), pp. 319–37 (p. 5). Nb. that there is a (somewhat arcane) debate in scholarship about the scope of potential persuasion, in particular if it should only apply to specific mental states e.g. excluding conviction to beliefs, and if it should encompass actions, see e.g. the discussion in J. Anthony Blair, 'Argumentation as Rational Persuasion', *Argumentation*, 26.1 (2012), 71–81 <https://doi.org/10.1007/s10503-011-9235-6>.

[312] See e.g. Clifford Nass and Youngme Moon, 'Machines and Mindlessness: Social Responses to Computers', *Journal of Social Issues*, 56.1 (2000), 81–103 <https://doi.org/10.1111/0022-4537.00153>.

[313] Rughiniș, Rughiniș, and Matei. Note however that this does not necessarily imply that the subject perceives the persuading element as having particular agency, see Cees Midden & Jaap Ham, *The Illusion of Agency: The Influence of the Agency of an Artificial Agent on Its Persuasive Power* 90–99 (2012), http://link.springer.com/10.1007/978-3-642-31037-9_8.

[314] See e.g. Richard Thaler, Cass Sunstein, and John Balz, 'Choice Architecture', in *The Behavioural Foundations of Public Policy* (Princeton University Press, 2013), pp. 428–39 <https://doi.org/10.1515/9781400845347-029>.

[315] See e.g. Kaptein and others.

[316] See e.g. Blair.

[317] Note that this is consistent with the definition given at the beginning of this entry, but may be out of the scope of some other conceptions of persuasive technology by individual theorists.

consciously is sometimes called *ambient persuasion*[318] or *unconscious persuasion*.[319] However, transparency is not necessarily a detriment to the effect of persuasive technology. Indeed some types of technology may rely upon the subject's knowledge of being observed, such as in many applications used for treatment purposes or the case of Fogg's surveillance strategies.[320]

Second, the persuasive quality imparted on information can vary in intensity. Persuasion that relies on exploitation of cognitive biases may be harder to "defend" against as opposed to persuasion based merely on more neutral information exposure. Examples for common mental peculiarities that may be exploited by persuasive technology pertain to human responses to perceived scarcity,[321] preferences for consistency in one's beliefs and reducing cognitive dissonance,[322] unproportionate loss aversion,[323] or the difficulty of pricing in sunk cost.[324] In this context, persistence is also a type of intensity modulator, as persistently interacting technology may erode an individual's self-control over time and thus be more "persuading".[325] Another measure of assertiveness is the extent in which the information is effectively personalized and targeted for the receiving individual.[326] Deploying sophisticated technology relying on data-laden profiles about the susceptibility of the target individual to influencing measures is prone to be more assertive.[327] This is salient in particular as different people (e.g. people of different genders or age) vary significantly in their responsiveness to certain persuasive strategies.[328]

---

[318] See e.g. Kaptein and others; Peter-Paul Verbeek, 'Ambient Intelligence and Persuasive Technology: The Blurring Boundaries Between Human and Technology', *NanoEthics*, 3.3 (2009), 231–42 <https://doi.org/10.1007/s11569-009-0077-8>; Saskia Maan and others, 'Making It Not Too Obvious: The Effect of Ambient Light Feedback on Space Heating Energy Consumption', *Energy Efficiency*, 4.2 (2011), 175–83 <https://doi.org/10.1007/s12053-010-9102-6>.

[319] See e.g. Peter A. M. Ruijten, Cees J. H. Midden, and Jaap Ham, 'Unconscious Persuasion Needs Goal-Striving', in *Proceedings of the 6th International Conference on Persuasive Technology Persuasive Technology and Design: Enhancing Sustainability and Health - PERSUASIVE '11* (New York, New York, USA: ACM Press, 2011), pp. 1–6 <https://doi.org/10.1145/2467803.2467807>.

[320] See e.g. Saskia Kelders, 'Involvement as a Working Mechanism for Persuasive Technology', in *PERSUASIVE 2015*, ed. by T. MacTavish and S. Basapur, 2015, pp. 3–14 <https://doi.org/DOI: 10.1007/978-3-319-20306-5_1>.

[321] Humans tend to emphasize value when they perceive the underlying item or service to be scarce, see e.g. S West, 'Increasing the Attractiveness of College Cafetaria Food: A Reactance Theory Perspective', *J Appl. Psychol*, 60 (1975), 656–58.

[322] See e.g. M Deutsch and H Gerard, 'A Study of Normative and Informational Social Influences upon Individual Judgment', *J Abnorm Soc Psychol*, 51 (1955), 629–636.

[323] See e.g. A Tversky and D Kahneman, 'The Framing of Decisions and the Psychology of Choice.', *Science*, 211 (1981), 453–458.

[324] See e.g. A Tversky and D Kahneman, 'Judgement under Uncertainty: Heuristics and Biases', *Science*, 185 (1974), 1124–31. Compare for all of these in more detail also R. CIALDINI, INFLUENCE, SCIENCE AND PRACTICE (2001), passim and Kaptein et al., *supra* note 2 also with more examples.

[325] See e.g. Jilles Smids, 'The Voluntariness of Persuasive Technology', 2012, pp. 123–32 <https://doi.org/10.1007/978-3-642-31037-9_11>. with reference to Mark Muraven and Roy F. Baumeister, 'Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle?', *Psychological Bulletin*, 126.2 (2000), 247–59 <https://doi.org/10.1037/0033-2909.126.2.247>.

[326] Maurits Clemens Kaptein, 'Formalizing Customization in Persuasive Technologies', in *PERSUASIVE 2015*, ed. by T. Mactavish and S. Basapur, 2015, pp. 27–38.

[327] See e.g. Maurits Kaptein and Dean Eckles, 'Selecting Effective Means to Any End: Futures and Ethics of Persuasion Profiling', 2010, pp. 82–93 <https://doi.org/10.1007/978-3-642-13226-1_10>.

[328] Orji Rita, Mandryk Regan, and Vassileva Julita, 'Gender, Age, and Responsiveness to Cialdini's Persuasion Strategies', in *PERSUASIVE 2015*, ed. by MacTavish T. and Basapur S., 2015, pp. 147–59.

Third, persuasive technology may attempt to effect change which benefits (or ought to benefit) primarily the information recipient, the technology provider or another party. The use of persuasive technology by an individual for the purposes of physical or mental health care clearly undergirds the persuasion with (legitimate) self-interest.[329] The same is not clear for the use of persuasive technology for the purposes of advertising, i.e. to effect a purchase by the information recipient. Using technology e.g. to avert an individual's propensity to vote in a political election then represents a clear conflict between the technology deployer's and the information recipient's interests.[330] persuasive technology benefiting an individual is not necessarily linked to that individual being aware of the deployment of said technology or the effect it has on them. However, there is an inherent tension in such applications connected to (contested) notions of legitimacy of paternalism, i.e. the notion of third-party interference in the interest of the person deprived.[331]

6.2.3 Intent

Persuasive technology is usually considered to require human intent to change attitudes or behaviours by using said technology.[332] If this is true, then technology must *inherit* human intent.[333] Fogg has suggested that intent may be inherited from the creator or designer of the technology (endogenous intent), by a provider of technology (exogenous intent) or by the subject of persuasion themselves (autogenous intent).[334] Again note that while many of the above insights are applicable to this inquiry, intent is not a necessary factor within this analysis

[329] See e.g. Tine Kolenik and Matjaz Gams, 'Persuasive Technology for Mental Health: One Step Closer to (Mental Health Care) Equality?', *IEEE Technology and Society Magazine*, 40.1 (2021), 80–86 <https://doi.org/10.1109/MTS.2021.3056288>; Thomas Fritz and others, 'Persuasive Technology in the Real World', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM, 2014), pp. 487–96 <https://doi.org/10.1145/2556288.2557383>. and a more systematic overview Oladapo Oyebode and others, 'Persuasive Mobile Apps for Health and Wellness: A Comparative Systematic Review', 2020, pp. 163–81 <https://doi.org/10.1007/978-3-030-45712-9_13>.

[330] See e.g. Mutale Nkonde and others, 'Disinformation Creep: ADOS and the Strategic Weapon-Ization of Breaking News', *Harvard Kennedy School Misinformation Review*, 2021 <https://doi.org/10.37016/mr-2020-52>.

[331] See generally for this tension GEORGE TSAI, 'Rational Persuasion as Paternalism', *Philosophy & Public Affairs*, 42.1 (2014), 78–112 <https://doi.org/10.1111/papa.12026>. and more specifically discussions with respect to Persuasive Technology e.g. Laura Specker Sullivan and Peter Reiner, 'Digital Wellness and Persuasive Technologies', *Philosophy & Technology*, 34.3 (2021), 413–24 <https://doi.org/10.1007/s13347-019-00376-5>; Andreas Spahn, 'And Lead Us (Not) into Persuasion…? Persuasive Technology and the Ethics of Communication', *Science and Engineering Ethics*, 18.4 (2012), 633–50 <https://doi.org/10.1007/s11948-011-9278-y>; Verbeek, 'Ambient Intelligence and Persuasive Technology: The Blurring Boundaries Between Human and Technology'.

[332] This notion may be traced back to definitions of persuasion in the fields of psychology and philosophy; it is not strictly necessary within the context of technology and some intuitive uses may encompass technology without a causal link to human intent. See for an affirmative viewpoint (with respect to requiring intent) on this issue BJ Fogg, *Persuasive computers, in* PROCEEDINGS OF THE SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS - CHI '98 225–232 (1998) and for a more contrarian viewpoint Johan Redström, 'Persuasive Design: Fringes and Foundations', in *Proceedings of the First International Conference on Persuasive Technology for Human Well-Being*, PERSUASIVE'06 (Berlin, Heidelberg: Springer, 2006), pp. 112–122.

[333] Note that this distinction is not always aligned with more practical concerns. A persuasive algorithm that effects specific behaviour in individuals that is exceeding the original intent of its creator, provider or user still elicits behaviour, and is still *persuasive*, even if it ought not to be classified as persuasive technology per se. See for this e.g. Alina Krischkowsky, Bernhard Maurer, and Manfred Tscheligi, 'Captology and Technology Appropriation: Unintended Use as a Source for Designing Persuasive Technologies', 2016, pp. 78–83 <https://doi.org/10.1007/978-3-319-31510-2_7>.

[334] Fogg, 'Persuasive Computers'.

(but may come as an important factor when considering ethical implications of autonomy constraints as outlined in Section 11.3.

6.3 Ethical Implications

Computer-assisted persuasion has been recognized as an area of ethical challenges very early on.[335] The main implications of persuasive technology tend to revolve around the notion of individual autonomy, i.e. the capacity of an individual to exercise self-government, and its derivative concepts, which makes engagement with this field fruitful for the purposes of this analysis.[336] Insofar persuasive technology is successful, and the imparted persuasion indeed effects a change of mental state or consequent action, the question arises if this change is still within the scope of an individual's own discretion or if an external decision has been ethically improperly imposed on them. Often these lines of ethical acceptability are denoted by a change of terminology. Technology that crosses the line may then deemed as coercive or manipulative as opposed to (merely) persuasive.[337] There is no definite consensus about the ethical boundaries of the field,[338] albeit some principles having been suggested by some,[339] but current discourse and intuition allows focus in particular on the concepts laid out below.

6.3.1 Voluntariness and Consent

A major consideration for the ethical use of persuasive technology, that is the use of persuasive technology most likely to preserve an individual's autonomy, centres around the notion of voluntariness and consent.[340] Voluntariness describes the absence of "stronger" forms of persuasion[341] such as manipulation or coercion combined with the intentional legitimation of the mental state or action difference effected by the technology. In other words, the individual ought to change their behaviour and participate in the persuasion process intentionally.[342] Voluntariness is then eroded both by lack of information, as informed consent becomes more difficult to exercise, and by increased assertiveness of the technology in question. This information is often understood widely; not only limited to the persuasion itself but also to

[335] Fogg, 'Persuasive Computers'.

[336] See e.g. Naomi Jacobs, 'Two Ethical Concerns about the Use of Persuasive Technology for Vulnerable People', *Bioethics*, 34.5 (2020), 519–26 <https://doi.org/10.1111/bioe.12683>. Note that the tension between persuasion and autonomy is not restricted to the domain of technology, see for an earlier discussion of this also David A. Strauss, *Persuasion, Autonomy, and Freedom of Expression*, 91 COLUMBIA LAW REV. 334 (1991) or in the domain of health care (where autonomy is considered as a central tenet in most matters and which arguably features the most in-depth discussions of autonomy beyond the domains of law and philosophy) Alisdair Maclean, 'Autonomy, Consent and Persuasion', *European Journal of Health Law*, 13.4 (2006), 321–38 <https://doi.org/10.1163/157180906779160274>.

[337] See e.g. Daniel Berdichevsky and Erik Neuenschwander, 'Toward an Ethics of Persuasive Technology', *Communications of the ACM*, 42.5 (1999), 51–58 <https://doi.org/10.1145/301353.301410>. Note that there is not a clear dividing line between these concepts and different authors have suggested different criteria to distinguish *acceptable* from *unacceptable* persuasive technology, see for this also Jacobs.

[338] See for a review of the available literature Raymond Kight and Sandra Burri Gram-Hansen, 'Do Ethics Matter in Persuasive Technology?', 2019, pp. 143–55 <https://doi.org/10.1007/978-3-030-17287-9_12>.

[339][339] See e.g. Fogg, 'Persuasive Computers'; Berdichevsky and Neuenschwander; Janet Davis, 'Design Methods for Ethical Persuasive Computing', in *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09* (New York, New York, USA: ACM Press, 2009), p. 1 <https://doi.org/10.1145/1541948.1541957>.

[340] Spahn.

[341] Again note that many definitions exclude coercion or manipulation from the descriptive scope of the term Persuasive Technology to begin with, but to use Fogg's words: "The line between persuasion and coercion can be a fine on, see Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do*, p. 21.f

[342] Jilles Smids, *The Voluntariness of Persuasive Technology* 123–132 (2012), http://link.springer.com/10.1007/978-3-642-31037-9_11.

encompass the intent behind the deployment of the persuading agent.[343] Highly assertive technology, e.g. by use of threats or applied force such as factual shaping of the interactive environment, may already be considered as coercive technology, meaning that voluntariness is not given anymore.[344] Consequently, the importance of giving (informed) consent to being subject to Persuasive Technology becomes more relevant as the assertiveness of the technology increases (and the user will face increasing difficulties breaking away from the behaviour preferred by the technology). Both of these concepts are strongly tied to the transparency of the persuasion process; persuasion that occurs without the subject's knowledge (and thus without their ability to extract themselves from being a target of persuasion) may require much stronger justification.[345] Barring sufficient justification (and sometimes even then), such technology is often denoted as *manipulative technology* as the individual is deprived of their agency in making conscious choices about their exposure.[346]

6.3.2 Privacy

As persuasive technology becomes more sophisticated and more often reliant on data of the subject or of groups the subject is a part of,[347] privacy and data protection concerns become more salient.[348] This is also of major practical relevance, as privacy and data protection are typically enshrined in legal regulation,[349] and again connects well to this thesis as autonomy and privacy are inherently intertwined.

6.3.3 Broadening of Persuasion Envelope

Another point of contention with persuasive technology pertains to mixed, or after-the-fact uses. An individual that consents (solely) to be subjected to technology meant to persuade him into a healthy eating habits cannot be said to have consented to the same technology being used to later persuade him to purchase other services by the technology provider.[350] In case of changes to the incentive-structure or use-envelope, it makes sense to also consider the *stickiness*

---

[343] See e.g. Berdichevsky and Neuenschwander, *supra* note 17; Bernardine M. C. Atkinson, *Captology: A Critical Review* 171–182 (2006), http://link.springer.com/10.1007/11755494_25.

[344] Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do*, p. 216.

[345] Just as above with respect to diverging interests between the subject and the technology deployer, this delta may be justified by some through notions of legitimate paternalism.

[346] Smids.

[347] There is a growing recognition within scholarship that individual identification is not necessarily required to target the very same individuals with high relevancy and accuracy, see e.g. Bart W. Schermer, *The limits of privacy in automated profiling and data mining*, 27 COMPUT. LAW SECUR. REV. 45–52 (2011), https://linkinghub.elsevier.com/retrieve/pii/S0267364910001767; Matthias Leese, *The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union*, 45 SECUR. DIALOGUE 494–511 (2014), http://journals.sagepub.com/doi/10.1177/0967010614544204; Brent Daniel Mittelstadt et al., *The ethics of algorithms: Mapping the debate*, 3 BIG DATA SOC. (2016).

[348] Privacy is often seen to serve as a prerequisite to autonomy, or at least as an enabling state hereto; at the same time (some concept of) privacy is considered to hold intrinsic value and therefore to warrant protection by nearly global consensus by itself in any case, see e.g. Benn, p. 241ff., Kupfer, p. 83 and as outlined in detail in Section 3

[349] This holds true in particular, as many applications of persuasive technologies may be found within the domain of e-health, and as such deal with sensitive data.

[350] The distinction between foreseeable and unforeseeable changes on how existing and originally consent-legitimized technology is used in the future has many parallels in the domain of data protection. Here, a horizon of data processing is sometimes assumed after which the individual cannot truly give (informed) consent anymore, as they do not possess sufficient information. See generally for the notion of consent with respect to data protection (here as exemplified in European Law) e.g. Stephen Breen, Karim Ouazzane, and Preeti Patel, 'GDPR: Is Your Consent Valid?', *Business Information Review*, 37.1 (2020), 19–24 <https://doi.org/10.1177/0266382120903254>.

of the technology in question, i.e. how easy or likely it is for an individual to extract themselves from being subject to the relevant technology. The more difficult a withdrawal from the use of or being subject to Persuasive Technology is, the stronger the underlying justification ought to be.

6.3.4 Veracity of Information

Further ethically relevant is the accuracy pertaining to information within the informational pipeline between the persuasive agent and the subject to be persuaded. Obviously, technology that persuades by using misleading or false information is ethically problematic, but lack of veracity with respect to information collected about the subject may also present ethical challenges. Insofar the user, based on their input or their observed behaviour, is profiled by the technology in question, and the profile is then used as a basis for further persuasion (or propagated to another algorithmic entity), an inaccurate or plainly wrong profile may lead to serious misalignment of the actual and intended persuasion, undermining whatever justification was originally available.[351]

6.4 Disambiguation

Contemporary discourse uses diverse terminology to describe different aspects of similar issues at the intersection of technology and autonomy.[352] This final section of the excursus highlights the overlap and differences between the phenomenon of Persuasive Technology and related concepts.

6.4.1 Design with Intent / Persuasive Design

Design with intent describes all design efforts intended to shape individual behaviour and hence serves as a superset of persuasive technology and other efforts intended to cause behavioural outcomes.[353] Persuasive technology is thus an example of design with intent.[354] Design with intent further exceeds the scope of persuasive technology as it also invariably encompasses coercive and manipulative design.[355] Strongly related but subordinate to this concept is persuasive design, a term which aims to describe the above but (originally) with diminished focus on intentionality (and potentially the exclusion of "harder" forms of persuasion such as coercion), recognizing that design may (always) impart shaping forces even without the intention of the designer.[356]

---

[351] See e.g. Spahn.

[352] See e.g. Roger Crisp, 'Persuasive Advertising, Autonomy, and the Creation of Desire', *Journal of Business Ethics*, 6.5 (1987), 413–18 <https://doi.org/10.1007/BF00382898>; Timothy Aylsworth, 'Autonomy and Manipulation: Refining the Argument Against Persuasive Advertising', *Journal of Business Ethics*, 175.4 (2022), 689–99 <https://doi.org/10.1007/s10551-020-04590-6>; Daniel Susser, Beate Roessler, and Helen Nissenbaum, 'Technology, Autonomy, and Manipulation', *Internet Policy Review*, 8.2 (2019).

[353] Dan Lockton, David H.T. Harrison, and Neville A Stanton, 'Choice Architecture and Design with Intent', 2009 <https://doi.org/10.14236/ewic/NDM2009.59>.

[354] Other types of design with intent may be found e.g. in manufacturing, urban planning or commerce, or any other domain in which the agent-environment is subject to intentional shaping, see for examples and more elaboration Dan Lockton, David Harrison, and Neville Stanton, 'Design with Intent: Persuasive Technology in a Wider Context', in *Persuasive Technology* (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 274–78 <https://doi.org/10.1007/978-3-540-68504-3_30>.

[355] Lockton, David Harrison, and Stanton.

[356] Redström.

6.4.2 Nudging

Nudging (in its original meaning) describes the imparting of influence on individuals through the deliberate shaping of contextual factors surrounding the choices that ought to be *nudged*. The concrete manifestation of these contextual factors is typically called the *choice architecture*.[357] Like persuasive technology, choice architecture is often built around (the same) specific human mental peculiarities. Indeed, many manifestations of persuasive technology deploy *nudges* by proactively shaping the (digital) choice architecture towards a specific intended outcome.[358] Similarly, many of the ethical issues raised by persuasive technologies parallel the issues raised by nudging.[359]

However, nudging does not necessarily require computer systems or other technology but can also be applied in analogue settings. There is also a conceptual difference in typical use-cases. While persuasive technology usually attempts to (positively) effect a change in behaviour through exposure, many examples of choice architecture face the nudged individual with pre-determined contextual factors prior to any interaction, relying on the inertia and passiveness of individuals and are thus embedded more deeply within the context of the choices to be made.[360]

6.4.3 Dark Patterns

Dark patterns denote a subset of persuasive technology (and subsequently a subset of persuasive design) in which intended actions of an individual are effected (or made more likely) by particular user-design elements, e.g. through the specific structure of a computer interface.[361] Examples of dark patterns include protracted procedures to withdraw from paid services or visual and procedural asymmetry in offering choices about data collection on websites through cookie-banners.[362] In contrast to persuasive technology as a whole, dark

[357] See e.g. Cass R Sunstein, 'Nudging: A Very Short Guide', *Journal of Consumer Policy*, 37.4 (2014), 583–88.
[358] Markus Weinmann, Christoph Schneider, and Jan vom Brocke, 'Digital Nudging', *Business & Information Systems Engineering*, 58.6 (2016), 433–36 <https://doi.org/10.1007/s12599-016-0453-1>.
[359] Just like persuasive technology is often thought to exclude the ethically more dubious variants of effecting compliant behaviour (e.g. manipulation or coercion), nudging is also usually considered to encompass only ethically unproblematic strategies; staying clear of any type of manipulations or coercion, see e.g. Mark Kosters and Jeroen Van der Heijden, 'From Mechanism to Virtue: Evaluating Nudge Theory', *Evaluation*, 21.3 (2015), 276–91 <https://doi.org/10.1177/1356389015590218>; Schmidt; Cass R Sunstein, 'The Ethics of Nudging', *Yale Journal on Regulation*, 32.2 (2015), 413–50.
[360] Most illustrative is the example of *default* or *preselected choices*, which tend to have a strong determinative link to the final choice affirmed or accepted by the individual. While persuasive technology undoubtedly can also use this strategy and exploit an individual's unwillingness to expend mental energy and time to change from a default state, nudging is often deployed even earlier in the process. For example, the practice of automatically enrolling individual's in a specific pension plan (bar any commitment to continue this arrangement) already creates consequences for the individual before any interaction between the individual and the respective choice architecture (and perhaps the persuasive technology used at a later stage in the persuasion process. See for the example used specifically e.g. Todd J Zywicki, 'Do Americans Really Save Too Little and Should We Nudge Them to Save More: The Ethics of Nudging Retirement Savings Symposium: The Ethics of Nudging - Evaluating Libertarian Paternalism: Implications, Extensions, and Applications', *Georgetown Journal of Law & Public Policy*, 14.Special 2016 (2016), 877–920.
[361] Colin M. Gray, Yubo Kou, and others, 'The Dark (Patterns) Side of UX Design', in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM, 2018), pp. 1–14 <https://doi.org/10.1145/3173574.3174108>; Arvind Narayanan and others, 'Dark Patterns: Past, Present, and Future', *Queue*, 18.2 (2020), 67–92 <https://doi.org/10.1145/3400899.3400901>.
[362] See e.g. Christoph Bösch and others, 'Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns', *Proceedings on Privacy Enhancing Technologies*, 2016.4 (2016), 237–54; Midas Nouwens and others, 'Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence', in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM, 2020), pp. 1–13 <https://doi.org/10.1145/3313831.3376321>.

patterns are deployed irrespective or in opposition to the interest of the subject of persuasion, and generally mean to overcome reluctance or aversion of an individual towards a certain action or choice, and are thus seen more problematic from ethical and legal standpoints.[363] Recently, dark patterns also have received increased attention and scrutiny from data protection watchdogs due such as the European Data Protection Board.[364] (As will be outlined later in Section 12, dark patterns present one of the most current attack points for the European regulator to protect autonomy within the digital domain).

6.5 Conclusion: From Persuasive Technology to Autonomy Constraints in the IoE

Research sub-question 2.1 concerned itself with the connection of persuasive technology to understanding autonomy and autonomy constraints in the IoE. As this section has shown, the domain of persuasive technology connects very well to the inquiry at hand and provides a wealth of research in which the effectiveness and ethicality of certain persuasion strategies (and as a result autonomy constraints under the pragmatic account of autonomy) have been analysed already. While there are differences in scope to be aware of, most notably that persuasive technology excludes by definition coercive and manipulative technology that also present autonomy risks, and that it does not necessarily mandate technology to be of sufficient complexity to warrant being included in the IoE, it still presents one of the richer fields to consider for this inquiry. The existing debate and academic findings and conceptualizations both structurally and ethically are useful beyond their initial scope and may inform investigations into other types of autonomy-constraining autonomy all the same. As a result, many of the same themes that were touched upon in this section will be reflected in subsequent analysis, proving both their relevance, and the analysis grounding in existing research at the same time.

[363] See e.g. Gregory Day and Abbey Stemler, 'Are Dark Patterns Anticompetitive? ', *Alabama Law Review*, 72.1, 1–46 <https://heinonline.org/HOL/P?h=hein.journals/bamalr72&i=11>; Davide Maria Parrilli and Rodrigo Hernandez-Ramirez, 'Re-Designing Dark Patterns to Improve Privacy', in *2020 IEEE International Symposium on Technology and Society (ISTAS)* (IEEE, 2020), pp. 253–54 <https://doi.org/10.1109/ISTAS50296.2020.9462197>; Diana MacDonald, 'Anti-Patterns and Dark Patterns', in *Practical UI Patterns for Design Systems* (Berkeley, CA: Apress, 2019), pp. 193–221 <https://doi.org/10.1007/978-1-4842-4938-3_5>; Ari Ezra Waldman, 'Cognitive Biases, Dark Patterns, and the "Privacy Paradox"', *Current Opinion in Psychology*, 31 (2020), 105–9 <https://doi.org/10.1016/j.copsyc.2019.08.025>; Than Htut Soe and others, 'Circumvention by Design - Dark Patterns in Cookie Consent for Online News Outlets', in *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (New York, NY, USA: ACM, 2020), pp. 1–12 <https://doi.org/10.1145/3419249.3420132>; Saul Greenberg and others, 'Dark Patterns in Proxemic Interactions', in *Proceedings of the 2014 Conference on Designing Interactive Systems* (New York, NY, USA: ACM, 2014), pp. 523–32 <https://doi.org/10.1145/2598510.2598541>; Colin M. Gray, Cristiana Santos, and others, 'Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective', in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM, 2021), pp. 1–18 <https://doi.org/10.1145/3411764.3445779>.

[364] European Data Protection Board, *Guidelines 3/2022 on Dark Patterns in Social Media Platform Interfaces: How to Recognise and Avoid Them*, 2022 <https://edpb.europa.eu/system/files/2022-03/edpb_03-2022_guidelines_on_dark_patterns_in_social_media_platform_interfaces_en.pdf>.

7 Fundamental Concepts: The Informational Pipeline and Vectors of Influence

7.1 The Informational Pipeline

To set the stage for the upcoming analysis, I will first consider the aforementioned *informational pipeline* as an apt domain of analysis. In particular, this concept describes the domain informational agents are embedded in and constitutive of. Simplified, this concept may be described as a chain of devices linked to each other spanning from the source of information to the individual receiver; but this is a simplistic assumption. More precisely, the informational pipeline consists of all *processes* and their manifestations that are used to transmit information. This includes the spheres of

- the information origin, such as
    - *explicit* or *real* information origins (e.g. actions of individuals and environmental phenomena), and
    - *derivative* or *synthetic* information origins (in particular information pertaining to the informational pipeline itself, etc.), as well as
    - all entities and artefacts in which this information is manifested (e.g. data storage devices, but also individuals and environmental factors)
- the transfer and transmission infrastructure, such as
    - sending and receiving processes (e.g. signal synchronization, modulation, amplification, error detection), and
    - matters of information processing (e.g. encryption and decryption) and again
    - all entities and artefacts in which these processes are manifest, and
- the information receiver, such as
    - intermediation and translation processes bridging the digital and human-accessible analogue spheres (e.g. presentation processes, audio-visual translations of digital data, etc.),
    - peculiarities of the receiving individual or entity, and
    - all entities and artefacts in which these processes are manifest.

We can imagine information *flowing* through this pipeline from a point of origin to a point of reception, and consequently consider an information transfer as a function of information and the informational pipeline it traverses. Notably, this definition is not purely consisting of technological processes or devices; indeed it includes non-digital processes (e.g. certain cognitive biases of the information receiver) or medium characteristics (e.g. if digital information is translated into a different medium e.g. in case of printing on paper), as well as environmental factors. Put bluntly, the informational pipeline is the actual or potential path an information takes and consists of all processes (and their manifestations) that affect that information transfer.

When considering the domain of the IoE, it is clear that the informational pipeline is not necessarily congruent but may have its start and endpoints outside of the IoE. Similarly, the informational pipeline may intersect the IoE only intermittently in case of more analogue information transfer processes that are interjected between digital processes. This necessitates a caveat for simplified notation: Where this thesis mentions the informational pipeline, it

generally does so with respect to its intersection with actual technology.[365] Upcoming sections will investigate the relevance of characteristics of the informational pipeline with respect to autonomy will be investigated further. In addition, this concept is also useful to derive a formal definition for informational agents.

7.2 Defining Vectors of Influence

When considering the impact of information technology, the entire informational pipeline is of interest. If and how a process can or will impart constraints onto an individual's autonomy may principally be affected by any traits of the agent, the information recipient and the characteristics of the information transfer. It is exactly this entire scope of analysis that is here characterized as *vector of influence* to better conceptualize autonomy constraints.

Explicitly, I define this concept here as follows:

- a vector of influence in a narrow sense are the compound phenomena that characterize a given interaction between an informational agent[366] and another entity that is capable of imparting change onto an individual's autonomy.

Information then flows along and is affected by the qualities of the respective vector of influence. For example, the capability of an algorithm to target individuals with personalized misinformation aligned with their cognitive biases and vulnerabilities characterizes the vector of influence along which the (mis)information is transferred just like the distrust of the recipient in the (mis)information due to media literacy or scepticism with respect to the origin of said (mis)information.

This concept can be expanded further to include the physical manifestations of informational agents:

- A vector of influence in a wide sense describes the informational agent, its procedural and physical structure and its capabilities (including the vectors of influence in a narrow sense at its disposal) to impart change onto an individual's autonomy more generally.

Under this definition, we may consider the intermediation artefacts used by informational agents (such as computers or smartphones), their design and their respective effects on the information transfer in addition to any (intrinsic) qualities of agent and recipient.[367] Connecting this to the elaboration of autonomy before, we can say that a vector of influence determines the constraining of autonomy (both actual and potential) by technology. More simply, we may say that a respective vector of influence *is* an autonomy constraint, if some of its qualities lead to an individual's autonomy being undermined. As technology becomes more powerful and more

[365] However, these lines can not always be precisely drawn. For example, vectors of influence can be affected by extra-technological factors such as an information recipient's trust into the information flow.
[366] Note that the concept of an informational agent will be defined in a later section. The reason why this definition includes a reference to informational agent lies within the usefulness of the term as a collective denominator of singular processes.
[367] For example, a navigation device outfitted with an informational agent that provides the relevant location-based services may be a vector of influence in a wide sense. (The type of information output, the design and characteristics of the output and interaction interface and the context of use of the location-based services device describe the vector of influence in a narrow sense.)

widespread, it is a trivial conclusion that the risk to individual autonomy as a result of these technological phenomena is increasing all the same.

Recalling the above definition of vectors of influence, we may distinguish them by their general objects of reference. First, aligning this with the previously introduced concept of the informational pipeline, we can generally situate these vectors in the nexus of information source, transfer processes or information recipient. The below analysis reflects this compartmentalization. The analysis of the general domain of information technology and the subsequent analysis of agent characteristics can be mapped to the segments of information origin and transfer processes. The analysis of information receival particularities maps well to the segment of the information recipient. The following sections follow this logic and investigate the vectors of influence most closely connected to the technology domain generally, to informational agents and to the information recipients in this order.

8 The Domain of Information Technology

When attempting to understand what role information technology plays in enabling and facilitating autonomy constraints, it makes sense to sense to start at the most fundamental vectors of influence: the characteristics of the information technology domain. Research sub-question 2.2 considers what characteristics of the IoE at large are potentially relevant for understanding the risk of technologically derived autonomy constraints. In pursuit of this question, this section outlines two sets of domain characteristics: physical characteristics and meta characteristics that are relevant when considering the inquiry at hand.

8.1 Physical Domain Characteristics

8.1.1 Factors of Relevance

Generally, importance and prevalence of the Internet seems to be increasing.[368] The term Internet of Everything (IoE) describes one of the more recent permutations of the Internet as the accumulation of technology and infrastructure that interweave (digitally) most or all aspects of human life.[369]

For the purposes of this thesis, one can start by understanding the evolving permutations of the Internet as (1) a set of devices serving as endpoints and interfaces that bridge the divide between the digital and the physical sphere, and (2) the connections between these devices. Internet prevalence and importance are determined by these factors. A change of the characteristic of the Internet is then caused by a change that pertains to these interfaces and their connections. This can happen due to different reasons that can be grouped as follows:

- the capability of devices that are connected to the Internet changes,
- the capability of the connection between such devices' changes (e.g. in speed or reliability), and
- the number of devices that are connected to the Internet changes; and hence the number of connections that make up the Internet changes, e.g. by introduction of new devices or by outfitting existing ("offline") devices with the capability of connectivity, thus taking them "online", but also by creating connectivity infrastructure that connects "offline but capable" devices.

And indeed, all three of these constitutive phenomena have been undergoing changes compared to a pre-information age.

8.1.2 Changes to Device Capability

It seems evident that since the inception of the early Internet, to the extent changes occurred over a period of time, these changes have been increases both in capabilities and numbers of devices and connections. Computers and similar devices have increased their theoretical and actual capabilities due to the increasing sophistication of the underlying hardware, e.g. the

[368] Instead of many reference for this truism, just see Naughton.
[369] See Sub-Section 1.1, and also generally CISCO.

increasing numbers of transistors used in integrated circuits as outlined in Moore's Law[370], as well as due to advances in computer science allowing the creation, testing and application of increasingly complex routines (i.e. *algorithms*). Increasing calculating speed of underlying devices can not only speed up existing computer operations but can open the door to novel forms of data processing. For example, the now omnipresent term "machine learning" was first introduced in 1959,[371] and the concept of artificial neural networks was discussed even before that.[372] However, widespread application of machine learning algorithms in the way they are now deployed at the time of writing were unfeasible not just due to the lack of depth of knowledge about such concepts but also due to the structure and speed of existing hardware. The increase in computer capability translated machine learning from a specimen of the theoretical realm to a highly relevant technology that is widely used today.[373]

8.1.3 Changes to Infrastructure Capability

The same is true when considering the change to the quality of interconnectivity of devices. Most importantly, bandwidth of Internet infrastructure is increasing steadily. This phenomenon is reflected in an observation named "Edholm's Law", which predicts a doubling of telecommunication bandwidth every 18 months.[374] Related to the increasing capability of bandwidth infrastructure, bandwidth usage is also increasing. The International Telecommunication Union estimates bandwidth usage growth for 2020 to amount to 38% (6 percentage points more than the previous year), and the trend has historically only gone upward.[375] Interconnectivity on a device-level can be achieved via wireless (WLAN, cellular networks, satellite, etc.) or wired (cable, fiber, powerlines, etc.) means.[376] Increasing capabilities of wireless networks have also led to a sharp increase of wirelessly connected devices after 1990, signalling a shift towards wireless internet access in many use cases, a phenomenon that has been dubbed the "wireless revolution".[377] This holds true for fully wireless devices that establish their own connection as well; by the end of 2021, market analyst

[370] See Gordon E. Moore, 'Cramming More Components onto Integrated Circuits', *Electronics*, 38.8 (1965), 114ff. Of course, Moore's Law is an empirical observation, not a definitive rule, and is expected to "end" between 2020 and 2030, see e.g. David Rotman, 'We're Not Prepared for the End of Moore's Law', *MIT Technology Review*, 2020.

[371] A. L. Samuel, 'Some Studies in Machine Learning Using the Game of Checkers', *IBM Journal of Research and Development*, 3.3 (1959), 210–29 <https://doi.org/10.1147/rd.33.0210>.

[372] See e.g. S. C. Kleene, 'Representation of Events in Nerve Nets and Finite Automata', in *Automata Studies. (AM-34)* (Princeton University Press, 1956), pp. 3–42 <https://doi.org/10.1515/9781400882618-002>; B. Farley and W. Clark, 'Simulation of Self-Organizing Systems by Digital Computer', *Transactions of the IRE Professional Group on Information Theory*, 4.4 (1954), 76–84 <https://doi.org/10.1109/TIT.1954.1057468>.

[373] See for an brief overview over the period of development of such concepts and the constraints to it Alexander L. Fradkov, 'Early History of Machine Learning', *IFAC-PapersOnLine*, 53.2 (2020), 1385–90 <https://doi.org/10.1016/j.ifacol.2020.12.1888>.

[374] S. Cherry, 'Edholm's Law of Bandwidth', *IEEE Spectrum*, 41.7 (2004), 58–60 <https://doi.org/10.1109/MSPEC.2004.1309810>.

[375] International Telecommunication Union, *Measuring Digital Development - Facts and Figures*, 2020, p. 10.

[376] There is nuanced middle ground between such types of connectivity; e.g. Steven Cherry distinguishes between three types of connectivity in the aforementioned publication: wireless, nomadic and wired. The middle concept describes connectivity that is somewhat bound to a location but within its sphere allows for wireless connectivity; e.g. in the case of a Wi-Fi Router, see Cherry. However, this distinction is not always necessary; it is entirely possible to see the nomadic devices itself as (likely) wired access point while further devices within its sphere can be characterized as wireless devices.

[377] T.S. Rappaport, 'The Wireless Revolution', *IEEE Communications Magazine*, 29.11 (1991), 52–71 <https://doi.org/10.1109/35.109666>. This is aided not just by added use-cases for devices in which a wired connection would invalidate the purpose of the device (e.g. mobile phones) but also by the fact that wireless connections increased their speed faster than wired connections, see Cherry.

firm Ericsson estimates that there will be 580 million 5G subscriptions, with an establishment rate that is quicker than the previous technology, leading to 3.5 billion subscriptions by 2026.[378]

8.1.4 Changes to Prevalence of Connected Devices

However, the most important factor when considering the relevance of the Internet likely is the change in number of connections and connected devices, and as a result of this connected persons. This change is driven by two separate trends: (1) an increase in the internet penetration rate and (2) the adoption of connectivity capabilities by devices of a class or function which previously did not utilize any form of connections.

8.1.4.1 Connectivity Growth by Increase of Traditional Internet Devices

Internet penetration rate measures how many individuals of a certain group (e.g. citizens of a country) have access to the Internet. This number has been increasing drastically since the inception of the Internet and its introduction in the public. Changes to penetration rate can be drastic. For example, according to some estimates, there were an estimated 407 million Internet users at the end of November 2000, which represented a doubling from the previous year, and a huge increase from the 26 million users in 1996.[379] The International Telecommunication Union estimates that at the end of 2019, about half of the world population was using the internet, with more than 69% of people aged between 15 and 24 years. Internet Access is dependent on geo-political factors[380] as well, and not uniform globally, e.g. total Internet penetration rate in Europe is at 83% (96% for youths), while Africa currently sits at 29% (40% for youths).[381] This phenomenon is dubbed the "digital divide".[382] Nonetheless, despite these differences and disparities, Internet penetration rate is increasing across all regions globally. This means not only that more and more people are connected to the Internet, but also that each of the people connected own or have access to at least one interned connected device. An increase in Internet penetration will always likely go hand in hand with an increase in available connected devices; either by a general increase in number of devices or potentially by a transfer of devices from one region to another. This is reflected in the numbers of connected devices, that are on the rise globally as well. Devices that have traditionally been the most common and conscious interfaces between their users and the Internet, i.e. personal computers and more recently mobile computing devices such as phones or tablets, are seemingly nearing a point of saturation, with a clear supermajority of people in many developed countries having access to such technology. However, due to the digital divide mentioned above, there may still be substantial demand in developing countries, that may boost the number of personal computers

[378] Ericsson, *Mobility Report*, 2021, p. 3f.
[379] Amy C. Bradshaw, 'Internet Users Worldwide', *Educational Technology Research and Development*, 49.4 (2001), 112–17 (p. 112) <https://doi.org/10.1007/BF02504952>.
[380] Indeed, scholars have found positive correlation between the Human Development Index, as assessed by the United Nations Development Program, and Internet penetration rate, see e.g. Ahmad Pratama and Moneer Al-Shaikh, 'Relation and Growth of Internet Penetration Rate with Human Development Level from 2000 to 2010', *Communications of the IBIMA*, 2012, 1–8 <https://doi.org/10.5171/2012.778309>.
[381] Similarly, there is a gender disparity in internet users, with 55% of the global male population but only 48% of the global female population having internet access. Only specific cohorts (i.e. the Americas and Small Island Development States, in short SIDS) show higher access for women than men. See International Telecommunication Union, p. 7ff.
[382] See e.g. Everett M. Rogers, 'The Digital Divide', *Convergence: The International Journal of Research into New Media Technologies*, 7.4 (2001), 96–111 <https://doi.org/10.1177/135485650100700406>; M. D. Chinn and R. W. Fairlie, 'The Determinants of the Global Digital Divide: A Cross-Country Analysis of Computer and Internet Penetration', *Oxford Economic Papers*, 59.1 (2006), 16–44 <https://doi.org/10.1093/oep/gpl024> and more broadly Pippa Norris, *Digital Divide* (Cambridge University Press, 2001) <https://doi.org/10.1017/CBO9781139164887>, passim.

or mobile connected devices such as phones.[383] Recalling the previous statement at the beginning of this section, this suggests the plausible conclusion that the increase in Internet penetration rate corresponds to an increase in connected devices which leads to an increase in the (prevalence and) importance of the Internet per se.

8.1.4.2 Connectivity Growth by Introduction of Non-Traditional Internet Devices

However, the increase of personal computers and mobile phones does not fully capture the omnipresence of the Internet in many regions and with respect to the exercise of life by many people, particularly in the "western world" at the time of writing. Instead, internet connectivity has also been added to classes of devices that previously did not feature any connective capability. These devices generally retain their previous main function but try to take advantage of an Internet Connection to add value in certain ways. Such devices are now often identified by the addition of the word "smart" as a qualifier. For example a refrigerator with connective capabilities (hence called a smart refrigerator, as this class of devices is not a traditional Internet device) will retain its main function. i.e. to refrigerate food (or other items) but adds Internet connectivity for the convenience of the user, e.g. to pull recipes from the Internet or maintain shopping lists.[384]

The superset of non-traditional Internet devices, or "smart" devices overlap with the use of the term "Internet of Things" (IoT). However, "smart" devices or devices that can conceivably be counted as belonging to the "Internet of Things" do not always have to be connected to the Internet per se. IoT-devices may be connected to each other in a certain environment without a connection between the network that they create and the network that is commonly understood to be the Internet at large. To continue the example used above, a smart fridge may be connected to a smart thermostat, a smart doorbell or a smart door lock, but their network may be closed and local, without any of them being connected to the "wider" Internet via an interfacing device such as a modem/router or if connected, utilizing that connection. In practice, this distinction is of less interest, as manufacturers of "smart" devices are heavily incentivized to access usage data.

Non-traditional Internet devices are also increasing in numbers. Their adoption is likely aided by the existing knowledge and infrastructure introduced and propagated by traditional Internet Devices, but also by their respective different use cases and often lower price points. [385] This is plausible from an intuitive standpoint: generally, people have use only for a limited amount of personal computers or phones and these devices used to be the primary connected devices, which puts a natural asymptotic limit onto the number of active devices any environment will accumulate and benefit from. The situation is different for non-traditional Internet devices however: First, these devices fulfil different roles and hence may replace existing "dumb" technology. For example, many households in the western world's will have access to a range of electric appliances that can be made "smart" and interconnected such as kitchen equipment (e.g. ovens, microwaves, fridges, water boilers, coffee machines, kitchen aids, blenders), hygiene equipment (e.g. washing machines, dryers, electric toothbrushes)), entertainment

[383] See for personal computers e.g. World Economic Forum, *Global Information Technology Report*, 2016, p. 11. The report has since been discontinued. For smartphones see e.g. Pew Research Center, *Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally*, 2019.
[384] Aurel-Dorian Floarea and Valentin Sgarciu, 'Smart Refrigerator: A next Generation Refrigerator Connected to the IoT', in *2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (IEEE, 2016), pp. 1–6 <https://doi.org/10.1109/ECAI.2016.7861170>.
[385] Ericsson, *IoT Connection Outlook* <https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/iot-connections-outlook>.

equipment (e.g. TVs, HiFi equipment, satellite receivers, game consoles), communication equipment (e.g. landline phones, mobile phones, answering devices), productivity devices (e.g. printers, fax machines, scanners) or even general furniture (e.g. light sources, reclining chairs, standing desks.).[386] On top of this, many other devices, such as sport equipment (e.g. watches, heartrate-monitors, ergometers), security equipment (e.g. security cameras) or pet paraphernalia (e.g. cat food portioners, automatic litterboxes) may lend themselves well to further connectivity. Even when assuming that the number of such devices and appliances will remain static[387], this would still represent a huge increase of interconnected devices per person or household compared to the more archetypical "one smartphone or one personal computer per person"-model that tracks traditional connected devices.

8.2 Domain – Meta Characteristics

The utility and power of an interconnected system is not merely defined by its physical manifestation. Instead one also ought to consider the more derivative phenomena of the physical paradigm changes outlined above. The changes in device and infrastructure capability and device prevalence have led to changes in how the domain can be perceived to function as a whole. Relevant to the scope of this inquiry, that is relevant to how autonomy is supported or undermined within the IoE, the following five factors are of particular relevance:

- intermediation of information
- immediacy of information transfer
- lower barriers to impart (mis)information
- anonymity or pseudonymity of information transfer participants
- plurality of information recipients
- personalized information exposure
- easy implementation of scalable, non-human actors.

8.2.1 Intermediation of Information

Information transferred digitally is, if not machine-readable in its content, machine-accessible in its meta-data. It is however not human readable without any additional tools. Digital information as it is transferred, processed or stored within a circuit of machines may manifest itself in binary states of electricity currents or on/off states within a transistor, in modulated signals in optical fiber cables or in more permanent form as entries within NAND flash storage

[386] For example, the average Austrian household in 2009 was believed to have between 17 and 29 of some of these electric appliances, see Sara Ghaemi and Guenther Brauner, 'User Behaviour and Patterns of Electricity Use for Energy Saving', in *6. Internationale Energiewirtschaftstagung an Der TU Wien [Proceedings]*, 2009. Similarly, a survey in the USA conducted in 2014 suggested between 10 and 17 home appliances per household depending on the number of people, see An Na Won and Won Hwa Hong, 'A Survey on Ownership of Home Appliances and Electric Energy Consumption Status According to the Number of Household Member', *Applied Mechanics and Materials*, 672–674 (2014), 2165–68 <https://doi.org/10.4028/www.scientific.net/AMM.672-674.2165>. (Of course studies like this are not directly comparable due to different scoping, but they do serve well here to illustrate the larger point made in the main text).

[387] This seems like an extremely conservative and ill-founded assumption for multiple reasons: Even if assuming that households will max out at a certain number of devices (and this number is already reached), there are still many consumer markets and households therein that do not have access to the same range of equipment for (inter alia) economic reasons. However, as these demographics are projected to develop increasing economic strength, it is likely that the demand for these type of appliances will increase. See for some of these projections (albeit through the lens of energy consumption connected to these devices more generally) e.g. Catherine Wolfram, Orie Shelef, and Paul Gertler, 'How Will Energy Demand Develop in the Developing World?', *Journal of Economic Perspectives*, 26.1 (2012), 119–38 <https://doi.org/10.1257/jep.26.1.119>.

or on magnetic discs in hard drives. None of this is directly accessible to the human senses. Before individuals interact with information stored, processed or transferred, it must be translated into a human-readable format via an interfacing device running some sort of intermediary agent, which gives the agent the opportunity to exert influence on its user. In short, all information transfer from the digital domain to a human recipient undergoes a process of *intermediation*[388] via some sort of interfacing device. This is not limited (or exhausted) by the intermediation e.g. of a personal computer but requires a translation into a medium in which human senses can perceive the information by themselves. Typical intermediation artefacts are screens for visual translation of information and speakers or headphones for translation of information into sound. Within this thesis, such interfaces will also be called *intermediating artefacts*.

8.2.2 Immediacy of Information Transfer

It is a trivial observation that a key characteristic of information technology is the speed of which information is transferred. The speed of transfer of information (within the domain of information technology) is dependent on a few factors. First, the transfer itself is limited by physical factors: based on current knowledge, it is plausible that no information can be transferred faster than the speed of light. Current real-life infrastructure such as fiber-optic cables can achieve around two-thirds of that speed, with further research into different materials ongoing that may close the gap further in the future. [389] Second, considering that information transferred over the internet is typically not a single binary state but a more complex pattern (e.g. binary code), bandwidth, i.e. the amount of information that can be transferred over a certain period of time, becomes an important metric when considering the limits of information transfer. However, as discussed previously, the bandwidth of Internet infrastructure has been increasing steadily as well.[390] This means that for many intents and purposes information transfer from endpoint to endpoint is almost instantaneous, in particular with respect to text or data communication.

However this kind of information transfer only reflects the speed of transfer between "Internet endpoints", leaving an "access gap" between human and machine. Put bluntly, information may be "immediately" transferred to an interfacing machine (i.e. an *intermediating artefact*), e.g. a personal computer, but the human user as ultimate information recipient may not receive it until some time has passed. However, this gap has also shrunk with interconnectivity on devices that are portable having also increased in prevalence and competency[391]. As individuals are more likely to use and carry with them increasingly powerful and fast machines that act as endpoints such as smartphones or IoT-devices, the speed of information transfer between the endpoint and the user can potentially decrease drastically.[392]

---

[388] Nb that this is different from similar economic phenomena within this domain, confusingly also often termed intermediation. Within this other context intermediation describe the establishment of some sort of marketplace or broker which sits between as service or goods provider and the end-costumer, see e.g. Richard Hawkins, Robin Mansell, and W Edward Steinmueller, 'Toward Digital Intermediation in the Information Society', *Journal of Economic Issues*, 33.2 (1999), 383–91..

[389] F. Poletti and others, 'Towards High-Capacity Fibre-Optic Communications at the Speed of Light in Vacuum', *Nature Photonics*, 7.4 (2013), 279–84 <https://doi.org/10.1038/nphoton.2013.45>.

[390] International Telecommunication Union, *Measuring Digital Development - Facts and Figures*.

[391] Rappaport; Ericsson, *Mobility Report*.

[392] This has implications about information consumption as well, for example see the establishment of the so called 24 hour news cycle Eric Bucy, Walter Gantz, and Zheng Wang, 'Media Technology and the 24-Hour News Cycle', in *Communication Technology and Social Change* (London/New York: Routledge, 2007).

When comparing this to pre-information age technology, the order of magnitude in transferring speed of information is immediately apparent. Illustrating this, we may compare the information transfer speed of written text by comparing a physical letter, travelling around the world along established postal routes and taking weeks or months, to the almost immediate sending and receival of an e-mail. It is also likely that the split between time of information spent in transit and time that information access is dependent only on the recipients' actions has changed. Considering again the voyage of a hypothetical letter for a substantial amount of time and then the time between the letter reaching its final destination (e.g. a post office or letterbox) and contraposing this to the time it takes the recipient to obtain and read the contents of the letter will generally reveal that the latter describes a shorter period than the former. Due to the almost instant nature of information transfer through information technology, this split may have reversed in many situations. An e-mail or text message sent is received by the service provider almost immediately, and in short succession also by the final recipient's endpoint device (with some exceptions, e.g. in case of subpar connectivity service reception). Instead, the time in which the message has been "sitting" at the point of intermediation waiting to be noticed and "read" by the recipient, while likely also short, has become the larger period in comparison.

8.2.3 Low Barriers to Impart Information

Individuals and institutions partaking in digital information transfer face comparatively little obstacles towards imparting information. For once, many of the Internet services necessary to transfer information do not have monetary costs to use beyond the purchase of an Internet endpoint, the subscription to the services of an Internet Provider and the nominal cost of electricity incurred by the use of Internet endpoint devices. This has implications both for individuals and for larger and more powerful entities such as institutional actors or businesses. For both, these costs may still be prohibitive in some cases; as mentioned before, total Internet penetration rate varies significantly around the world in favour of "western" societies along the aforementioned "digital divide ". Similar purchasing power may not be able to acquire the same services from ISPs everywhere, either because of price differences or availability constraints. Nonetheless, despite these differences and disparities, Internet penetration rate is increasing across all regions globally, suggesting that the cost obstacles to access is generally decreasing. Beyond these initial costs, differences in monetary impact between on- and offline variations of similar information transfers become clear, e.g. sending letters will incur the cost of postage while sending e-mails or internet text messages (e.g. via WhatsApp, Telegram, Signal, etc.) will typically not.[393] Lower barriers also bring with them increased incentive to use the information transfer process grown more attractive this way. Individuals can use the digital domain to interact with each other's, entertain themselves or use it for professional reasons, allowing for a clear increase of connection potential compared to their pre-digital selves. Economic entities such as businesses can use the digital domain to propagate information about their services and goods, if not more effectively then perhaps to a greater audience.[394]

[393] See e.g. Karen Church and Rodrigo de Oliveira, 'What's up with Whatsapp?', in *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '13* (New York, New York, USA: ACM Press, 2013), p. 352 <https://doi.org/10.1145/2493190.2493225>.

[394] This notwithstanding, as the field of online advertising embedded e.g. in open online media is becoming more saturated and the open online media providers have begun to finetune their information-ranking algorithms towards their own economic incentives, the cost of effective information transfer has likely increased again in many cases. For example, a business posting on a social network about its sales is likely not prioritized (anymore) by the social network's algorithm, if the business does not pay for prioritized ranking (like with general advertising). However, this observation concerns the discoverability of the information compared to

8.2.4 Plurality of Information Recipients

All Information transfer may be dyadic or multiadic. Dyadic transfer describes the transfer from one entity to another, while multiadic transfer describes the transfer from one entity to a plurality of other entities. Multiadic information transfer is not a new phenomenon, indeed the offline media-landscape addresses many recipients by definition. Newspapers, radio and TV-channels do not represent pinpointed communication targeted at single individuals but rather are a fanned-out way of communicating information. However, multiadic information transfer has become relatively more accessible than dyadic information transfer. Naturally, this is connected to the previous characteristic of the digital domain, i.e. low barriers to impart information: put bluntly, as it gets easier to contact people, it gets easier to contact *many* people. Pre-internet, the necessary infrastructure cost scaled somewhat with recipients, as one needed more printing presses or radio towers to reach more and more people. Physical mailing to a large group of people was (and still is) significantly more expensive than digital transfer of mass e-mails. Publication of a tweet or blog post utilizing existing social network providers is "free" compared to the cost of buying an announcement slot in a traditional newspaper.[395] The factors of immediacy and reduced obstacles discussed above lend themselves well to increased access to a large group of recipients for information transfer. This applies not only to sharing original information but also to the re-sharing of information that one has received themselves.

8.2.5 Anonymity or Pseudonymity of Information Transfer Participants

In many situations information transfer may be possible under anonymity or pseudonymity, perceived or factual. One can distinguish different types and layers of anonymity, and from an end-user's perspective much of this type of assessment is likely contextual.[396] Within many of the Internet's (human-readable) information exchange fora (e.g. social networks in all permutations, messenger services, video games, etc.), individuals can usually conceal their identity to a certain degree, and may have varying reasons to do so, including their preference for privacy.[397] Many social networks deliberately accept or tolerate individual's acting under false names or handles, or will not enforce strictly their terms of service mandating that their

---

other competing information-"nuggets" that are weighed by the algorithm when considering how to construct the visual interface. Even if the information is de-ranked in a way that only a small amount of people actually see it, this still means that there was a negligible cost to reach the few people it did actually reach, still making the barriers to impart information quite low *compared to a pre-information age*. For some analysis and context (albeit potentially with limited quality as an academic source) about e.g. Facebook's / Meta's approach to organic and inorganic growth see e.g. Scott W H Young and others, 'The Social Media Toll Road: The Promise and Peril of Facebook Advertising', *College \& Research Libraries News*, 75.8 (2014), 427–34; Marshall Manson, 'Facebook Zero: Considering Life after the Demise of Organic Reach', *Social@ Ogilvy White Paper*, 2014, 1–7.

[395] Of course, the costs of sending e-mails to a group of a huge group of people compared to a single individual are still different: the underlying connectivity infrastructure must still be set up accordingly (e.g. more storage space in more e-mail accounts, more internet-cables reaching to more households), but the costs are both smaller compared to the pre-digital alternatives, and they are typically not priced in immediately at the point of imparting information. It is usually free to send an e-mail to one or multiple persons, the incremental, structure-related costs are born then by the community of digital users, individuals and larger entities alike.

[396] See for a similar distinction between "technical" and "social" anonymity STEPHEN C. HAYNE and RONALD E. RICE, 'Attribution Accuracy When Using Anonymity in Group Support Systems', *International Journal of Human-Computer Studies*, 47.3 (1997), 429–52 <https://doi.org/10.1006/ijhc.1997.0134>.

[397] Ruogu Kang, Stephanie Brown, and Sara Kiesler, 'Why Do People Seek Anonymity on the Internet?', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM, 2013), pp. 2657–66 <https://doi.org/10.1145/2470654.2481368>; Kimberly M. Christopherson, 'The Positive and Negative Implications of Anonymity in Internet Social Interactions: "On the Internet, Nobody Knows You're a Dog"', *Computers in Human Behaviour*, 23.6 (2007), 3038–56 <https://doi.org/10.1016/j.chb.2006.09.001>.

user's remain identifiable. Similarly, using messenger services often requires solely a phone number, and e-mail addresses do not have to display the human behind the service. This does not mean that these digital avatars are truly anonymous in a technical sense; an individual behind an obscure e-mail address may be identified given enough effort, access or technical tools. But this type of uncloaking is unrealistic in the context of many short-lived interactions between individuals within the digital domain.[398] This also has implications for how the medium of information technology is used, interacted in and perceived by other users. For example, imparting wrong information comes with the risk of a negative reaction by the recipient when the lack of veracity is discovered. If nothing else, a loss of reputation concerning the truthfulness of the information imparted by the respective individual may follow, with more severe consequences being conceivable [399]. Anonymity or pseudonymity can shield from consequences or the perception thereof (lowering a non-monetary cost factor of sorts) and may lead to a more "loose" assessment of the veracity of information. This perceived loss of individuality, as a result of anonymizing environments, is often termed deindividuation.[400]

This type of "realistic" anonymity of an information source can be extended beyond the archetypical human Internet user. Just as it is difficult to identify a specific individual as the information source behind a digital avatar, it may be difficult for an information recipient to trace specific information to an informational agent as well. At the same time as contextual anonymity vis-à-vis other human users is maintained, anonymity vis-à-vis informational agents such as profiling processes is often yielded. As individuals leave behind them rich data trails, they may be (sufficiently) identified through some informational agent, even though they were sufficiently cloaked from any type of human intuition.[401]

---

[398] For example, two users of a social network, using fake names, arguing in a comment-section of an incendiary posting may never interact again after their heated exchange, and neither will likely have access to (or perhaps motivation to use) decloaking knowledge or tools to establish their contrarian's identity. A recipient of an unsolicited (spam)-email is unlikely to ascertain the true identity of the sender (insofar that is even possible and the sender was another human individual).

[399] Beata Arcimowicz, Katarzyna Cantarero, and Emilia Soroko, 'Motivation and Consequences of Lying. A Qualitative Analysis of Everyday Lying', *Forum: Qualitative Social Research*, 16.3 (2015) <https://doi.org/https://doi.org/10.17169/fqs-16.3.2311>.

[400] S. D. Reicher, R. Spears, and T. Postmes, 'A Social Identity Model of Deindividuation Phenomena', *European Review of Social Psychology*, 6.1 (1995), 161–98 <https://doi.org/10.1080/14792779443000049>; Paul Benjamin Lowry and others, 'Why Do Adults Engage in Cyberbullying on Social Media? An Integration of Online Disinhibition and Deindividuation Effects with the Social Structure and Social Learning Model', *Information Systems Research*, 27.4 (2016), 962–86 <https://doi.org/10.1287/isre.2016.0671>; Serena Coppolino Perfumi and others, 'Deindividuation Effects on Normative and Informational Social Influence within Computer-Mediated-Communication', *Computers in Human Behaviour*, 92 (2019), 230–37 <https://doi.org/10.1016/j.chb.2018.11.017>.

[401] See for a few example strategies (without claim of exhaustiveness) Mudhakar Srivatsa and Mike Hicks, 'Deanonymizing Mobility Traces', in *Proceedings of the 2012 ACM Conference on Computer and Communications Security - CCS '12* (New York, New York, USA: ACM Press, 2012), p. 628 <https://doi.org/10.1145/2382196.2382262>; Xuan Ding and others, 'A Brief Survey on De-Anonymization Attacks in Online Social Networks', in *2010 International Conference on Computational Aspects of Social Networks* (IEEE, 2010), pp. 611–15 <https://doi.org/10.1109/CASoN.2010.139>; Huaxin Li and others, 'Privacy Leakage via De-Anonymization and Aggregation in Heterogeneous Social Networks', *IEEE Transactions on Dependable and Secure Computing*, 17.2 (2020), 350–62 <https://doi.org/10.1109/TDSC.2017.2754249>; Jianwei Qian and others, 'Social Network De-Anonymization and Privacy Inference with Knowledge Graph Model', *IEEE Transactions on Dependable and Secure Computing*, 16.4 (2019), 679–92 <https://doi.org/10.1109/TDSC.2017.2697854>; Shouling Ji, Prateek Mittal, and Raheem Beyah, 'Graph Data Anonymization, De-Anonymization Attacks, and De-Anonymizability Quantification: A Survey', *IEEE Communications Surveys & Tutorials*, 19.2 (2017), 1305–26 <https://doi.org/10.1109/COMST.2016.2633620>; Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez

8.2.6 Personalization, Fragmentation and Erosion of Privacy

Pre-Internet, most multiadic information transfer was not personalized towards the individual recipients. Now, information transfer is often highly personalized, i.e. information is customized, and individualized by matching certain categories of information to an individual's profile based on their stated and implied preferences. [402] One of the most illustrative examples of personalization within the digital domain is advertising. Some research suggests that higher (perceived) personalization is not necessarily the most effective in steering towards a desired behaviour (e.g. purchases or engagement rate) [403], and individuals may exhibit reactance when low-trust actors deploy high personalization. [404] From a point of stipulative definition however, this may just be seen as improper personalization, as low-trust actors' optimized and hence most personalized information transfer may be one of insufficient depth of personalization. Content personalization is not restricted to direct monetarization efforts but can also affect the display of information pertaining to the essential core of an information service. To this end, newspaper websites may tailor their content based on reader profiles in a similar way.[405]

Similarly, many social networks have moved away from a purely chronological display of content to a more personalized feed to optimize for an engagement metric.[406] This phenomenon is aided by a fragmentation of information, particular on participative channels of information distribution such as social networks [407]. As "nuggets" of information become smaller and more numerous, personalization algorithms have more opportunity and material to personalize information display accordingly. Unfortunately, this type to fragmentation (and interjection or introduction of non-relevant information) seems to carry negative implications for information comprehension by the affected individuals,[408] likely increasing the susceptibility to misinformation as cognitive load of the information recipient increases.

Finally, information technology can also personalize the legal imposition on the individual directly, e.g. by changing not just how information is presented but also what conditions the

---

del Prado Cortez, 'De-Anonymization Attack on Geolocated Data', *Journal of Computer and System Sciences*, 80.8 (2014), 1597–1614 <https://doi.org/10.1016/j.jcss.2014.04.024>; Yingxia Shao and others, 'Fast De-Anonymization of Social Networks with Structural Information', *Data Science and Engineering*, 4.1 (2019), 76–92 <https://doi.org/10.1007/s41019-019-0086-8>; Adriano Di Luzio, Alessandro Mei, and Julinda Stefa, 'Mind Your Probes: De-Anonymization of Large Crowds through Smartphone WiFi Probe Requests', in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications* (IEEE, 2016), pp. 1–9 <https://doi.org/10.1109/INFOCOM.2016.7524459>.

[402] Alan L. Montgomery and Michael D. Smith, 'Prospects for Personalization on the Internet', *Journal of Interactive Marketing*, 23.2 (2009), 130–37 <https://doi.org/10.1016/j.intmar.2009.02.001>.

[403] Alexander Bleier and Maik Eisenbeiss, 'Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where', *Marketing Science*, 34.5 (2015), 669–88 <https://doi.org/10.1287/mksc.2015.0930>.

[404] Alexander Bleier and Maik Eisenbeiss, 'The Importance of Trust for Personalized Online Advertising', *Journal of Retailing*, 91.3 (2015), 390–409 <https://doi.org/10.1016/j.jretai.2015.04.001>.

[405] Neil Thurman and Steve Schifferes, 'The Future of Personalization at News Websites', *Journalism Studies*, 13.5–6 (2012), 775–90 <https://doi.org/10.1080/1461670X.2012.664341>; Vadim Lavrusik, 'Washington Post Announces Personalized News Aggregation Site', *Mashable*, 2011.

[406] Robert Bodle, 'Predictive Algorithms and Personalization Services on Social Network Sites: Implications for Users and Society', in *The Ubiquitous Internet* (London/New York: Routledge, 2014).

[407] Tonglin Jiang, Yubo Hou, and Qi Wang, 'Does Micro-Blogging Make Us "Shallow"? Sharing Information Online Interferes with Information Comprehension', *Computers in Human Behaviour*, 59 (2016), 210–14 <https://doi.org/10.1016/j.chb.2016.02.008>.

[408] Richard E. Mayer and others, 'Increased Interestingness of Extraneous Details in a Multimedia Science Presentation Leads to Decreased Learning.', *Journal of Experimental Psychology: Applied*, 14.4 (2008), 329–39 <https://doi.org/10.1037/a0013835>; David R. Karger and William Jones, 'Data Unification in Personal Information Management', *Communications of the ACM*, 49.1 (2006), 77–82 <https://doi.org/10.1145/1107458.1107496>.

individual may act under. For example, dynamic pricing algorithms may set a factual default for each individual, determining their action potential and legal obligations should individuals engage in a purchase. Even more direct, legal contract rules (or their defaults) that attach to a certain transaction may be personalized by informational agents.[409]

Again, the increase in capability and use cases of informational agents, and capability of interconnective infrastructure together with improvements in data analysis suggests that individuals leave behind "data trails" that are both more comprehensive in a quantitative sense as well as more informative generally. This likely creates an economic incentive to maximize access to such information [410], which in turn may conflict with the individual's privacy. [411] On the other hand, existing notions of privacy protection, e.g. as enacted through legislation, may only provide limited reprieve as technology also "outsmarts" the legal boundaries in some cases. For example, privacy may still be lost in practice if data is collected or assigned to an individual merely on the basis of group observations which are still sufficiently accurate.[412] Of course, privacy is generally seen as inextricably linked or even constitutive of autonomy,[413] tying this domain characteristic close to the object of inquiry in this thesis.

8.2.7 Easy Implementation of Scalable, Non-Human Actors

The domain characteristic of information intermediation demands that information transferred via the Internet must be translated into a human-readable format via an interfacing device running some sort of intermediary agent, which gives the agent the opportunity to exert influence on its user. But algorithmic intermediary agents also work farther removed from these Internet endpoints. Many parts of information transfer processes that require somewhat complex actions benefit from the use of advanced intermediary agents, in particular on the level of aggregation and distribution. For example, aforementioned personalization of information, e.g. through a sort-of recommender system, requires the parsing of increasingly large amounts of datasets, the creation of profiles of the information recipient and the matching of information and these profiles.[414] These algorithms are becoming more sophisticated and their underlying artefacts are becoming more potent. These kind of intermediary agents have distinct advantages over potential human competitors that could be used for the same tasks. In situations in which their application is feasible, a number which is likely to increase, they are orders of magnitude faster and they are easily portable and scalable. Installing a new intermediary actor on an informational pipeline, under the condition of the task being appropriate for the underlying

[409] See e.g. Ariel Porat and Lior Jacob Strahilevitz, 'Personalizing Default Rules and Disclosure with Big Data', *Mich. L. Rev.*, 112 (2013), 1417. This combines the notions of default-setting (as an example of choice architecture) and immediate legal consequences.

[410] Shoshana Zuboff, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization', *Journal of Information Technology*, 30.1 (2015), 75–89 <https://doi.org/10.1057/jit.2015.5>.

[411] Dan Cvrcek and others, 'A Study on the Value of Location Privacy', in *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society - WPES '06* (New York, New York, USA: ACM Press, 2006), p. 109 <https://doi.org/10.1145/1179601.1179621>; Anne S.Y. Cheung, 'Location Privacy: The Challenges of Mobile Service Devices', *Computer Law & Security Review*, 30.1 (2014), 41–54 <https://doi.org/10.1016/j.clsr.2013.11.005>; Stephen B. Wicker, 'The Loss of Location Privacy in the Cellular Age', *Communications of the ACM*, 55.8 (2012), 60–68 <https://doi.org/10.1145/2240236.2240255>.

[412] As pointed out by already by Anton Vedder, 'KDD: The Challenge to Individualism', *Ethics and Information Technology1*, 1 (1999), 275–81 <https://doi.org/https://doi.org/10.1023/A:1010016102284>. See also Section 3.4 for this discussion.

[413] Benn; Kupfer; Andreas Tsamados and others, 'The Ethics of Algorithms: Key Problems and Solutions', *AI & SOCIETY*, 2021 <https://doi.org/10.1007/s00146-021-01154-8>.

[414] Paul Resnick and Hal R. Varian, 'Recommender Systems', *Communications of the ACM*, 40.3 (1997), 56–58 <https://doi.org/10.1145/245108.245121>.

technology may be done with fewer resources than placing a new human actor into the same place. This holds true in particular in cases where there is little adjustment that has to be made to existing algorithms that can be copied and ported onto a new informational pipeline. Indeed, effective and commercially available agents may *envelope* future informational pipelines, that is effecting them to be constructed so that they may be ported onto them easily due to the (economic) interests of the pipeline stakeholders.

As stated before, these agents are not necessarily autonomy neutral. The use of these intermediate informational agents does not necessitate that autonomy constraints will flourish. However, informational agents, often serve to maximize metrics (such as engagement rate of the information recipients in online open media) that are not related to the individual autonomy (e.g. their access to veracious information) per se. [415]

8.2.8 Interim Conclusion

In answering the question of what the most important characteristics of the IoE are, that are potentially relevant for understanding the risk of autonomy constraints, this section has outlined two sets of factors that are of particular relevance. On the one hand, the underlying domain and all autonomy constraints emanating from it are directly affected by the capabilities and prevalence of the devices forming it and the infrastructure connecting said devices. On the other hand, we can derive particular vectors of influence, i.e. fulcrums which facilitate autonomy constraints, which serve as meta characteristics, i.e. characteristics that are best attributed to the domain itself. Here, the thesis has identified seven vectors: the (mandatory) intermediation of information, immediacy of information transfer, low barriers to impart information, plurality of information recipients, anonymity or pseudonymity of information transfer participants, personalization, fragmentation and erosion of privacy and easy implementation of scalable, non-human actors. Each of these enable or facilitate constraints of an individual's autonomy on the three dimensions discussed previously. This fits into the previous discussion: assuming that technology is not autonomy-neutral, the above domain characteristics highlight, that exposure to autonomy-undermining (and supporting) phenomena is increasing as informational agents increase in both capability and number. Given easier propagation of information, extent of reach and often personalized delivery to human recipients, and given its autonomy-undermining qualities, as will be explained below), the risk to an individual's autonomy is increasing as a result of the shift of information domain characteristics. However, due to the fundamental nature of these vectors, while they are undoubtedly causal for autonomy constraints, attribution of them to specific agencies (as opposed to classifying them as "environmental factors") may be difficult still.[416] The analysis thus moves now to vectors of influence that are more easily attributable.

[415] Ozan Candogan and Kimon Drakopoulos, 'Optimal Signaling of Content Accuracy: Engagement vs. Misinformation', *Operations Research*, 2020, opre.2019.1897 <https://doi.org/10.1287/opre.2019.1897>; Sebastián Valenzuela and others, 'The Paradox of Participation Versus Misinformation: Social Media, Political Engagement, and the Spread of Misinformation', *Digital Journalism*, 7.6 (2019), 802–23 <https://doi.org/10.1080/21670811.2019.1623701>.

[416] This is discussed in greater detail in Section 11.3.2.1.

9 Agents and Agent Characteristics

This thesis investigates the exercise of, and the constraints upon individual autonomy within the domain of the IoE. After previous analysis and contextualization of the domain of interest and the concept of autonomy, we may now approach the technical counterparts of human subjects, users and information recipients. The second, and arguably most important, set of vectors of influence connects to processes on the informational pipeline that are not mainly connected to the underlying domain or the sphere of the information recipient, but lie in between them. This section aims to address research sub-question 2.3, which asks about the characteristics of these processes relevant for understanding the risk of autonomy constraints. The object of interest in this chapters are thus informational agents and their characteristics.

The analysis will proceed as follows. Section 9.1 will elaborate on the concept of agency and its holders and connects this notion to potential qualified processes within the IoE-domain. Section 9.2 then provides a definition for a specific type of agent, the informational agent, which serves as the attachment point and object of reference for the vectors of influence of interest in this chapter. Based on this use, the analysis crystallizes the necessary prerequisites for a process to be considered as an informational agent, in particular agency, compactness, complexity and the capacity to impart persuasion, discusses their implications and compares them to different approaches to define agents. On the basis of this, Section 9.3 then outlines the vectors of influence attaching to informational agents, which are primarily the (relative) immediacy of information processing, selective superior reasoning capabilities, opacity of algorithmic decision-making, data reliance and susceptibility to bias, their complex status between mistrust and deference, their potential to facilitate and amplify information transfer and their placement, function and effectuation obscurity. Section 9.4 offers a brief conclusion.

9.1 Agents and Agency

Within this text, the term *informational agent* or *agent of the IoE* is used to denote this counterpart-concept, but previous mentions of the term have relied upon context instead of explicit definitions. This section aims to rectify this previously introduced opacity.

9.1.1 Agents as Entities with Agency

In its most fundamental form, an agent is an entity that has the capacity to *act*, or in other words has (the capacity for) *agency*.[417] The concept of agency is not dissimilar but distinct from the concept of autonomy. While autonomy describes self-government under fulfilment of specific requirements (e.g. coherence and integrity of underlying mental states, sufficient information, etc.), the concept of agency is more specific and deals solely with actions.[418] Contemporary theoretical scholarship has developed different theories as to what requirements such entity ought to meet. According to what is sometimes dubbed the *causal* or *standard theory* of agency,

[417] See for this and some of the following conceptual taxonomy (including later references) Markus Schlosser, 'Agency', ed. by Edward N Zalta, *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University, 2019).

[418] In particular, individuals may be non-autonomous with respect to specific actions, but as they still *act*, they still possess agency. Consider e.g. the notion of an individual that is addicted to consumption of social media: While their autonomy is clearly constrained (in the meaning of this thesis the constraints are of relational and intrinsic nature), their capacity to act is not; the consumption itself is evidence of agency.

an entity can exercise (intentional) agency (and may be thus considered an agent) if it has the capacity to act intentionally, and agency is exercised through intentional actions or at least actions that derive from intentional actions.[419] While some theorists understand this definition to limit (intentional) agency to human actors (as it may require underlying mental states of intention), others take a wider stance,[420] ascribing pseudo-mental states to non-human actors[421] or not requiring mental states at all.[422] However, intentional agency is not required to recognize entities as agents within this inquiry, and even within standard agency theory, different (non-intentional) types of agencies are accepted, which includes artificial agency.[423] And indeed, the use of the term *agent* within the domain of computer science is well established, if not necessarily well-defined. [424] This underlines, that the term *agent* is at least not inappropriate to describe the autonomy-relevant phenomena within this context.

9.1.2 From Algorithms to Artificial Intelligence

Returning back to the concept of an agent, i.e. an entity that acts with agency, we can use the concept of the informational pipeline to inform a possible delimitation. Clearly, of interest here are informational agents, that is entities that act with agency within or as a function of the domain of the IoE. We may therefore narrow the scope of what ought to be an agent to entities (or more precisely processes) that are part of the informational pipeline as it intersects with the IoE.[425] The question then remains what scope, complexity and impact these digital processes ought to have to be considered agents. Naturally, many complex processes can be atomized into sub-processes themselves, begging the question on which level of abstraction to attach the agent-denominator.

The main building block within this domain are algorithms.[426] An algorithm is a type of instruction or process, divided into separate steps that is modelled around and applied to a specific situation or problem to be solved.[427] Connected to this, and usually confusingly also denotated as an algorithm is the digital realization of this type of instruction, i.e. the computer program or similar that applies the information within the algorithmic concept. To distinguish these two, we may also call the latter a computer program. Algorithms (of both kind) can be of varying complexity and dependencies, requiring different types and amounts of inputs, and progress towards similar or different outputs or results at potentially different speeds. Algorithms can be dependent on other algorithms and incorporate them, with both the overall algorithm and its nested sub-algorithm qualifying as an algorithm by themselves.

---

[419] Markus Schlosser. See also for further discussion Markus E. Schlosser, 'Agency, Ownership, and the Standard Theory', in *New Waves in Philosophy of Action* (London: Palgrave Macmillan UK, 2011), pp. 13–31 <https://doi.org/10.1057/9780230304253_2>. Of course, the definition of *intentional* is also a matter of no consensus.

[420] Schlosser calls these postures *instrumentalist* or *realist* stances, see Markus Schlosser.

[421] Danniel Dennett, *The Intentional Stance, Cambridge* (Cambridge: MIT Press, 1987), chap. 2.

[422] Xabier E. Barandiaran, Ezequiel Di Paolo, and Marieke Rohde, 'Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action', *Adaptive Behaviour*, 17.5 (2009), 367–86 <https://doi.org/10.1177/1059712309343819>.

[423] Of course, this is not true within all theories and exceptions exist.

[424] As observed by Krzysztof Cetnarowicz, 'From Algorithm to Agent', 2009, pp. 825–34 <https://doi.org/10.1007/978-3-642-01973-9_92>.

[425] This distinction is necessary, as the informational pipeline encompasses non-digital environments as well, as mentioned above.

[426] Of course, there are more atomistic elements within the digital domain. For example, we may consider data or information itself or the concept of (low-level) programming languages, the physical structure of transistors, etc. as building blocks of the digital domain. However, it is with algorithms that we can start ascribing them (from the outside) a status that may somewhat resemble or fulfil the requirements of agency.

[427] 'Algorithm', *Merriam-Webster Dictionary*, 2022.

Higher in abstraction and on the opposite side of the spectrum of complexity and impact than algorithms (and programs) is the concept of artificial intelligence (AI). Artificial intelligence describes a technological entity (comprised of algorithms and the like) which is thought as a whole to accede to some (higher) standards of behaviour, reasoning or compliance with rationality. The exact scope of the term is extremely contentious, with some theorists distinguishing between weak and strong AI based on their abilities and complexity, of which only weak AI have been achieved so far.[428] Between these limits on the plane of complexity and impact lie the entities that ought to be grasped by the term agent.

9.1.3 Necessity of Intrinsic Qualities

However, it is not a necessary conclusion that the complexity and ability that are intrinsic to an entity should be a deciding factor for its denomination as agent. As mentioned above, theoretical scholarship dealing with the issue of agency sometimes disagrees, if intrinsic requirements ought to be fulfilled to describe the exercise of the same. This is highlighted in an assessment procedure commonly known as the *Turing-Test*. Suggested by Alan Turing in 1950, it describes a process dubbed the *imitation game* in which a computer process and a human individual are observed and compared to each other. Both attempt to convince a (human) interrogator through textual representation of language that they are human, and the computer process *passes* this test if the interrogator cannot distinguish between the two based on the written responses to their questions.[429] With this procedure, Turing suggested to divide computer processes on the basis of their ability to *think*. However, it is immediately obvious that the imitation game does not directly rely on any intrinsic feature of the computer process, but instead assesses the outcome. In other words, the imitation game is rather agnostic about the internal information processing capabilities of the computer process. Rather, its output (and hence impact) on the outside, non-digital world is considered of importance. Critical then is seemingly not if computers really do *think* but if they *appear to think*. The same approach can be applied to the issue of agency. The refusal of recognizing the agency of entity based on its internal characteristic may be in line with how we understand human agency on the basis of concepts such as rationality, free will or the existence and prevalence of certain mental states. However, this distinction may be academic in nature if the entity is clearly perceived as an agent by interactors or observers that are not fully aware of these intrinsic qualities.

9.2 Defining Informational Agents

In accordance with the above considerations, this thesis adopts a pragmatic conception of informational agent (*as is relevant for autonomy*) in the digital sphere.[430] Hence, within this thesis an informational agent (or just agent) is

- a non-human process within the informational pipeline that intersects with the IoE, that
- is sufficiently compact and delimited from an *outside* point of view, as well as
- may be reasonably ascribed agency from an *outside* point of view, i.e. may generally be perceived to have

[428] See Stuart Russel and Peter Norvig, *Artificial Intelligence: A Moden Approach* (Pearson, 2020), p. 31f; Pei Wang, 'On Defining Artificial Intelligence', *Journal of Artificial General Intelligence*, 10.2 (2019), 1–37 <https://doi.org/10.2478/jagi-2019-0002> with many further references. The penultimate representation of such an entity is sometimes called general artificial intelligence or singularity.
[429] A M Turing, 'Computing Machinery and Intelligence', *Mind*, 59.236 (1950), 433–60.
[430] For consideration of and comparison with existing agent conceptions see below under Section 9.2.7

        - (self-imposed or otherwise-derived) intent or targets,
        - sufficient deliberative (i.e. non-trivial information processing-) capabilities, and
        - sufficient potential to shape and effect information based on intent and deliberation,
    - is sufficiently complex from an *outside* point of view, and
    - has the potential to impart persuasive qualities onto information,
        - i.e. has the potential to constrain or support an individual's autonomy as a result of either direct interaction or other handling of the information flow along the informational pipeline.

Let us consider these characteristics in turn.

9.2.1 Informational Pipeline-Bound Non-Human Process

The first characteristic is derived from the previously discussed concept of the informational pipeline. As their name suggests, informational agents deal with information. Typically, this means that information is received, processed and transferred on to some effect for some purpose. As a result, they must be part of the informational pipeline. This characteristic serves as the coarsest filter to divide between agents generally and informational agents specifically, of which only the latter are of interest here. However, as mentioned previously, the informational pipeline is not purely digital (or technological) in nature. Indeed, human behaviour may very well be a process of information transmission, and as such they too are eligible agents on the informational pipeline. For example, a human operator may enter data into a computer, train an artificial intelligence or interpret the outcome of an algorithm, before this information is further transferred and processed. Strictly speaking then, we may distinguish between human and non-human informational agents. However, this diction is unnecessarily cumbersome, and the main scope of research of this text are non-human informational agents. We therefore deploy a non-human qualifier, meaning that the term informational agent (where not specified otherwise) refers to non-human processes.

The fact that an informational agent must be part of the informational pipeline implies another, hidden, characteristic. Such an agent cannot be insularly isolated from all environments, and it must interact with its environment (in particular the processes or entities before and after it). A process that cannot impart change onto its environment, cannot really be said to have the capacity to act, and hence lacks agency. For example, a highly advanced algorithm for the purposes of timekeeping (not connected to any other processes) is clearly not an agent (but may be part of an agent). Similarly, porting an algorithm that processes visual inputs to steer a movable robot in pursuit of some predetermined or dynamic goal (and as a result perhaps an agent) onto a device that has neither visual sensing nor movable parts clearly does not result in an agent.

9.2.2 Perceived Compactness

Within the informational pipeline, different processes may be interwoven and dependent on each other. For example, the information processing in the context of an individual receiving a bank loan may consist of information input by the applicant, data correction or cleansing through human or automated means, reformatting of the information, processing by an

advanced algorithm (e.g. a neural network) to determine likeliness of default, [431] comparison of the resulting risk factor with other data-points and internal policy, consideration of case-by-case peculiarities and concluding in the affirmative or negative depending on this process. Assuming these processes meet our other requirements, which combination of these processes are informational agents, and which are merely sub-processes of agents?

The interna of informational agents is rarely fully (if at all) accessible for observers or individuals interacting with them. Information processing tends to happen in the background, with the individuals affected only receiving information at certain points in the process. Continuing the above example, the applicant likely is partial to information only when entering their loan-related information and when receiving the outcome. Hence, from the viewpoint of the applicant, they are interacting with a black box that processes data for the purposes of securing a loan, and not with a bundle of processes that are in nuanced interplay. Assuming that strictly speaking only one part of this black box meets our requirements of agency (perhaps the final decision-making algorithm weighing all the information), we may be inclined to specify that only this algorithm is the agent in question. But obtaining knowledge about these interna is often unrealistic and implausible, not just for researchers but in particular for lay-persons interacting with the agent. It seems sensible then to use the distinguishable black box as the concept to which the term agent attaches. Insofar an individual can perceive limits of an information process to a sufficient degree, these are the limits of what the agent consists of. In other words, observers perceive the agent as sufficiently *compact*, and distinguishable from other processes in the informational pipeline.

This approach necessitates that we recognize that the perceived compactness (and as a result the status of agent) is somewhat subjective. Different human observers may have different insights into information processes. The black box these individuals perceive is therefore different as a result. The software-engineer that has conceived and implemented the above loan-assessment algorithm clearly has deeper insight into the entire process than a typical bank customer, and may perceive the compactness of the algorithm to be narrower (even though his interaction with it may be similar). Here two considerations may be of help. First, even with uncharacteristic, specialized knowledge, separating processes in between which interactions do not occur seems unnecessary. From the point of the software engineer, their special knowledge about the make-up of the risk-assessment processes does not change the fact that between application and receiving their decision no input can be made. We should understand perceived compactness then not just on the basis of knowledge but also on the basis of interaction. Second, it is worth keeping in mind that a combination of processes that qualify as agents (in some context) can make up another agent. In other words, a multi-agent system may be considered as an agent under a different context (such as from the viewpoint of an individual with different perceptions). As an example we can consider a risk-assessment algorithm in which the algorithm, after processing the individual trove of data, gives an intermediate result that can be accessed by the applicant, and can be fed with additional data, both of which may be done through some sort of technical process (e.g. a specific web-interface). However, if the applicant does not have this knowledge (e.g. they have not read the supporting documentation when agreeing on the risk assessment procedure), the algorithm proceeds with default

[431] See e.g. Ming-Chun Tsai and others, 'The Consumer Loan Default Predicting Model – An Application of DEA–DA and Neural Network', *Expert Systems with Applications*, 36.9 (2009), 11682–90 <https://doi.org/10.1016/j.eswa.2009.03.009>; Herbert L. Jensen, 'Using Neural Networks for Credit Scoring', *Managerial Finance*, 18.6 (1992), 15–26 <https://doi.org/10.1108/eb013696>; Marion O Adebiyi and others, 'Secured Loan Prediction Systens Using Artificial Neural Network', *Journal of Engineering Science and Technology*, 17.2 (2022), 854–73.

information. Clearly, the informed applicant has before them two black box processes, divided by an intermissive opportunity to interact. Depending on how the purpose and processing is divided between these black boxes (to the knowledge of the applicant), the informed applicant may perceive these two processes as separate, compact agents. At the same time, the uninformed applicant will not perceive any difference between these two processes and consider them joined as compact. The same would be functionally true for an informed applicant who has lost the ability to access the intermediate result and provide additional information (e.g. because they have lost the password necessary to access the web interface).

In summary, the compactness of an agent describes its delimitation and scope as distinguishing factors of other processes within the same informational pipeline. When it comes to perception of this compactness, we should consider not only knowledge about the structure of the agent, but also the nature of meaningful interaction.

9.2.3 Ascribable Agency

Informational processes can take myriad forms and complexities. In order to be ascribed agency to them, observers should be able to derive three characteristics.

First, the agent must have some sort of intent, goal or target.[432] This intent will typically be inherited. In this case, the intent is provided by something external to the agent such as the maximization of a certain metric. For example, an algorithm structuring the feed of a social media website may have the goal to maximize a metric that tries to assess user engagement. A recommender system deployed on an e-commerce site, suggesting other wares to purchase, may try to maximize the conversion rate of these suggestions to actual purchases. Note that these goals are typically proxies for the actual target,[433] which can lead to deviations of the agent from truly intended behaviour (and corresponding ethical challenges).[434] But of course, agents can also set their own goals to a certain extent. Agents may rotate between different inherited goals based on some criteria (e.g. environmental input), or determine their own goals if they have the capacity to do so. In many instances of using machine learning, the inheritance of intent is muddled. If the deployer of a technology uses an algorithm (i.e. the machine-learning algorithm) to determine the best proxy metrics for another algorithm (e.g. the aforementioned recommender system), the high-level intent may still be user engagement, but the more concrete goals of the recommender system and its processes may be more granular or arcane. This also means that the deployer may have limited knowledge over the proxy metrics that the machine learning algorithm as deemed to be most appropriate, so the link between the algorithmic intent and the deployers (human) intent becomes tenuous. This has implications for ascribing moral responsibility.[435]

[432] Compare this to the requirement of intent for persuasive technology in Section […].
[433] See for a good overview Rachel Thomas and David Uminsky, 'The Problem with Metrics Is a Fundamental Problem for Ai', *ArXiv Preprint ArXiv:2002.08512*, 2020.
[434] See e.g. Anya E R Prince and Daniel Schwarcz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data', *Iowa L. Rev.*, 105 (2019), 1257; Richard Warner and Robert H Sloan, 'Making Artificial Intelligence Transparent: Fairness and the Problem of Proxy Variables', *Criminal Justice Ethics*, 40.1 (2021), 23–39. However, even if the proxies are inaccurate for the true goal of the deployer of the agent, the intent of the agent is still actually inherited, as the metric was chosen by the deployer nonetheless.
[435] Mittelstadt and others.

Second, the agent must have sufficient deliberative capabilities to process information. Deliberation requires the agent to have access to its environment[436] and the means to analyse the environment generally and in light of its goal.[437] In other words, the agent must encompass a representation of its environment that synthesizes information into a reduced form which is accessible to its reasoning process. This process in turn must be sufficient to take on this representation and through some computation determine which actions to take. The bar which this level of deliberation has to meet is difficult to determine.

Third, the agent must have sufficient potential to shape and effect information based on its intent and deliberation. The concept that agents cannot be insular was already highlighted in Section 9.2.1, where we found that isolated entities without input or output potential cannot be considered agents. But this requirement goes beyond mere passive or receptive embedment into an environment. A process that is not isolated but merely forwards information does not have any potential to change its environment beyond stopping to work as intended.

All of these requirements must be fulfilled at the same time. For example, an entity with the purpose of recommending movies to its users with sophisticated profiling capabilities that, due to a software bug, only recommends the same movies over and over, lacks the impact of its deliberation onto its actions, and is thus not an agent. Similarly, the entity with the same purpose that lacks any type of deliberation processes and merely selects a movie randomly out of some list is not an agent for lack of these deliberative capacities and likely also for lack of any graspable intent.

9.2.4 Sufficient Complexity

The first two requirements cast a very wide net. Artificial Intelligence, complex recommender systems and sophisticated profiling algorithms are clearly encompassed by them, but so are simple if-then algorithms weather forecasting software and calculators. This is intuitively unsatisfying. Clearly, the term agent should describe an entity of some sufficient complexity. But what does that notion encompass? One aspect of complexity, that is complexity of the agent's deliberative potential, is subsumed by the concept of agency, that was highlighted above. But other complexity remains. In particular, agents must be perceived to display some extent of unpredictability or fuzziness. This is the other side of the coin of the previous requirement of compactness. Whereas compactness is a procedural characteristic, complexity is its substantive counterpart. Insofar interacting users can reliably and with utmost accuracy predict the outcome of an agent's behaviour as well as the minutiae of actions the agents itself, its agency clearly collapses into a deterministic process from the viewpoint of the observer. In practice, this means that the process of making agents more explainable also threatens their status as agents under this definition.

9.2.5 Imparting of Persuasion

Agents can process information towards many different outcomes. To be of interest for this inquiry, that is for the analysis of how informational agents can constrain an individual's autonomy, they must have the potential to indeed affect an individual's autonomy. However, this is not a very selective criterion. Considering the elaborations in Section 5 and the examples

[436] An agents access to its environment is technically already part of its status as a part of the informational pipeline.
[437] See e.g. Michael Wooldridge and Nicholas R Jennings, 'Intelligent Agents: Theory and Practice', *The Knowledge Engineering Review*, 10.2 (1995), 115–52.

on how autonomy can be constrained in a multitude of fashions (and on the three dimensions outlined), most processes of the informational pipeline will accede to this requirement. Similarly, the excursus into the domain of persuasive technology in Section 6 has also highlighted how omnipresent the use of technology is that has persuasive qualities. It is noteworthy again that in order to persuade, agents are not required to be situated at the end of the informational pipeline where they interact directly with a user. Instead, they can impart some persuasive quality to the information that is then further transported by other processes which then interact with a user. For example, consider persuasion conducted by an agent that creates profiles of online users of a website, observes their online behaviour and proactively selects advertising content that uses that individual's cognitive bias to its fullest. This highly effective advertising material is then forwarded to a human operator who prints the material out and sends them to the individual to be persuaded per physical mail. Clearly, the agent has conducted the persuasion even if it has not interacted with the ultimate information recipient. Ultimately, the agents that are considered in this thesis will be persuasive agents in the meaning of the above sections.

9.2.6 Pseudo-Agents or Agents Proper?

Insofar a process fulfils the requirements outlined in the previous text, we may call it an agent. But of course there are epistemological challenges when it comes to advanced technologies. First, as complexities of these processes increases, ascertaining that these requirements are fulfilled or not becomes more difficult. Second, some of these (in particular the requirement of compactness, complexity and even the distinguishing between these two) seem to be relative from an observer's position. The very nature of agent's status as a black box means that if that box is truly impervious to outside observation, we cannot ascertain that the processes within this box actually fulfil all of the requirements. To distinguish processes that truly fulfil all of these requirements and agents that seem to fulfil these requirements we can denote the first group as agents proper and the second group as pseudo agents. This creates conceptual clarity, but at the cost of more cumbersome notation and little gain. Instead, this text subsumes both agents proper and pseudo-agents into the same category, and denotes them both as agents. As the set of processes that would qualify as pseudo-agents is clearly a superset of the set of processes that qualify as agents proper, the working definition used in this text truly targets processes that are perceived as agents, not merely agents proper themselves. This is reflected in the qualifying statements next to the requirements of complexity, agency and compactness that these requirements ought to be perceivable for an external observer. This also means that the bar of these requirements is effectively lowered. It is not necessary that a process reaches a certain level of sophistication; instead it must only be perceived as having reached that level, e.g. by the individual interacting with it or being affected by it. Ultimately, this approach then moves away from the more stringent accounts of agency and to a more encompassing concept. This is in line with the aim of this thesis. Just as persuasion and autonomy constraints can be manyfold and imposed by many different entities in many different contexts, so the term used for those entities should be similarly encompassing. With these considerations, we arrive, now fully justified, at a meaning of the term agent that is also sufficiently reflective of its colloquial use and meaning within the domain of computer science, while espousing the more rigorous requirements of philosophical agents.

9.2.7 Comparison

There are other approaches on how to define agents generally and in the context of technology. Many recognize that attempts are unlikely to yield one-hundred percent satisfiable results and

"fail around the edges".[438] To give a sense of different approaches to definition, this section outlines three of the variations of agent conceptions that lead to different classification results.

Many conceptions deal with *autonomous* agents, or require at least some sort of (intrinsic) autonomy. [439] This connects closely to the requirements of complexity and agency used in the present model For example, Franklin and Graesser demand of an agent that it "is situated in and is part on some environment", which is similar to the point made in section 9.2.1 but also that "no other entity is required to fed it input or to interpret and use its output". However, they recognize that agents do require information to be in some sort of accessible format, remarking that when the environment is changed, the process in question may no longer be an agent. However, they note that other entities are also part of the environment, so insofar the environment is conducive to the agent in question, it seems that it does feed the agent input in some sense of the word. The argument made in section 9.2.1 does not go as far. Similarly, Franklin and Graesser require agents to exhibit temporal continuity, with computer programs that only act when called upon not fulfilling their requirements for agents. [440] Again, this stance is not adopted here. As the present approach to define agents does not rely on their intrinsic agency but on their externally perceived agency, temporal permanence is only necessary to the extent that entities interacting with the agent require it to ascribe it agency. Agents may be ascribed agency externally even if they, under the approach of Franklin and Graessar, do not have sufficient actual (autonomous) agency.

The type of action potential is also subject to differences of viewpoints. For example Woolridge and Jennings required an agent to have social ability, reactiveness, proactiveness (and autonomy). Here, reactiveness describes perceiving and responding to changes in the agent's environment. Proactiveness goes further and requires agents "taking the initiative" in pursuit of their goal. Social ability describes the capacity of an agent to interact with other agents. [441] Some of these requirements do not seem to be of utmost importance for the present inquiry. Communication between other agents (or humans) is too strong of a requirement. It is sufficient if agents can communicate with other non-agent processes that in turn communicate with humans. Agent-agent communication is not necessary at all for our purpose. Proactiveness is certainly indicative of a stronger concept of agency. But as mentioned before, the digital domain is usually only accessible through intermediary interfaces for individuals. Even if agents "roam" proactively in the digital domain, an individual's exposure is gated by their use of an intermediary interface. In a way, most cases in which an agent and an individual interact are subject environmentally imposed reactiveness. The agent can (usually) only interact, if the individual enters the digital domain. On the other hand, even proactiveness will rarely exist in an interaction vacuum. Arguments can be made to classify parts of an agent's environment as input prompting a reaction, invalidating any notion of true proactiveness.[442] Ultimately, excluding non-proactive algorithmic entities that are still complex and intent-driven seems unnecessary for analysing their impact on autonomy. Lastly, the notion of reactiveness is subsumed by agents being part of the informational pipeline, as they must be necessarily connected to their environment and hence perceive it in some sort.

---

[438] Stan Franklin and Art Graesser, 'Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents BT - Intelligent Agents III Agent Theories, Architectures, and Languages', ed. by Jörg P Müller, Michael J Wooldridge, and Nicholas R Jennings (Berlin, Heidelberg: Springer Berlin Heidelberg, 1997), pp. 21–35.
[439] See e.g. Jose C Brustoloni, *Autonomous Agents: Characterization and Requirements* (Citeseer, 1991); Franklin and Graesser.
[440] It is noteworthy, that their definition does explicitly reference autonomous agents, which does not fully align with the conception of agents propagated here. See Franklin and Graesser.
[441] Wooldridge and Jennings.
[442] This of course mirrors the debate of human agency in the light of potential determinism.

Furthermore, the concept that agents require some sort of deliberative competence is well recognized. However the type of competence is subject to debate.[443] One point of contention is if artificial agents ought to have deliberative qualities that are comparable to humans. This is often described as a set of mental qualities or mental states.[444]

Finally, this text would be incomplete without a reference to the concept of artificial agents established by Floridi and Sanders. They have considered three properties that are deemed relevant for agenthood, namely interactivity, autonomy and adaptability, based on which the agent may or may not also be a moral agent. As with the conceptions described above, the scoping of this text neither relies on intrinsic autonomy of technology, which is why this conception is not utilized here directly. It does however hold value when considering the moral responsibility in later sections of the text.

9.2.8 Implications

In practice, we can consider most types of informational processes to be either agents themselves or sub-processes of an agent. Hence, we may for the rest of this inquiry describe e.g. an information ranking algorithm or a social credit scoring application as agents without yielding to ambiguous terminology anymore. However, we can also see that this definition is not fully inclusive of phenomena within the digital domain. Comparing this to some examples of persuasive technology (such as dark patterns) in the previous sections makes clear that many of these do not necessarily rise to the level of compactness, delimitation and complexity. A dark pattern that creates its effect (purely) on the basis of cognitive biases connected to human perception can hardly be ascribed agency. However, it may be part of a bigger information process that could legitimately be described as an informational agent.

This type of analysis also allows to avoid some of the trickier delimitation problems that often arise when dealing with autonomy and constraints imposed on individuals pertaining to that autonomy. As mentioned above where the pragmatic account of autonomy was introduced, autonomy is better not understood as a binary process; indeed constraints may be imposed (or not imposed) on multiple dimensions of autonomy.[445] Indeed, it seems very likely that *some* constraint on autonomy is present at any given time.[446] At the same time, these constraints can also be imposed by a multitude of factors, informational factors among them. By ascribing agents a wide definition, e.g. perhaps encompassing more reflexive or passive information processes such as dark patterns, this means naturally that one has to consider many more potential sources of autonomy constraints. Combined with the insight stated above that

[443] See also Section […] above.
[444] John McCarthy, *Ascribing Mental Qualities to Machines.*, 1979.
[445] Here, this elaboration concerns the pragmatic account of autonomy, but the same problem arises with other accounts of autonomy as well.
[446] For example, informational autonomy requires that information that may undergird an individual's decision making is available to them in a form that is true, non-misleading and accessible. But it is not difficult to imagine in which these situations cannot all be fulfilled at the same time. As individuals have only limited mental capacities to process information, the trade-off between sufficient and overwhelming information exposure (irrespective of the veracity of the information) is limiting informational autonomy in some way in any case. A certain quantity of information exposure may simply be the optimum with respect to preservation of an individual's autonomy, with less information exposure meaning that more actions are foregone due to the individual not being aware of his ability to take them, and more information exposure meaning that the individual is mentally overloaded which diminishes his ability to process any information further. Even in this case, the individual is clearly not fully exposed to his decision potential, and so we may consider his autonomy (necessarily) diminished.

generally all environmental aspects will have *some* determinative effect on an individual and their autonomy, this threatens to quickly devolve any analysis. The result then may either be a hopelessly borderless area of inquiry or the simple statement that no-one is autonomous based on most everything around them. This is hardly a desirable result. Instead, a proper delimitation of agent (and autonomy), as is hopefully conducted here, allows a useful scope for analysis. Just as autonomy here is not considered binary (and hence likely unachievable), agents are not merely any factor that can constrain autonomy (and hence be omnipresent throughout all aspects of environment). As the scope narrows appropriately, so (again hopefully) the clarity of this analysis and its findings increase and sharpen.

9.3 Agent Characteristics

Having obtained a working definition of an informational agent, the next question that requires addressal are the peculiarities and idiosyncrasies of informational agent as opposed to other autonomy-constraining entities or forces such as humans, animals or environmental factors. Informational agents are distinguishable from other autonomy-constraining factors by the above definition and in particular by virtue of the domain they inhabit.

The next area of interest for this inquiry is then naturally how these agents differ from (and perhaps how they are similar to) other entities or factors that are *not* informational agents. In other words, to understand informational agents and their impact on individual autonomy, it may very well be relevant to have a sound understanding of their characteristics. Here, we may *prima facie* distinguish between intrinsic characteristics, i.e. characteristics that are directly derived from the structure of the informational agent and the domain it is embedded in, and other characteristics, e.g. characteristics that pertain to certain outcomes effected by the informational agent.

9.3.1 Intrinsic Characteristics or Outcome-Characteristics?

As part of the informational pipeline and the underlying domain of information technology, informational agent's characteristics are (at least partly) determined by these underlying structures. We may recall some of the biggest paradigm shifts within the underlying domain to be intermediation of information, immediacy of information transfer, lower barriers to impart (mis)information, anonymity or pseudonymity of information transfer participants, plurality of information recipients. personalized information exposure, and easy implementation of scalable, non-human actors. As informational agents are deeply embedded within this domain *and* often serve as part of its interfacing and output-process, many of these characteristics apply to agents as well or are at least mirrored in corresponding attributes. For example, just as information technology generally works at blazing speeds beyond intuitive human comprehension, so do the processes within informational agents, and perhaps the agent as a whole. Indeed, the very characteristic of easy implementation of scalable, non-human actors identified earlier pertains directly to the concept of informational agents: it is these very agents and their deployment opportunities that have become a defining characteristic of information technology as a whole. In other instances, we may recognize some sort of agent as a tool of execution that is relevant to a domain characteristic. For example, increased personalization requires some sort of informational agent that parses information and creates the necessary links between data points to alter the information exposure that is ultimately imposed on an individual. In other words, the process of personalizing is (part of) an informational agent, and as such may be seen as a relevant characteristic.

However, it is not necessary to limit the consideration of agent characteristics to their intrinsic qualities. Instead, one may consider their *impact* from certain external viewpoints, e.g. the ethical challenges they pose or the economic impact they may have. Existing scholarship has recognized this. For example, in their widely cited paper, Mittelstadt et al. identified the following ethically relevant characteristics of algorithms (i.e. agents within the dictum of this thesis for most purposes): inscrutability, adaptability, insight from inconspicuous data and irrelevance of anonymity.[447] Clearly some of these characteristics can also be seen as intrinsic (e.g. inscrutability may be considered as a degree of obfuscation within the internal processes of an informational agent), and not only be considered as outcome specific (e.g. the ethical or legal challenges the agent incurs as a result of their inscrutability). Nonetheless, the classification hints at a dogmatically different approach that is also in line with the previous text. Just as we have adopted an external view on how to define informational agents, their characteristics can be validly described using external observations, as the external impact these agents have (e.g. such as economic impact or ethical challenges brought forward by them) are (one of) the relevant mean(s) in which they can be understood.

9.3.2 Autonomy-Relevant Characteristics of Informational Agents

In the end, any list of characteristics of informational agent is but a selection of idiosyncrasies, plausible perhaps but likely incomplete. Similarly futile is to cleanly divide characteristics of agents themselves and interaction characteristics that describe the interplay of information between informational agents and their (human) recipients (which is the subject of the next section). Nonetheless, the following is an attempt to provide an overview over relevant informational agent characteristics. As the previous section has touched upon, we can distinguish between intrinsic characteristics and outcome-characteristics. This text identifies selective superior reasoning capabilities, (relative) immediacy of information processing, complex status between mistrust and deference, data reliance and susceptibility to bias and opacity of algorithmic decision-making, their potential to facilitate and amplify information transfer and their placement, function and effectuation obscurity as part of the set of intrinsic characteristics of informational agents.

9.3.2.1 (Relative) Immediacy of Information Processing

As previously described, changes in the underlying domain and the informational pipeline of which informational agents are a part of have created a situation in which information is processed and transferred at speeds not achievable by humans in most cases. We may recall that this immediacy can be seen as consisting of two factors. First, physical infrastructure (such as optic fiber between devices but also intra-device infrastructure such as a data bus or similar) has dramatically increased information transfer time as opposed to pre-information age technology, and this infrastructure is still evolving.[448] Second, and even more relevantly, processing of this information has become much speedier. Considering a certain amalgamation of processes on the informational pipeline as a black-box, the processing time to get from input to output data will generally increase with the computational capability of this black-box (as well as with the sophistication of the processes within). Of course, adopting an *outside*-view, or *accepting* this black-box approach is often to recognize it as an informational agent insofar the previously described requirements are met. As the information technology becomes faster, so do informational agents in processing information, and as a result in other derivative tasks.

[447] Mittelstadt and others.
[448] See again International Telecommunication Union, *Measuring Digital Development - Facts and Figures*. and Poletti and others.

As this text adopts an *outside* view of agents and their characteristics, it may be illustrative to compare actions that can be taken by informational agents and individuals alike. For example, face recognition (of familiar faces) is thought to take a human brain 140 ms,[449] while some face-recognition algorithms (with respective hardware) may confirm recognition (possibly with even higher accuracy) at 2.4ms.[450] Similarly, when pitted against undergraduate or postgraduate students by researchers, a machine-learning process was not only more accurate but also significantly faster in classifying research abstracts by their academic domain; with the algorithm requiring 5 seconds to classify 247 abstracts as opposed to the more than 2 hours the fastest human classifier needed for the same task.[451] (Of course, these comparisons, and in particular any times referenced on the side of algorithms, etc. are but a snapshot in time and highly dependent on the hardware used to perform the algorithm. However, with the existing sophistication of existing hardware, and the prospect of further advances in this field, as opposed to the comparatively glacial pace of evolution when it comes to "human hardware"[452], the maintaining of existing dominance of automated information processing in certain fields, and expansion of said dominance into other (but not necessarily all fields) is almost a given.)

9.3.2.2 Selective Superior Reasoning Capabilities

As a derivative of the raw speed of (parallel) computations that algorithms may undertake, informational agents can access underlying, hidden, information within a dataset that is, for all practical purposes, inaccessible to humans. This characteristic may be exhibited in varying intensities.

In certain tasks, informational agents may conduct tasks with a higher success rate than human individuals. For example, much research has been conducted with respect to the application of artificial intelligence in the medical field, e.g. algorithmic analysis of electrocardiograms (ECGs).[453] ECGs are a commonly performed test that is subject to complexity and human accuracy in performing an ECG is limited.[454] At the same time, informational agents (usually of a machine-learning variety) have already been found to exceed human accuracy in

[449] G. Barragan-Jason, M. Cauchoix, and E.J. Barbeau, 'The Neural Speed of Familiar Face Recognition', *Neuropsychologia*, 75 (2015), 390–401 <https://doi.org/10.1016/j.neuropsychologia.2015.06.017>.
[450] See e.g. Xiujie Qu and others, 'A Fast Face Recognition System Based on Deep Learning', in *2018 11th International Symposium on Computational Intelligence and Design (ISCID)* (IEEE, 2018), pp. 289–92 <https://doi.org/10.1109/ISCID.2018.00072>.
[451] Yeow Chong Goh and others, 'Evaluating Human versus Machine Learning Performance in Classifying Research Abstracts', *Scientometrics*, 125.2 (2020), 1197–1212 <https://doi.org/10.1007/s11192-020-03614-2>.
[452] For example, Worden has suggested that only a fraction of a bit *per generation* is changed in brain genetic information, which seems glacial indeed compared with the previously discussed notion of Moore's Law; see R.P. Worden, 'A Speed Limit for Evolution', *Journal of Theoretical Biology*, 176.1 (1995), 137–52 <https://doi.org/10.1006/jtbi.1995.0183>.
[453] Zachi I Attia and others, 'An Artificial Intelligence-Enabled ECG Algorithm for the Identification of Patients with Atrial Fibrillation during Sinus Rhythm: A Retrospective Analysis of Outcome Prediction', *The Lancet*, 394.10201 (2019), 861–67 <https://doi.org/10.1016/S0140-6736(19)31721-0>; E. Tatara and A. Cinar, 'Interpreting ECG Data by Integrating Statistical and Artificial Intelligence Tools', *IEEE Engineering in Medicine and Biology Magazine*, 21.1 (2002), 36–41 <https://doi.org/10.1109/51.993192>; Demilade Adedinsewo and others, 'Artificial Intelligence-Enabled ECG Algorithm to Identify Patients With Left Ventricular Systolic Dysfunction Presenting to the Emergency Department With Dyspnea', *Circulation: Arrhythmia and Electrophysiology*, 13.8 (2020) <https://doi.org/10.1161/CIRCEP.120.008437>; Ana Mincholé and Blanca Rodriguez, 'Artificial Intelligence for the Electrocardiogram', *Nature Medicine*, 25.1 (2019), 22–23 <https://doi.org/10.1038/s41591-018-0306-1>.
[454] See for an in-depth meta-analysis e.g. David A. Cook, So-Young Oh, and Martin V. Pusic, 'Accuracy of Physicians' Electrocardiogram Interpretations', *JAMA Internal Medicine*, 180.11 (2020), 1461 <https://doi.org/10.1001/jamainternmed.2020.3989>. According to this research, the median accuracy for cardiologists (!) was 74.9 %, and lower for all other personnel queried including practicing physicians.

diagnosing (certain) heart-related problems.[455] Similarly, informational agents have been shown to provide fast and accurate identification services e.g. with respect to breed of animals such as dogs[456] or cats[457], or medicinal plants[458]. Image recognition and classification generally seems to be a task well suited for informational agents (within limitations) as evidenced by the ImageNet Large Scale Visual Recognition Challenge, in which results of different image classification software are compared. In 2015, an informational agent (in this case a convolutional neural network) surpassed the accuracy of competing human classifiers for the first time.[459] This capability seemingly extends to drawn pictures (by humans) as well, as an informational agent was able to identify objects in drawn sketches with higher accuracy than human participants in an experimental setting.[460]

In other tasks, the speed and/or accuracy of informational agents allow the successful completion of that task that would be infeasible for human individuals. Profiling and personalization at scale may be theoretically possible but becomes practically infeasible when dealing with larger groups of individuals to serve information. For example, parsing personal information *en masse*, creating digital profiles and utilizing this information for personalized and targeted advertising is only possible through the use of artificial agents. This becomes even likely, when allowing for a varying array of personal information sources (e.g. health-data producing wearables).[461]

It is important to note, that these superior capabilities of informational agents are not universal, and limited to specific domains of comparisons. Classifying algorithms may be highly successful with respect to their specific setting, but useless outside of their domain. Informational agents also tend to rely on identifying correlation. As Brian Bergstein, executive editor of the MIT Technology review has summarized, informational agents tend to have difficulties with, among other things, grappling with understanding causation.[462] The superior

---

[455] Joon-Myoung Kwon and others, 'Comparing the Performance of Artificial Intelligence and Conventional Diagnosis Criteria for Detecting Left Ventricular Hypertrophy Using Electrocardiography', *EP Europace*, 22.3 (2020), 412–19 <https://doi.org/10.1093/europace/euz324>.

[456] Sandra Varghese and S Remya, 'Dog Breed Classification Using CNN', in *Security Issues and Privacy Concerns in Industry 4.0 Applications* (Wiley, 2021), pp. 195–205 <https://doi.org/10.1002/9781119776529.ch10>; Kenneth Lai, Xinyuan Tu, and Svetlana Yanushkevich, 'Dog Identification Using Soft Biometrics and Neural Networks', in *2019 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2019), pp. 1–8 <https://doi.org/10.1109/IJCNN.2019.8851971>.

[457] See e.g. Xiaolu Zhang, Luyang Yang, and Richard Sinnott, 'A Mobile Application for Cat Detection and Breed Recognition Based on Deep Learning', in *2019 IEEE 1st International Workshop on Artificial Intelligence for Mobile (AI4Mobile)* (IEEE, 2019), pp. 7–12 <https://doi.org/10.1109/AI4Mobile.2019.8672684>; Tita Karlita and others, 'Cat Breeds Classification Using Compound Model Scaling Convolutional Neural Networks.', 2022 <https://doi.org/10.2991/assehr.k.220301.150>.

[458] See for a high-level analysis Kalananthni Pushpanathan and others, 'Machine Learning in Medicinal Plants Recognition: A Review', *Artificial Intelligence Review*, 54.1 (2021), 305–27 <https://doi.org/10.1007/s10462-020-09847-0>.

[459] Kaiming He and others, 'Deep Residual Learning for Image Recognition', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–78. and Kaiming He and others, 'Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification', in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–34.

[460] Qian Yu and others, 'Sketch-a-Net: A Deep Neural Network That Beats Humans', *International Journal of Computer Vision*, 122.3 (2017), 411–25.

[461] See e.g. Davide C Orazi and Greg Nyilasy, 'Straight to the Heart of Your Target Audience: Personalized Advertising Systems Based on Wearable Technology and Heart-Rate Variability', *Journal of Advertising Research*, 59.2 (2019), 137–41.

[462] Brian Bergstein, 'What AI Still Can't Do', *MIT Technology Review*, 2020 <https://www.technologyreview.com/2020/02/19/868178/what-ai-still-cant-do/>.

reasoning capabilities of informational agents is hence selective, not universal, and specific to the domain and context the agent is applied to.

9.3.2.3 Opacity of Algorithmic Decision-Making

As pointed out by Mittelstadt et al. as a key epistemic concern, the (deductive or inductive) process of an informational agent to come to a certain conclusion based on a given set of information can be opaque and inaccessible and/or incomprehensible for human individuals. This opacity may be both due to factual characteristics of the agent as well as due to the peculiarities of the individual for which the agent's decision-making process is opaque. Understanding, predicting and reproducing algorithmic processes may be inaccessible to a laymen user but possible for a domain expert such as a software engineer or data scientist. However, there may very well be elements of an informational agents that shield its decision-making process even from the mind of familiar domain experts.[463] In this case, the decision-making process is not merely subjective but objectively opaque.[464]

Subjective opacity can be outlined on multiple axis. Opacity may be present e.g. with respect to the rules (i.e. code) set within the structure of the informational agent, their representation on different levels of abstraction or their general predictability. The below figure outlines this relationship in a simplified form.

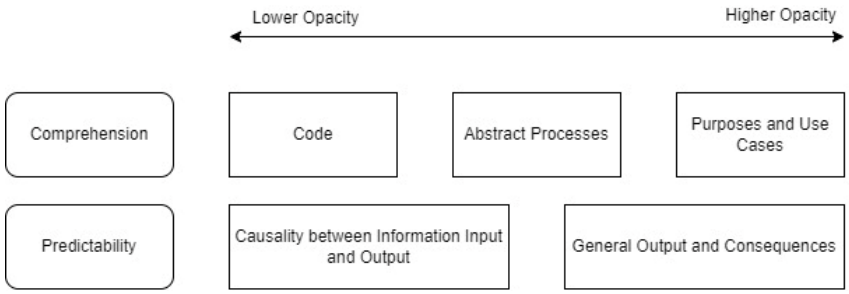

Figure 3

The computer code of an informational agent represents the most granular structure subject to potential opacity. It may be subject to secrecy (e.g. because it is proprietary intellectual property), or freely available. In the latter case, the complexity and legibility of the code may impose transparency obstacles: The higher the complexity of a given set of instructions, the less accessible these instructions will likely be. Similarly, at the same level of complexity, well-written code may often be easier to understand than worse-written code.[465] Of course, most

[463] There is a subtle distinction between this and the previously elaborated characteristic. Superior reasoning capabilities mean that algorithmic agents (in certain situations) succeed in information processing and consequent actions where humans cannot, but this does not necessarily imply that humans cannot understand the information processing that this entails or predict its outcomes. The matter of interest here is instead that algorithmic agents reasoning is not transparent enough (even if the information processing is within the reach of human individuals).

[464] See for this distinction Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms', *Big Data & Society*, 3.1 (2016), 205395171562251 <https://doi.org/10.1177/2053951715622512>. Burrell also suggests a third type of opacity (intentional corporate or state secrecy), but this seems like a dogmatic mismatch.

[465] See for recognition of this issue e.g. Raymond P.L. Buse and Westley R. Weimer, 'A Metric for Software Readability', in *Proceedings of the 2008 International Symposium on Software Testing and Analysis - ISSTA '08* (New York, New York, USA: ACM Press, 2008), p. 121 <https://doi.org/10.1145/1390630.1390647>; Mohammed Akour and Bouchaib Falah, 'Application Domain and Programming Language Readability Yardsticks', in *2016 7th International Conference on Computer Science and Information Technology (CSIT)*

laypeople are unlikely to penetrate the opacity that even comparatively simple computer code incurs due to their unfamiliarity with the programming language's syntax and semantics.

Individuals without the technical knowledge of understanding programming languages *per se* may nonetheless have the ability to understand the general processing pipeline of information within an informational agent. Abstracting software and programming commands to natural-language concepts may ensure (more abstract) transparency and lessened opacity, again lessened either by secrecy with respect to these processes or by complexity of (abstracted) processes that, in their entirety, are difficult to understand.[466]

Somewhat on the higher ends of abstraction is a purely predictive approach. One may consider informational agents as "transparent" and not opaque if they are (as a black box) understandable in their purpose and predictable in their output and effect. The majority of online advertising targets are very unlikely to have deep understanding of the underlying source code or even the abstracted processes from data collection, profiling, data brokering and advertisement bidding. They may however have an intuitive understanding that an informational agent embedded within e.g. a social network website presents them with personalized advertising based on previously consumed content. [467] Of course, the same may be true in the other direction as well. Individuals may very well have (or be given) insights to (parts of) the source code of an informational agent, but at the same time may not be aware of its use case or the information that is used as input, relativizing the figure shown above.[468]

Lack of knowledge about the workings, sources and purposes, and hence the opacity of informational agents is generally seen as a risk to privacy,[469] and as a result to individual's autonomy, underlining the salience of this characteristic to the inquiry at hand. In response to the ethical and practical concerns this agent characteristic raises, research is conducted to allow deeper insights into agent decision-making processes under the label of explainable AI (XAI).[470]

(IEEE, 2016), pp. 1–6 <https://doi.org/10.1109/CSIT.2016.7549476>; Todd Sedano, 'Code Readability Testing, an Empirical Study', in *2016 IEEE 29th International Conference on Software Engineering Education and Training (CSEET)* (IEEE, 2016), pp. 111–17 <https://doi.org/10.1109/CSEET.2016.36>.

[466] Here, this may again be alleviated by utilizing even higher levels of abstraction.

[467] Cf Andrew D Selbst and Solon Barocas, 'The Intuitive Appeal of Explainable Machines', *Fordham L. Rev.*, 87 (2018), 1085. The above reference also claims to outline why this purely superficial and/or intuitive level of understanding may be insufficient for many purposes, as inscrutability and non-intuitiveness do not always fully align with each other.

[468] One may argue that context itself is already necessary to truly "understand" processes at the code level, see e.g. Alberto Bacchelli and Christian Bird, 'Expectations, Outcomes, and Challenges of Modern Code Review', in *2013 35th International Conference on Software Engineering (ICSE)* (IEEE, 2013), pp. 712–21 <https://doi.org/10.1109/ICSE.2013.6606617>.

[469] See e.g. Catherine Dwyer, 'Task Technology Fit, the Social Technical Gap and Social Networking Sites', *AMCIS 2007 Proceedings*, 2007, 374.

[470] Lindsay Wells and Tomasz Bednarz, 'Explainable Ai and Reinforcement Learning—a Systematic Review of Current Approaches and Trends', *Frontiers in Artificial Intelligence*, 4 (2021), 48; Hani Hagras, 'Toward Human-Understandable, Explainable AI', *Computer*, 51.9 (2018), 28–36; Andreas Holzinger, 'From Machine Learning to Explainable AI', in *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, 2018, pp. 55–66.

9.3.2.4 Data Reliance and Susceptibility to Bias

Informational agents process information, and the quality and veracity of the information[471] parsed through the agent is correspondingly impactful. Informational agents rely on data being accessible to them, and operationalize this data towards a certain purpose. As a result, if data is not available or suffers from shortcomings, the data reliance of informational agents leads to the informational agent and its successful operation suffering shortcomings as well. Data reliance is then inextricably linked to data quality. Data quality in turn may suffer from not meeting sufficiency standards when it comes to (e.g.) accuracy, timeliness, precision, reliability, currency, completeness, relevancy, accessibility, interpretability, compatibility, security, coherence, validity and appropriate information format.

Two concepts of data reliance and risks to it seem particularly important. First, an agent's input might only be insufficient or incorrect information. The phrase "garbage in, garbage out" is often used to describe the phenomenon that data of insufficient quality tends to lead to output that is of similar insufficient quality, and as a result unreliable. Second, the agent itself may be structurally determined by insufficient or incorrect information. For example, ,f the informational agent encompasses processes that are the result of machine-learning, the rules and processes it applies to informational input may have themselves be shaped by information that is flawed. If the data set on which machine learning was applied to reflected biases, the agent may propagate the biases itself.[472]

Data reliance is salient to individual's autonomy as individuals tend to not have full control over their data. Individuals may not be aware of the extent of data an informational agent has at its disposal, or may not be aware of the inferences the informational agent may receive from this data. In case of chained agents, that is agents which output is then transferred as input to other agents, the reliance of the second agent on the information of the first may lead the second agent to take quasi-normative actions to the detriment of the individual. The individual however will often find it difficult to parse this information flow and determine the exact point of interaction necessary to defend against the autonomy constraint (which is closely connected to the aforementioned opacity of the agent), all of which would constitute an informational autonomy constraint.

9.3.2.5 Complex Status between Mistrust and Deference

Informational agents process information towards some sort of (intended) end or target. The output of this agent may be used by other informational agents, less complex processes or human individuals or groups to act. Finance institutions may rely on algorithmically calculated

[471] The terms data and information here are used interchangeably. This is not a necessity. Typically, information is seen as hierarchically higher whereas data is a more atomistic concept. On an abstract level, information (of any veracity) can be understood as well-formed, meaningful data (see Luciano Floridi, *Philosophy and Computing: An Introduction* (London/New York: Routledge, 1999); John Mingers and John Brocklesby, 'Multimethodology: Towards a Framework for Mixing Methodologies', *Omega*, 25.5 (1997), 489–509 <https://doi.org/10.1016/S0305-0483(97)00018-2>); whereas data may be understood as „a putative fact regarding some difference or lack of uniformity within some context (see e.g. Luciano Floridi, 'Semantic Conceptions of Information', *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University, 2019).

[472] See e.g. McKenzie Raub, 'Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices Comment', *Arkansas Law Review*, 71.2, 529–70 <https://heinonline.org/HOL/P?h=hein.journals/arklr71&i=549>. The academic debate and literature on bias in big data is legion and beyond the scope of this inquiry.

default risks when considering if they should grant a loan to an individual.[473] Within the context of criminal proceedings, bail-related or sentencing decisions may be taken in consultation to a predictive algorithm attempting to quantify the risk of re-offending.[474] Companies may use algorithms to sift through high numbers of applications, using automated procedures to screen candidates,[475] or to use them internally to track productivity of employments and base promotion decisions on that basis.[476] Logistic companies may determine salary incentives and penalties to their truck drivers based on an algorithm's verdict on how fast and efficient the driver transported its haul.[477] Informational agents exist in a tensioned state between being considered as trusted sources of information (and analysis) and being considered as insufficient, biased, impersonal and non-reliant actors. Individuals can fail to rely upon such information appropriately. Overreliance risks increasing the impact of algorithmic biases on the decisions made and can create ethical challenges, e.g. through loss of attributable responsibility for algorithmic decision making.[478] This deference to algorithmic decision making can be due to multiple reasons, most important of which is usually considered to be trust. The more an algorithmic agency is trusted (both analytically and affective-emotionally)[479], the more likely reliance on algorithmic decision making is afforded. Trust-

[473] See e.g. S Samsir, S Suparno, and M Giatman, 'Predicting the Loan Risk towards New Customer Applying Data Mining Using Nearest Neighbor Algorithm', in *IOP Conference Series: Materials Science and Engineering*, 2020, DCCCXXX, 32004; Yi He, Jian-chao Han, and Shao-hua Zeng, 'Classification Algorithm Based on Improved ID3 in Bank Loan Application', in *Information Engineering and Applications* (Springer, 2012), pp. 1124–30; Rafik Khairul Amin, Yuliant Sibaroni, and others, 'Implementation of Decision Tree Using C4. 5 Algorithm in Decision Making of Loan Application by Debtor (Case Study: Bank Pasar of Yogyakarta Special Region)', in *2015 3rd International Conference on Information and Communication Technology (ICoICT)*, 2015, pp. 75–80.

[474] Cf. Jason Tashea, 'CALCULATING CRIME: Attorneys Are Challenging the Use of Algorithms to Help Determine Bail, Sentencing and Parole Decisions', *ABA Journal*, 103.3 (2017), 54–59; Avi Feller and others, 'A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased against Blacks. It's Actually Not That Clear', *The Washington Post*, 17 (2016).

[475] Cf. Nathan R Kuncel, David M Klieger, and Deniz S Ones, 'In Hiring, Algorithms Beat Instinct', *Harvard Business Review*, 92.5 (2014), p32--32; Miranda Bogen and Aaron Rieke, 'Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias', 2018.

[476] See e.g. Brent Noack and others, 'Big Data Analytics in Human Resource Management: Automated Decision-Making Processes, Predictive Hiring Algorithms, and Cutting-Edge Workplace Surveillance Technologies', *Psychosociological Issues in Human Resource Management*, 7.2 (2019), 37–42; David T Newman, Nathanael J Fast, and Derek J Harmon, 'When Eliminating Bias Isn't Fair: Algorithmic Reductionism and Procedural Justice in Human Resource Decisions', *Organizational Behaviour and Human Decision Processes*, 160 (2020), 149–67. Concerns about rampant bias are especially prevalent within this class of algorithmic agents, see only Isobel Asher Hamilton, 'Amazon Built an AI Tool to Hire People but Had to Shut It down Because It Was Discriminating against Women', *Business Insider*, 2018.

[477] See for this discussion e.g. Dominic Loske and Matthias Klumpp, 'Intelligent and Efficient? An Empirical Analysis of Human--AI Collaboration for Truck Drivers in Retail Logistics', *The International Journal of Logistics Management*, 2021; Alvin J Williams and BabuP George, 'Truck Drivers--the under-Respected Link in the Supply Chain: A Quasi-Ethnographic Perspective Using Qualitative Appreciative Inquiry', *Operations and Supply Chain Management: An International Journal*, 6.3 (2014), 85–93; Matthias Klumpp, 'Automation and Artificial Intelligence in Business Logistics Systems: Human Reactions and Collaboration Requirements', *International Journal of Logistics Research and Applications*, 21.3 (2018), 224–42.

[478] See for this discussion the section "Traceability leading to moral responsibility" in Mittelstadt and others. The attribution of moral responsibility is hotly contested and complex. An in-depth investigation is beyond the scope of this section.

[479] See e.g. J. D. Lee and K. A. See, 'Trust in Automation: Designing for Appropriate Reliance', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46.1 (2004), 50–80 <https://doi.org/10.1518/hfes.46.1.50_30392>. Note that these factors may be assigned different weights based on their context and applicable environment. A commercial environment will likely rely more on analytical trust, e.g. the algorithmic agent's probability to meet performance targets, while a more casual environment the affective-emotional trust will likely increase in importance. For an example of commercial environments see

diminishing factors, such as privacy concerns or awareness of vulnerability, can increase reactance of individuals against informational agents, and decrease reliance.[480] This may be indeed detrimental in cases where the use of informational agents would be beneficial. Dependency on informational agents is most visible in a class of informational agents dubbed "recommender systems", which predict the alignment of certain objects (e.g. goods for purchases, TV shows to watch, food to order, etc.) with the preferences derived from a profile of an individual (or the group that individual belongs to).[481] Here again, trust plays an essential role in acceptance of the system and consequent incorporation into the individual's decision-making process.[482] As a result, informational agents (and their deployers) are often incentivized to maximize trust for minimal reactance and maximal reliance.[483] Complicating matters, establishing trust seems to play into subjective elements of human perception and interaction, and may be subject to human cognitive biases. For example, the (visual) interface design of an informational agent may already impact how much trust is awarded to it.[484] This highlights that trust may not always be warranted, even if present. In cases where trust is established by questionable means, e.g. if credibility is communicated without factual basis, or emphasized through the use of persuasive technology, trust becomes weaponized as an autonomy constraint.

9.3.2.6 Facilitation and Amplification of Information Transfer

As outlined in Section 8.2, the information domain allows the connection of a plurality of information sources and information recipients and sports increasingly lower barriers to impart information generally. No doubt, some of this is due to the increased physical interconnectivity that was identified in Section 8.1, but within this infrastructure, harmonizing standards and processes (and as a result informational agents) along the informational pipeline facilitate the ease of information flow and achieve said level of connectivity. For example, the Extensible Messaging and Presence Protocol, or XMPP, allows informational agents to facilitate direct

---

e.g. Christoph Keding and Philip Meissner, 'Managerial Overreliance on AI-Augmented Decision-Making Processes: How the Use of AI-Based Advisory Systems Shapes Choice Behaviour in R&D Investment Decisions', *Technological Forecasting and Social Change*, 171 (2021), 120970 <https://doi.org/10.1016/j.techfore.2021.120970>.

[480] Cf. Qi Chen and others, 'Understanding Consumers' Reactance of Online Personalized Advertising: A New Scheme of Rational Choice from a Perspective of Negative Effects', *International Journal of Information Management*, 44 (2019), 53–64 <https://doi.org/10.1016/j.ijinfomgt.2018.09.001>.

[481] Milano et al define recommender systems as "class of algorithms that address the recommendation problem using a content-based or collaborative filtering approach, or a combination thereof", see Silvia Milano, Mariarosaria Taddeo, and Luciano Floridi, 'Recommender Systems and Their Ethical Challenges', *AI & SOCIETY*, 35.4 (2020), 957–67 <https://doi.org/10.1007/s00146-020-00950-y>.

[482] See e.g. Sarita Herse and others, 'Do You Trust Me, Blindly? Factors Influencing Trust Towards a Robot Recommender System', in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (IEEE, 2018), pp. 7–14 <https://doi.org/10.1109/ROMAN.2018.8525581>.

[483] This is not a necessary conclusion in all cases. Indeed, there may very well be situations in which critical engagement with the information outputted by algorithmic agents is not just warranted but preferred. For example, on commercial settings, over-reliance on potentially faulty calculations is hardly desirable by the deployer. However, in many situations the interest of the deploying entity and the interests of the interaction targets are not aligned; e.g. in online advertising situations, the deployer of profiling and advertising algorithms is interested in high engagement and realization of sales of goods and services, and likely less in if the purchase of these goods really were in the best interest of the individual targeted. Insofar purchase rates do not decrease because of interest mismatch, the faulty calculations of the profiling agent are commercially irrelevant for the deployer (in this moment in time).

[484] Nava Tintarev and Judith Masthoff, 'A Survey of Explanations in Recommender Systems', in *2007 IEEE 23rd International Conference on Data Engineering Workshop*, 2007, pp. 801–10. See also Joseph B. Lyons and others, 'Shaping Trust Through Transparent Design: Theoretical and Experimental Guidelines', 2017, pp. 127–36 <https://doi.org/10.1007/978-3-319-41959-6_11>.

messaging and contact list maintenance.[485] Building on this, instant messaging applications such as WhatsApp or Telegram may allow cheaper and faster transfer of (specific types of) information (e.g. multimedia content) than previous processes (such as the ill-fated MMS) to an increasing number of recipients.[486] Similar facilitative communication standards exist in the domain of Voice over IP (VOIP).[487]

Beyond facilitation by enabling certain channels of information transfer, processes drawing from the wealth of information within the domain can also help to amplify the effectiveness of communication. As potential information recipients are typically attached to the domain, they are subject to constant observation (e.g. through the use of cookies)[488] and are subsequently easier to interact with in a personalized form. The general notion of personalization has already been recognized as a underlying domain characteristic in Section 8.2.6, but effective personalization capabilities of an agent can serve as a powerful amplifier of the information transferred.[489] This is also of interest with respect to the main inquiry at hand. If information by itself can be used to impose autonomy constraints, then its facilitation and amplification may also facilitate and amplify the constraint imposed or raise the information to the level of an autonomy constraint to begin with.

9.3.2.7 Placement, Function and Effectuation Obscurity

As is pointed multiple times within this text, one of the main domain characteristics underlying informational agents is the fact that all information must be intermediated to be readable for a (human) information recipient). Recalling that the informational pipeline describes the flow of information from information origin to recipient, and that information flows through a technological medium, meaning it needs to be reinterpreted for the recipient to perceive it, this means that the information transfer processes are typically somewhat of a black box for participants. During the stage at which information passes through the domain of information technology, it is potentially subject to observation and alteration processes and hence agents that are equipped with such. However due to the general opacity of such programs, the addition of additional processes or agents that observe or alter the information that is transferred is not necessarily visible for information recipients or third parties as illustrated in Figure 4.

---

[485] Adrian Hornsby and Rod Walsh, 'From Instant Messaging to Cloud Computing, an XMPP Review', in *IEEE International Symposium on Consumer Electronics (ISCE 2010)*, 2010, pp. 1–6.

[486] Readers may remember that MMS where usually accounted for by message when it comes to billing them against the user, whereas instant messaging is typically billed only for the data used, which can be agnostic to the number of recipients as this distribution is done at a central level of the messaging service provider.

[487] E.g. Opus used in WhatsApp, SILK used in Skype or AAC-LD used in FaceTime, see https://en.wikipedia.org/wiki/Voice_over_IP

[488] Note that these type of observative measures are incredibly powerful. Cahn et al. have suggested that even if only selecting for 1 % of entities that place cookies on the web, these cookies can aggregate information across 75% of web sites, see Aaron Cahn and others, 'An Empirical Study of Web Cookies', in *Proceedings of the 25th International Conference on World Wide Web*, 2016, pp. 891–901.

[489] The effectiveness of personalization e.g. in online advertising is by now well established. However, this does not necessarily imply that higher personalization always leads to higher effectiveness of the information transfer. Indeed, information recipients may become wary when recognizing personalization e.g. when they feel the underlying personal data has been collected without their consent. This is sometimes dubbed the personalization paradox. Similarly, effectiveness can be greatly affected by the context of the information embedding such as timing and placement. See e.g. Hisham Abdel Monem, 'The Effectiveness of Advertising Personalization', *Journal of Design Sciences and Applied Arts*, 2.1 (2021), 114–21; Bleier and Eisenbeiss, 'Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where'; Elizabeth Aguirre and others, 'Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness', *Journal of Retailing*, 91.1 (2015), 34–49 <https://doi.org/10.1016/j.jretai.2014.09.005>.

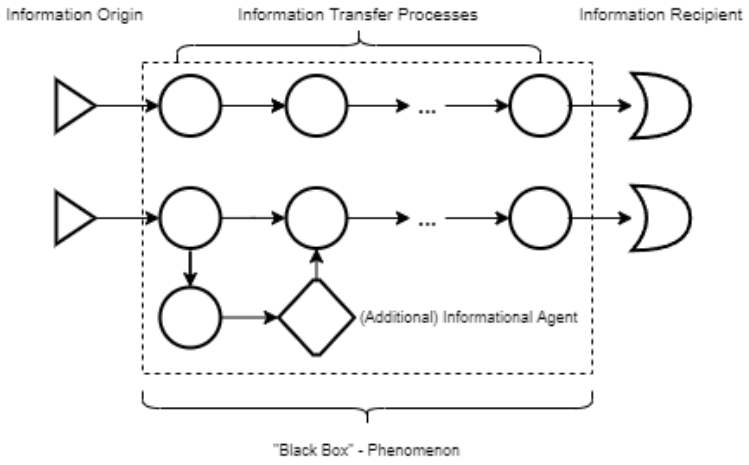

Figure 4

As a result, informational agents are characterized by a certain level of obscurity in many situations in which they are part of an informational pipeline. This obscurity can be manifest in different situations.

Uncertainty can be introduced with respect to the placement (or "existence" of informational agents. Where no direct interaction between an informational agent and the information recipient takes place, agents may be utilized for purposes of mere observation e.g. of the online behaviour of an information recipient. While there may be legal obligations forcing the disclosure of the use of such agents, particularly in the European Union under the GDPR,[490] there may be little factual control mechanisms. The same applies to information flows which general observation is disclosed, but the depth of observance and analysis (and its purpose) conducted of this information is not. For example, deployers of informational agents in an e-commerce setting may have incentives both to use effective agents with observational capabilities (e.g. for use of profiling, A-B testing or personalized display of information) and to hide them from potential consumers as to not loose trust[491], and consequentially sales.[492] In cases where direct interaction happens (e.g. the observed information is directly translated into a personalized interface), the link between observed data and interaction (e.g. the fact or degree of personalization) is also outside of the view of the information recipient. Lastly, in cases of indirect interaction, individuals are even further removed from the placement of an informational agent. In cases where the information recipient is not the individual whose information is analysed (e.g. when a financing institution consults an informational agent with respect to establishing a credit score about an individual), they interact with a third entity whose actions are guided by informational agents about which the individual will have limited

[490] While the GDPR implications have been discussed already within this text, there is also discussion outside of the European Union. For incentives and issues within an USA context see e.g. Sylvia Lu, 'Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence ', *Vanderbilt Journal of Entertainment & Technology Law*, 23.1, 99–160 <https://heinonline.org/HOL/P?h=hein.journals/vanep23&i=111>.

[491] Alisa Frik and Luigi Mittone, 'Factors Influencing the Perception of Website Privacy Trustworthiness and Users' Purchasing Intentions: The Behavioural Economics Perspective', *Journal of Theoretical and Applied Electronic Commerce Research*, 14.3 (2019), 89–125.

[492] See for a quantitative assessment e.g. Marcus D Odom, Anand Kumar, and Laura Saunders, 'Web Assurance Seals: How and Why They Influence Consumers' Decisions', *Journal of Information Systems*, 16.2 (2002), 231–50.

information on at best. On the other hand, customers may intentionally interact with a recommender system to receive personalized information, validating the interaction with their approval or consent or approving of a system because they perceive it as added value, in which case their transparent placement may be advantageous.[493]

As a result, informational agents are characterized by the fact that their obscurity, i.e. visibility to information recipients, or information-affected entities, is not (fully) within the power of those entities. Instead, these must rely on legal incentives or commercial obligations of the deployers of informational agents for disclosure of the placement, function, and effect of informational agents.

9.4 Intermediate Conclusion

The comparatively high density of this section warrants revisiting the most important aspects highlighted herein. The overarching motivation that is addressed in this thesis is the question of how individual autonomy can be exercised and constrained by informational agents. The first third of the text has outlined what it might mean to be autonomous. This section now has followed research sub-question 2.3, which demanded an investigation into the vectors of influences connected to processes neither of the domain nor the information recipient. To this end, it has introduced entities beyond the individual and put the focus on analysis on such entities that introduce particular risks to individual autonomy. Within this text, these origins of risks have been denoted as informational agents. To derive at satisfactory definition of this concept, this section has built upon the concept of the informational pipeline, an abstraction of information transfer processes made up by the spheres of information origin, transfer and transmission infrastructure and information recipient.[494] It is on between the endpoints of this pipeline that we can locate informational agents. At the same time, this text has introduced complexity thresholds, with simple algorithms on the lower end and fully fledged artificial intelligence at the other end of the spectrum.[495] The plane that is spanned by these two axes of logistical and complexity-characteristic seems intuitively populated by the entities of interest. On the basis of this, the text has given a definition of informational agent that relies partly on perceptions of complexity (and agency, naturally) as seen *from the outside*.[496] However, it is noteworthy that while this definition rounds out the arguments presented herein and attempts to close methodological gaps within this thesis, its adoption is not strictly necessary for all the findings made herein. Adopting a different take on what ought to be considered an agent within the IoE-sphere should not substantially change the persuasiveness of the other arguments presented herein.

While the definition of the term informational agent may be of abstract relevance, the characteristics of such agents is even more pertinent. The text has focused on a few out of many characteristics that may be ascribed to informational pipeline processes and has focused on characteristics that seem particularly pertinent to individual autonomy. Characteristics identified (and contextualized vis-à-vis some of their autonomy implications) were the immediacy of information processing, selective superior reasoning capabilities, opacity, data

[493] See for a very detailed investigation into the experience of affected individuals by such informational agents Bart P. Knijnenburg and others, 'Explaining the User Experience of Recommender Systems', *User Modeling and User-Adapted Interaction*, 22.4–5 (2012), 441–504 <https://doi.org/10.1007/s11257-011-9118-4>. Other examples of potentially trust enhancing informational agents may be found e.g. in the domain of fraud detection.
[494] See Sub-Section 7.1
[495] See Sub-Section 9.1.2
[496] See Sub-Section 9.2

reliance and bias susceptibility and the complex position of informational agents in tension between mistrust and deference, their property of facilitating and amplifying information transfer and the obscurity awarded to their placement, function and effectuation.[497]

The avid reader will have already picked up on the fact that the selection and ordering of characteristics is somewhat circular and ambiguous. Within this text, characteristics of the overarching domain[498] and of the entities within the domain have been introduced with considerable overlap. This seems unavoidable, as the entities (i.e. agents) are not just inhabiting the domain but are constitutive elements of the domain itself. As mentioned shortly in previous text,[499] the assignation of certain characteristics either to the domain or to entities is often a matter of choice and not necessarily exclusive. However, as with the definition of informational agents above, adopting a different assignation is hopefully unlikely to threaten the persuasiveness of the text at large; indeed the stipulative ordering and contextualization of abstract concepts within the intersection auf autonomy and the sphere of the IoE is for many purposes an accessibility-device and not a prescriptive effort. In any case, both the domain and agent characteristics are closely connected, as outlined in the below Figure.

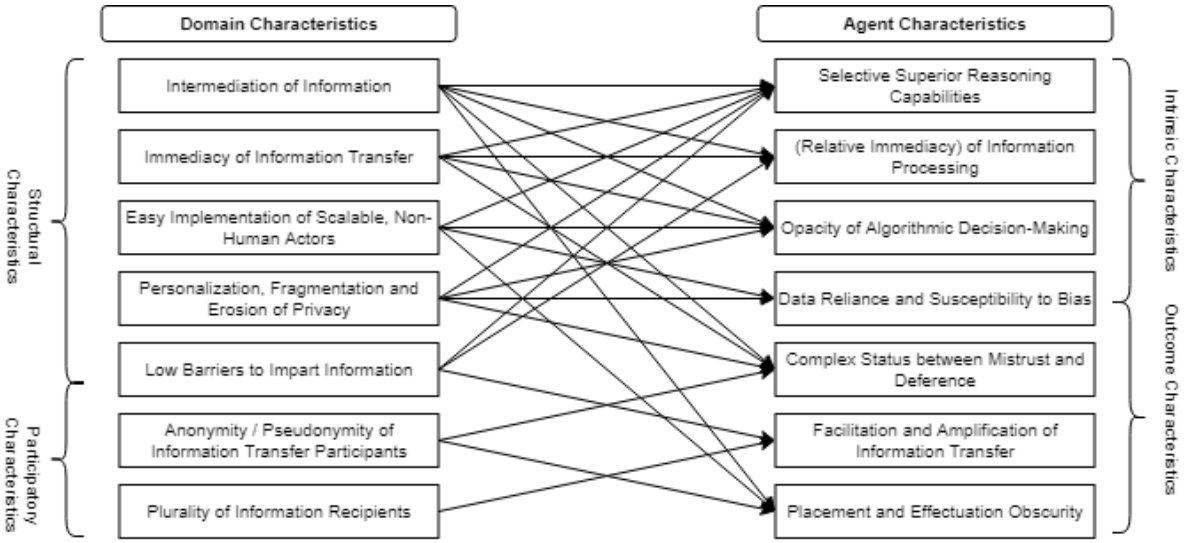

Figure 5 – Connection between Domain Characteristics and Agent Characteristics

[497] See Sub-Section 9.3.2
[498] See Section 0
[499] See again Sub-Section 9.3.2

10 Vectors of Influence in the Context of Information Receival

Recalling the above definition of vectors of influence, we may distinguish them by their general objects of reference. First, aligning this with the previously introduced concept of the informational pipeline, we can generally situate these vectors in the nexus of information source, transfer processes or information recipient. The previous sections have investigated the previously identified domain of information technology and agent characteristics as vectors of influence, generally situated in the first two regions of the informational pipeline, analysis of which was also conducted above.

This section will hence focus on vectors of influence that pertain most closely to the information recipient and answer the research sub-question 2.4, which asked what typical patterns and phenomena characterizing risk to an individual's autonomy exist that can be contributed to the context of information receival and are mainly connected to the sphere of the information recipient. Analysis will be guided by the following structure:[500] Section 10.1 outlines the connections between the present vectors of influence and the previously discussed informational agents and their characteristics. Section 10.2 briefly touches upon the two sources of instructive examples this section uses, namely the domain of misinformation and the domain of location-based services. Sections 10.3, 10.4, 10.5, identify three vectors which correspond closely to the information recipient, namely the medium of information transfer, the (perceived) veracity, accuracy, trustworthiness and assertiveness of the information transfer, and the transparency and completeness of information. Section 10.6 deals with the phenomenon of assignation of actionable data, highlighting how the information recipient and the constrained individual do not have to be identical. Section 10.7 offers a conclusion for this section and the entire Part II of this thesis.

10.1 Introduction

I suggest here that the following interaction characteristics are of particular interest and relevance for the purpose of this analysis (and in general with respect to understanding autonomy constraints imposed by technology): (1) the medium of information transfer and its characteristics, (2) the perceived accuracy, precision, trustworthiness, and assertiveness of information, (3) transparency to the user with respect to the information collection process of the agent and alternative action potentials available to the user (where applicable), and (4) the assignation of actionable data to a human user.

[500] The avid reader is encouraged to forgive the "flat" structure and lack of nested sub-section which puts it at aesthetic odds with the rest of the thesis.

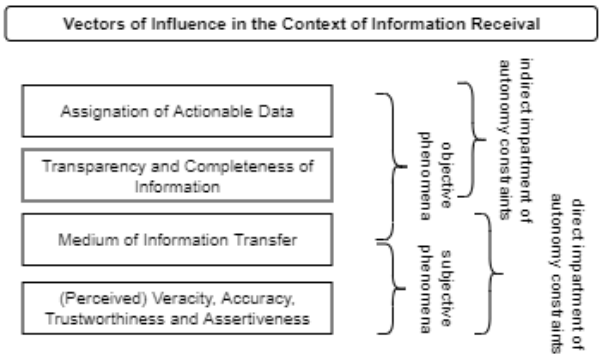

Figure 6 - Visualization of Vectors of Influence in the Context of Information Receival

As above in Figure 5, it can be useful to understand these vectors to be closely connected to the previously established. Indeed, as information transfer is a process that is invariably connected to the agents present on the informational pipeline their characteristics reflect upon the information receival-specific vectors of influence as well., as outlined in the below figure.

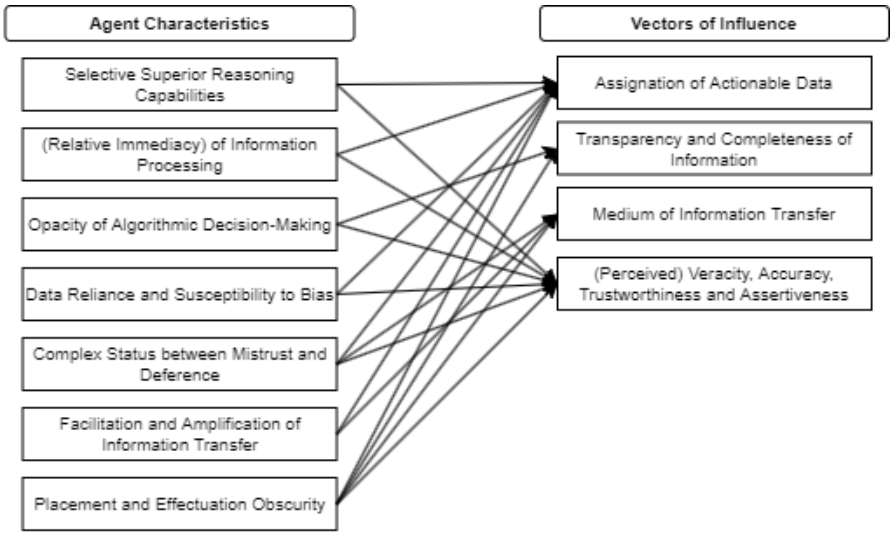

Figure 7 - Connections between Agent Characteristics and Vectors of Influence in the Context of Information Receival

In addition to this compartmentalization, we can differentiate further between such vectors as outlined in the figures above. The most salient differentiation that appears sensible is to distinguish between objective and subjective phenomena. As the information recipient is ultimately a human individual, this means that their receival and processing of the transferred information is subject to assessment and filtering through limited human perception, and subject to biases, heuristics, simplifications and other subjective factors.

We may also differentiate between interaction qualities of informational agents and the ultimate information recipient. Agent-human interaction can be characterized on different dimensions. Most fundamentally, informational agents can affect an individual's autonomy

twofold. The informational agent may interact with the individual, e.g. the user, directly, by using some sort of interface of transmitting information between the informational agent and the user. This will encompass most if not all services that are consciously enabled by a user for the purposes of benefitting directly from the informational agent. For example, a digital assistant on a smartphone responding to a query for information interacts with an individual directly via (audio-) visual outputs. However, an informational agent may also be non-interactive from the side of the affected individual. Contextual data may be collected by an informational agent that does not benefit the individual directly, or at all. Often, there may be informational friction, with the individual not in full control or knowledge about the data collected at any given time. Informational agents that belong to the class of agents most removed of user-interaction are e.g. those that assign data to the individual for the purposes of profiling, which in turn can be used for the purposes of (personalized) information transfer.

10.2 Practical Context

To contextualize the coming analysis, I use here two domains of interest to generate examples for the interaction characteristics and persuasion-imparting processes that are of interest to the coming chapters, namely the domain of misinformation and the domain of location-based services. For completeness' sake, both are extensively described in Annex 10.2a and Annex 10.2b. In short, LBS describe a concrete application of information processing, while misinformation can be seen as almost a meta-domain. Both do share exposure to increasing public attention. For example, misinformation has become a topic of increased interest in the wake of the Cambridge Analytica-scandal, in which targeted (and perhaps misleading) online advertising was allegedly used based on the strong profiling of the social media site Facebook.[501] Location-based services have similarly received increased attention, with recent concerns including leaks of location data originating e.g. from Muslim prayer apps[502] or discovery of secret military bases by analysing running data from a sport tracking application.[503] This approach also helps connect the (abstract) formulation of vectors of influence further to existing academic debate, as well as providing worthwhile context apart from the stereotypical

10.3 Medium of Information Transfer

Interaction between an informational agent and an individual are necessarily intermediated, usually through some type of interface. The imparting of information may be conducted e.g. via audio-visual or haptic cues (e.g. via a screen, voice announcement, etc.). In particular, visual interfaces may engage with a multitude of known biases, with respect to both perception and cognition that the user may display [504], the most obvious of which is the implicit or explicit ranking of the displayed information; a choice likely to influence the user [505]. Misinformation may be (and often is) favourably ranked when calculating the order of appearance of postings

[501] Christophe Olivier Schneble, Bernice Simone Elger, and David Shaw, 'The Cambridge Analytica Affair and Internet-Mediated Research', *EMBO Reports*, 19.8 (2018), e46579.
[502] See e.g. https://www.vice.com/en/article/xgz4n3/muslim-app-location-data-salaat-first
[503] See e.g. https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases.
[504] Emily Wall, John Stasko, and Alex Endert, 'Toward a Design Space for Mitigating Cognitive Bias in Vis', in *2019 IEEE Visualization Conference (VIS)* (IEEE, 2019), pp. 111–15 <https://doi.org/10.1109/VISUAL.2019.8933611>.
[505] Mohd Taib, 'Loophole Allows Bias in Displays on Computer Reservations Systems', *Aviation Week & Space Technology*, 132.7 (1990), 137; Batya Friedman and Helen Nissenbaum, 'Bias in Computer Systems', *ACM Transactions on Information Systems*, 14.3 (1996), 330–47 <https://doi.org/10.1145/230538.230561>.

on a social media-*newsfeed* based on the underlying informational agent's incentives (e.g. engagement rate) and displayed in a way that is (increasingly) appealing to the interacting users. Due to the existing constraints of accessing this information e.g. via a limited-size screen, sufficiently highly ranked misinformation can then displace legitimate information, hence diminishing informational autonomy. Insofar prominent placement of participative markers (e.g. likes, shares, etc.) charge misinformation with additional authority for the information recipient, this may impose social pressure or engage cognitive biases that defer to a perceived majority, both relational autonomy constraints (see for some of these cognitive heuristics [506]. Ranking is also used in the context of LBS, in particular when it comes to route-finding devices. Depending on which options are shown and how they are shown, users can be nudged into taking certain routes instead of others as a derivative result of the visual medium and its design that is used to intermediate information. For example, the wide-used routing service Google Maps recently introduced a mode of navigation focused on maximizing "sustainability" by highlighting separate routes that provide the best fuel efficiency (as opposed to competing metrics such as distance travelled or time spent).[507] Depending on how this information is ranked and presented, perceptive and cognitive biases interact with this information. Interfaces may also be designed deliberately hostile to increase mental load of the user, which in turn may make them more susceptible to biases [508] and diminishing informational autonomy all the same. In general, information transfer is conducted in a way that follows the design of the relevant technology's (e.g. LBS') front-end. To the extent it is interactive, the interface may be understood as an example of choice architecture, i.e. a decisional infrastructure "organizing the context in which people make decisions" [509]. Of course, choice architecture is generally capable of altering individual's behaviour to a certain extent, i.e. the aforementioned nudge [510]. This is potentially even more effective if the choice architecture is personalized via the data collected on the user [511]. Using the language of this text, the medium of information transfer constitutes a vector of influence that may constrain an individual's autonomy in particular on their informational dimension, e.g. via informational overload in case of hostile design,[512] and on their relational dimension, in particular by imposing factual and mental constraints, e.g. by the exploitation of cognitive or perceptual biases. At the same time, very effective (and potentially value-laden) interface design and the (mis)information it transports may qualify as an intrinsic autonomy constraint, as it may insert or affect the individual's mental states, e.g. via creating an assumption of value through the display of ranked information or context of social value, whereas these impart influence on the individual's preferences.

[506] Metzger & Flanagin, 2013)
[507] Note that these metrics can sometimes be misleading in themselves which would impose another informational autonomy constraint, see e.g.
[508] Sarah R. Allred and others, 'Working Memory and Spatial Judgments: Cognitive Load Increases the Central Tendency Bias', *Psychonomic Bulletin & Review*, 23.6 (2016), 1825–31 <https://doi.org/10.3758/s13423-016-1039-0>.
[509] Thaler, Sunstein, and Balz.
[510] Kosters and Van der Heijden.
[511] Omer Tene and Jules Polonetsky, 'Big Data for All: Privacy and User Control in the Age of Analytics', *Northwestern Journal of Technology and Intellectual Property*, 11.5 (2013), 239–72.
[512] See for this concept e.g. Benjamin Scheibehenne, Rainer Greifeneder, and Peter M. Todd, 'Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload', *Journal of Consumer Research*, 37.3 (2010), 409–25 <https://doi.org/10.1086/651235>; Petra Persson, 'Attention Manipulation and Information Overload', *Behavioural Public Policy*, 2.1 (2018), 78–106; Alexander Chernev, Ulf Böckenholt, and Joseph Goodman, 'Choice Overload: A Conceptual Review and Meta-Analysis', *Journal of Consumer Psychology*, 25.2 (2015), 333–58 <https://doi.org/10.1016/j.jcps.2014.08.002>.

10.4 (Perceived) Veracity, Accuracy, Trustworthiness and Assertiveness

Veracity of information is generally considered as an important quality of the same. This holds true for the subjective assessment of the interacting information recipient as well: quality of information is perceived to be of paramount relevance.[513] It seems a trivial conclusion that perceiving information as more accurate or truthful, creates derivative expectations such as trust in the information or credibility in the information provider.[514] Indeed, information sources that are seen as highly trustworthy (albeit perhaps from a subjective viewpoint) may significantly affect the impact the (mis)information transferred has for the information recipient.[515] As a consequence, individuals may act upon that (lack of) trust or credibility. Considering the example of a navigation algorithm, it is intuitively apparent that users will act upon it likelier, the more they trust the accuracy of the algorithm in question. Insofar an individual acts upon (and hence trusts in) inaccurate or misleading information they are likely experiencing constraints on their informational autonomy, as they must exercise their autonomy on the basis of faulty information. This issue is compounded by the fact that individuals may display deference to algorithmic decision-making. [516] (This may be mitigated by creating accurate expectations of the receiver with respect to the information transferred. For example, where the informational agent is in itself the source and not merely the conveyor of information it may disclose information about the quality of the output, e.g. a human-readable confidence score which can prevent non-justified reliance on the information in question.) Intrinsic autonomy constraints may further be imposed by an overly effective or assertive informational agent spreading misinformation: Informational agents may insert preferences or information into the individual's decisional structure through information transfer along their vectors of influence [517], and so can utilize faulty information for these purposes as well. (At the same time, particular assertiveness of an information source may be used to detect misinformation in the first place.) [518]

10.5 Transparency and Completeness of Information

The interaction of informational agents and individuals usually exhibits an information asymmetry. With the increasing prevalence of personalization the incentives for wide-ranging data collection are growing correspondingly. As the informational (intermediary) agents do not (or cannot) always discern the veracity of the information intermediated as a result of

[513] Min Kyung Lee, 'Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management', *Big Data & Society*, 5.1 (2018), 205395171875668 <https://doi.org/10.1177/2053951718756684>.
[514] (see for these concepts e.g. Kelton, Fleischmann, & Wallace, 2008)
[515] Sara Pluviano, Sergio Della Sala, and Caroline Watt, 'The Effects of Source Expertise and Trustworthiness on Recollection: The Case of Vaccine Misinformation', *Cognitive Processing*, 21.3 (2020), 321–30 <https://doi.org/10.1007/s10339-020-00974-8>.
[516] Eric Bogert, Aaron Schecter, and Richard T. Watson, 'Humans Rely More on Algorithms than Social Influence as a Task Becomes More Difficult', *Scientific Reports*, 11.1 (2021), 8028 <https://doi.org/10.1038/s41598-021-87480-9>.
[517] Christopher Burr, Nello Cristianini, and James Ladyman, 'An Analysis of the Interaction Between Intelligent Software Agents and Human Users', *Minds and Machines*, 28.4 (2018), 735–74 <https://doi.org/10.1007/s11023-018-9479-0>; Katja de Vries, 'Identity, Profiling Algorithms and a World of Ambient Intelligence', *Ethics and Information Technology*, 12.1 (2010), 71–85 <https://doi.org/10.1007/s10676-009-9215-9>.
[518] Hannah Rashkin and others, 'Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking', in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2931–37.

personalization, introducing false or misleading information threatens to widen this asymmetry. Different situations relevant to an individual's autonomy are conceivable. The affected individual may lack information about the data collection and analysing processes they are subjected to, and the consequences with respect to personalization that they may entail, which constitutes an informational autonomy constraint by itself. Lack of transparency may also undermine the trust of individuals into the informational agent [519], which can impede individuals consciously using the relevant technology to exercise their autonomy. In other words, the individual may be *inhibited* from exercising their autonomy via use of a technology in question due to a lack of trust in the very same technology, which would qualify as a relational-mental constraint in the previously discussed framework. Inhibition of individuals to exercise their autonomy freely because of concerns related to the algorithmic observation, a relational-mental constraint of sorts, may be counteracted by establishing trust into the informational agent [520] and allows the individuals exercise of their autonomy with respect to data collection and processing that may impede or support their autonomy in the future. [521] In connection with LBS deference as discussed above, lack of information about other viable recommendations (.g. an alternative but equally viable route that is not displayed on a navigational device) constrains the individual in their ability to autonomously choose an alternative action, e.g. the alternative route. Indeed, limiting exposure to potential venues of autonomy exercise imposes relational-factual constraints in addition to any informational constraints, and is a well-recognized source of diminished autonomy. [522]

Misinformation within an informational pipeline tends to radiate out and affect aspects of an individual's autonomy beyond mere informational autonomy. As already hinted at above, an important area in which this becomes visible is the use of recommender systems. Recommender systems of all types already suffer from a sort of principal-agent paradox, in which the interests of the provider and the information recipient are not necessarily aligned. [523] (The closer the results of the recommender systems are aligned with the true interests of the users, the more likely they are to help curb informational autonomy constraints; the more these systems reflect other interests to the detriment of the true interest of the users, the more likely they are to impose informational autonomy constraints.) However, even the most well-intended recommender system (that may have the corresponding trust from its users) may turn harmful towards its user's autonomy when "poisoned" with misinformation. Such systems may have implications for third parties as well; the respective informational agents may have tacit stakeholders that benefit or depend on the recommender system for their exercise of autonomy [524].

[519] Robert H. Sloan and Richard Warner, 'When Is an Algorithm Transparent? Predictive Analytics, Privacy, and Public Policy', *IEEE Security & Privacy*, 16.3 (2018), 18–25 <https://doi.org/10.1109/MSP.2018.2701166>.
[520] Donghee Shin and Yong Jin Park, 'Role of Fairness, Accountability, and Transparency in Algorithmic Affordance', *Computers in Human Behaviour*, 98 (2019), 277–84 <https://doi.org/10.1016/j.chb.2019.04.019>.
[521] Sloan and Warner.
[522] Ansgar Koene and others, 'Ethics of Personalized Information Filtering', 2015, pp. 123–32 <https://doi.org/10.1007/978-3-319-18609-2_10>; Mariarosaria Taddeo and Luciano Floridi, 'How AI Can Be a Force for Good', *Science*, 361.6404 (2018), 751–52 <https://doi.org/10.1126/science.aat5991>.
[523] Kartik Hosanagar, Ramayya Krishnan, and Liye Ma, 'Recomended for You: The Impact of Profit Incentives on the Relevance of Online Recommendations', in *ICIS 2008 Proceedings*, 2008, p. 31.
[524] Milano, Taddeo, and Floridi.

10.6 Assignation of Actionable Data

The impact of information processing is not limited to the primary information recipient. Information (and misinformation) can effect their behaviour in ways that impose autonomy constraints further *downstream* of the causal chain.[525] First, they may amplify and facilitate the spread of misinformation from individual to individual. However, informational agents may also themselves be recipients of misinformation. Informational agents engaging in profiling activities collect significant amount of data both about an individual and their relationship to their environment and other entities; this data may be used to gain further (predictive) insights into the respective individual and their behaviour [526]. These profiles may then be used by other entities to make decisions affecting the individual, such as excluding them from purchasing certain products or enjoying certain services, e.g. in the field of finance or insurance [527], or in a more intensive form as part of a social scoring system [528]. In other words, data is assigned to individuals, on the basis of which future consequences may be incurred. If the information received and processed by such entities is incorrect or misleading, and the informational agent is incapable of correcting for this, it's consequent decisions and their impact on an individual's autonomy remain fuelled by misinformation. The previously described models classifies these structural processes as relational-factual autonomy constraints by imposing real-life limits onto the action potential of a person.

10.7 Conclusion

This section has dealt first and foremost with answering research sub-question 2.4, namely what the typical patterns and phenomena are that characterize risk to an individuals autonomy that are mainly attributable to the sphere of the information recipient. This section has, through recourse to practical examples, identified four such factors: the medium of information transfer and its characteristics, the perceived accuracy, precision, trustworthiness, and assertiveness of information, transparency to the user with respect to the information collection process of the agent and alternative action potentials available to the user (where applicable), and the assignation of actionable data to a human user.

With the analysis of vectors of influence in the context of information receival we have also finished the systematic appraisal of phenomena that affect the impartment of persuasive qualities. As a result research question 2, which served as the overarching guidance for Part II of this thesis, has been sufficiently addressed as well. Information technology plays a critical role in enabling and facilitating autonomy constraints, and it does so in three different sets of vectors of influence. Domain-specific, agent-specific and information receiver-specific particularities all play a role on how autonomy constraints can be imparted. Differences within these vectors of influence have the potential to affect information flow and the imparting of persuasion on such information flow in a major way (insofar they are mutable at all). As some of these vectors are becoming more powerful (e.g. through increased connected device

[525] Naturally, with each causal link the impact that may be prescribed to the initial phenomenon of misinformation becomes more tenuous. This is particularly true if the causal links are themselves individuals capable of autonomous actions, as their ethical responsibility may take precedence.
[526] Muhammad Bilal and others, 'Social Profiling: A Review, Taxonomy, and Challenges', *Cyberpsychology, Behaviour, and Social Networking*, 22.7 (2019), 433–50 <https://doi.org/10.1089/cyber.2018.0670>.
[527] (see e.g. Cevolini & Esposito, 2020)
[528] Ulrich Hoffrage and Julian N. Marewski, 'Social Scoring Als Mensch-System-Interaktion', in *Social Credit Rating* (Wiesbaden: Springer Fachmedien Wiesbaden, 2020), pp. 305–29 <https://doi.org/10.1007/978-3-658-29653-7_17>.

prevalence, stronger analytics or decreased scepticism and privacy preferences), the risk of an individual experiencing a constraint to their ability to exercise autonomy increases.

All previously discussed types of vectors of influence stand in close connection to each other. Domain characteristics are reflected in the characteristics we may ascribe to informational agents, and vectors of influence attributable to an information recipient's sphere are present because of the interacting informational agent all the same. It remains important to consider the distinction between these types of vectors only as a first assignment and not a hard line. While their distinction is useful to divide relevant characteristics between different sets of phenomena, and they can give a good approximation for how to attribute autonomy constraints as outlined in Section 11.3.2.1, they remain strongly interconnected. To this end, the below figure visualizes the amalgamation of factors previously identified.

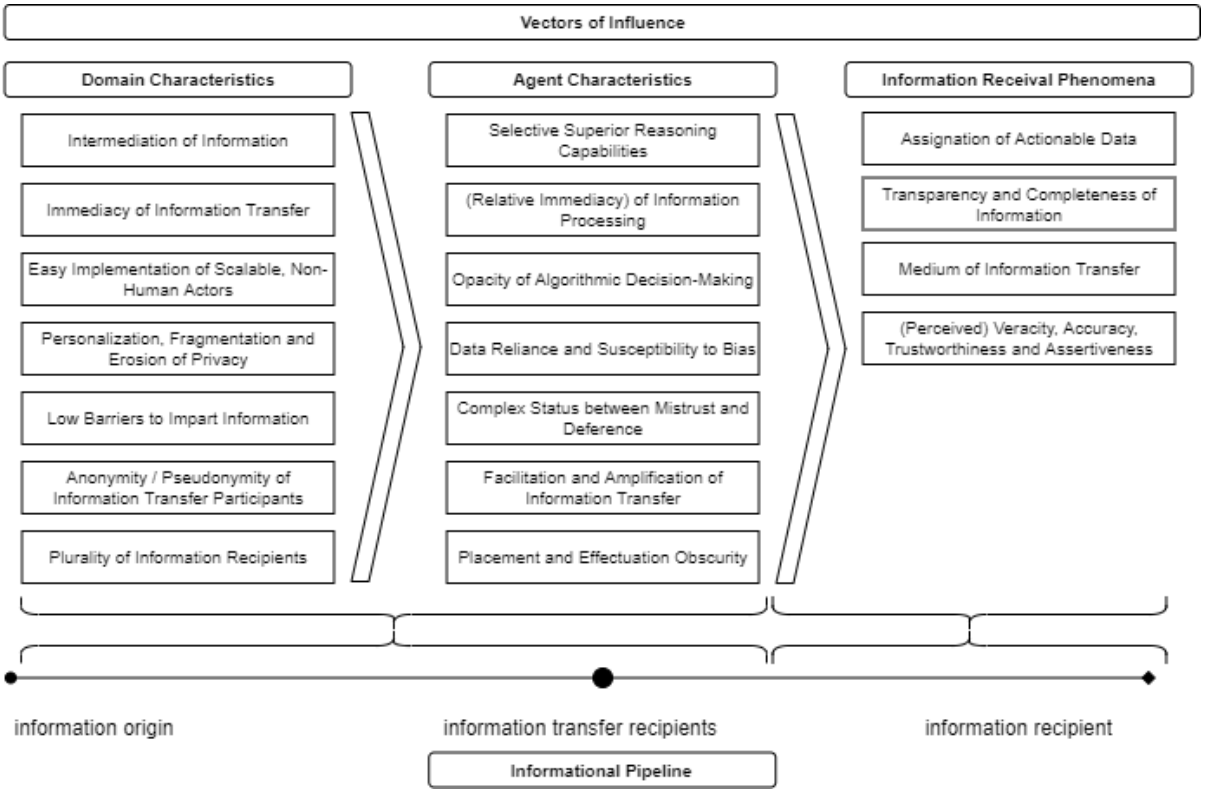

Figure 8 - Overview over Vectors of Influence

Part III: Regulation & Ethics

The final part of this thesis deals with the operationalisation of the conceptual superstructures previously outlined, namely the concepts of the pragmatic account of autonomy and the vectors of influence and investigates (further) the implications of technology-derived autonomy constraints from an ethical and legal viewpoint. Previous sections have already touched upon both of these briefly. Ethical concerns have already been reflected in many of the existing elaborations. Ethical intuition serves as the main motivation behind this research and the utility of the previous theories will have to be measured in how they help understand and address autonomy concerns. The legal domain was also one of the explicit benchmarks to inform and test plausibility against when outlining the pragmatic account of autonomy in Section 4 and 5.

Section 1.6 has outlined that there are two main goals of autonomy research. The primary goal is to understand or conceptualize autonomy for its own sake, while the secondary goal is to apply this knowledge to protect autonomy where appropriate and needed. While the previous research has been first and foremost in service of that primary objective, this section shows their utility by contributing to the secondary objective. Section 11 deals more intensively with the ethical implications of autonomy-constraining autonomy. The section outlines in detail how the pragmatic account of autonomy and the concept of vectors of influence can aid in establishing breaches of ethical obligations and their attribution. Section 12 expands on Section 4's analysis of autonomy in the legal domain, and traces the shift of autonomy recognition and protection in the European Data Laws. The analysis finds that there is a perceptible change from process-driven informational privacy protection to explicit recognition of decisional privacy and hence autonomy. Finally, Section 13 considers how the research of this thesis may be applicable in different cultural contexts and what type of implications transcultural deployment of autonomy-relevant technology may have.

11 Ethical Implications of Autonomy-Constraining Technology

11.1 Introduction

The relevance of the research undertaken within this thesis originates partly from (intuitive) ethical concerns that have arisen out of the risks perceived to be connected with information technology. As was already mentioned in Section 1, there has been significant concern voiced by public institutions, watchdogs, stakeholders, political bodies and academics alike. While the first two parts of this thesis have taken this as a justification to suggest an overarching account of what we ought to understand when we bemoan the risk of technology to an individual autonomy, this section aims to engage more closely with the ethical issues raised in view of the framework that has been established previously. While not the main focus of this thesis, touching upon the implications the existing analysis has on ethical analysis of autonomy-constraining autonomy helps show the utility of the previous research and puts it into some context with the main intuitive concerns that have motivated the writing of this text to begin with.

The fundamental assumption when considering the ethical impact of technology on an individual's autonomy is the belief that autonomy by itself is valuable and generally worth

protecting; this has been referred throughout the thesis as the omni-relevance of autonomy. Within the first part of this thesis, this assumption has been made plausible in particular by references to legal rules that each aim to protect some aspects of what we ought to understand by the concept of autonomy.[529] In addition to this, we can also assume plausibility of this statement through reference to the works of established researchers in the field of autonomy research and ethics in general: Many theorists that have engaged with the question of moral relevance of autonomy suggest that autonomy is indeed morally relevant and ought to be preserved.[530] When it comes to the operationalisation of such concepts (in the domain of technology), again there seems to be widespread concurrence; e.g. many of the existing proposed ethical frameworks for artificial intelligence consider autonomy (and freedom) as a core value.[531] Finally, there seems to be pragmatic evidence that autonomy (in some form or another), and the self-perception of exercising it, is actually beneficial to the individual themselves. As pointed out by André et al, individuals perceiving themselves to acting autonomous consider this favourable; for example they may experience utility from being able to self-attribute the results of their choices, perceiving themselves as exercising agency and subsequently deriving meaning from both.[532] From a methodological perspective we may thus validly assume that the capacity of an individual to exercise autonomy and its corresponding exercise itself is valuable and worth safeguarding. This generally allows the baseline assumption that "more autonomy is better than less autonomy" everything else being equal and barring exceptional edge cases.[533]

These considerations are embedded in a wider context related to the ethical analysis of technology and the ethical analysis of information. Typical lines of inquiries in this context start with considering technology as an extension of human behaviour and tend to denote it as a "tool", and then add additional layers of complexity to account for the interdependencies between technology, individuals and society at large.[534] Typical research in this domain is then

[529] The same naturally holds true when accounting for a narrower conception of autonomy, as the required findings for plausibility would become narrower as well.

[530] See e.g. Young 1982; Dworkin 1988; Darwall 2006; Sneddon 2013.

[531] Anna Jobin, Marcello Ienca, and Effy Vayena, 'The Global Landscape of AI Ethics Guidelines', *Nature Machine Intelligence*, 1.9 (2019), 389–99. See for a discussion in context of the European Strategy for Data also 12.

[532] André and others.

[533] Note e.g. some of the instances introduced in the below Section 13.2.3, as well as the notion of negative impacts of exercising autonomy, as elaborated by André and others. In particular, André et al identify trade-off conflicts, option attachment and guilt incurred from choices between multiple non-ideal options as sources of concern, e.g. with reference to the following research: Steven M. Shugan, 'The Cost of Thinking', *Journal of Consumer Research*, 7.2 (1980), 99 <https://doi.org/10.1086/208799>; Ravi Dhar and Klaus Wertenbroch, 'Self-Signaling and the Costs and Benefits of Temptation in Consumer Choice', *Journal of Marketing Research*, 49.1 (2012), 15–25 <https://doi.org/10.1509/jmr.10.0490>; Ziv Carmon, Klaus Wertenbroch, and Marcel Zeelenberg, 'Option Attachment: When Deliberating Makes Choosing Feel like Losing', *Journal of Consumer Research*, 30.1 (2003), 15–29 <https://doi.org/10.1086/374701>. Another highly relevant case identified is choice overload, which conforms closely with the concept of information overload identified in the pragmatic account of autonomy in Section 5.4.3.5, see e.g. Sheena S. Iyengar and Mark R. Lepper, 'When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?', *Journal of Personality and Social Psychology*, 79.6 (2000), 995–1006 <https://doi.org/10.1037/0022-3514.79.6.995>; Scheibehenne, Greifeneder, and Todd; Chernev, Böckenholt, and Goodman.

[534] See e.g. Ronald L. Sandler, 'Introduction: Technology and Ethics', in *Ethics and Emerging Technologies* (London: Palgrave Macmillan UK, 2014), pp. 1–23 <https://doi.org/10.1057/9781137349088_1>. For a more poignant case study (including analogue technology such as the stirrup (!) see e.g. John Danaher and Henrik Skaug Sætra, 'Technology and Moral Change: The Transformation of Truth and Trust', *Ethics and Information Technology*, 24.3 (2022), 35 <https://doi.org/10.1007/s10676-022-09661-y>. For insights into the interdependency between technology and moral systems see in particular the short book Peter-Paul Verbeek,

often concerned either with the ethical or moral quality of certain types of technology or with matters of concern that arise out of the current or potential future use of technology.[535]

The current section relates more closely to the second of these lines of inquiries, albeit recognizing that these cannot fully be separated.[536] The question of the moral or ethical quality of an informational agent[537] and the technical processes that encompass any relevant vectors of influence are not of direct concern here. Instead this section focuses on the resulting situations which trigger, as previously mentioned, intuitive concerns. However, this section does not aim to provide exhaustive ethical analysis of autonomy constraints. Instead, it complements the existing abstract frameworks with certain principal ethical assumptions, highlights the most important areas of concern that may arise out of their application and the peculiarities of autonomy as conceptualized previously and provides context for ethical analysis on the basis of whichever framework may be seen as appropriate.[538]

This section aims to answer research sub-question 3.1, namely how and under consideration of which factors an ethical analysis of technology-imposed autonomy constraints should be conducted and to what extent the pragmatic account of autonomy can be useful in this context. The analysis in this section proceeds as follows. Section 11.2 outlines key concepts and terms that are needed to complement the previous findings for the purposes of ethical analysis and outlines a general gradient of ethical permissiveness of autonomy constraints. Section 11.3 investigates particular attachment points of the pragmatic account of autonomy and the previously outlined vectors of influence. Finally, Section 11.4 elaborates on some other ethical implications that arise out of the matters discussed within this thesis.

11.2 Prescriptive Prerequisites

11.2.1 Ethic Neutrality of the Pragmatic Account of Autonomy?

We can use the findings of Section 5 and apply the pragmatic account of autonomy to a situation to determine the types of autonomy constraints imposed onto an individual. But while

Moralizing Technology: Understanding and Designing the Morality of Things (University of Chicago press, 2011). One of the interesting criticisms against viewing technology as a mere tool is that the decision of non-use of the technology or reverting to a stage without a given technology may in some cases be simply not (practically) possible, see Jonas Hans, 'Technology as a Subject for Ethics', *Social Research*, 49.4 (1982), 891–98 <http://www.jstor.org/stable/40971222>.

[535] In a way this means that technology "unearths" dormant problems that may have always be considered as ethically problematic but were just not part of any relevant discourse as they only apply if the technology in question is present. In this context, some authors also use the word "create" with respect to the ethical issues arising in connection with technology, see e.g. Richard O. Mason, 'Applying Ethics to Information Technology Issues', *Communications of the ACM*, 38.12 (1995), 55–57 <https://doi.org/10.1145/219663.219681>.

[536] This was already noted and elaborated on in Section 1.6

[537] This is consistent with previous analysis. For example, Section 9.2.7 has briefly outlined why this thesis does not adapt the previously existing conception of "moral agents".

[538] A specific positioning beyond the most intuitive implications seems imprudent at the current stage. Not only is there only limited consensus as to which autonomy constraints are morally impermissible or justifiable, there is little epistemic recourse to validate any posture taken here apart from the plausibility reasoning given already multiple times in this thesis. The development of a particular set of ethical principles is hence left for another time to respect the original scope of this thesis.

this account was somewhat informed by (pseudo-)[539] ethical frameworks that justify its relevance, it does not yield a definite ethical assessment.[540] Consider the following example:

Example 3.7 *C is a non-smoker, partly due to health reasons. After seeing an advertisement on a social media platform for cigarettes that highlights alleged social benefits and status that come with the consumption of cigarettes, they start to consume cigarettes regularly, despite his health-related considerations. Prior to the advertisement, they would not have considered this purchase.*

The pragmatic account of autonomy allows us to identify this as an autonomy constraint that is partly intrinsic (through mental state insertion by a foreign entity with attributable agency) and partly relational (through (pseudo-) societal pressure.[541] However, just because this insertion may create incongruence between the individual's mental states and that they feel social pressure does not quite mean that the constraint is indeed unethical. Consider instead the counterexample:

Example 3.8 *C is a smoker, partly due to (perceived) social benefits and status they believe comes with it. After seeing an advertisement on a social media platform that highlights the health advantages of ceasing smoking and replacing cigarettes with a nicotine patch, they wean of smoking and purchase nicotine patches instead, despite still believing in the social benefits and status awarded by smoking cigarettes. Prior to the advertisement, they would not have considered this purchase.*

While this counter example is very similar, the intuitive judgment becomes less clear. In both cases, a competing preference is inserted that is at odds with the existing preferences that the individual possesses. Nonetheless, there seems to be a qualitative difference: In one example the individual is led to behaviour that is objectively unhealthier than the alternative, while in the other example the opposite is the case.[542] This and similar differences are not always reflected in the pragmatic account of autonomy, indeed they may not be inherent to the very concept of autonomy at all,[543] so the ethical evaluation of such potential autonomy constraints requires recourse to other evaluative frameworks besides the general assumption of autonomy

[539] See Section 1.5

[540] However it would be false to state that the pragmatic account of autonomy is inherently and completely ethically neutral. In particular, the general assumption of omnirelevance of autonomy charges the account as a whole with ethical relevance. In particular, barring any other factors, constraining an individual's autonomy on any of the pragmatic account's dimensions is not ethically neutral when taking the aforementioned assumption. In other words, the pragmatic model clearly divides phenomena in ethically unproblematic and ethically problematic sets.

[541] See Sections 5.4.1.2.1, 5.4.2.3

[542] Adapting the above examples in a way that removes the autonomy constraint by allowing sufficient integration of the inserted preference (e.g. by conscious reflection of the individual) may be considered to ease this tension; clearly once certain behaviour is consciously validated, moderate amount of self-harm or foregoing beneficial outcomes is within the ethically acceptable realm of actions an individual can take (autonomously). Again, this also highlights the advantage of taking a differential view in which autonomy and morality are separated, unlike the Kantian conception that was explored in Section 2.3.1.1. It seems appropriate to describe actions that are sufficiently reflected as autonomous, even if they are deeply unethical.

[543] See e.g. the argument made by Sarah Buss: "…the capacity for autonomous agency—autonomy—is neither the key to justifying moral constraints nor the key to identifying what the content of these constraints must be" in Buss.

omnirelevance. This section outlines some relevant concepts that can inform the ethical evaluation of autonomy constraints.

11.2.2 Inviolable Principles

The concept of individual self-government and its constraints as described in the pragmatic account of autonomy spans a wide range from simple exposure of misinformation over changes to mental states to physical restraints. Within those, most will find that some autonomy constraints are generally ethically unacceptable regardless of any other circumstances.[544] When these principles are translated into the legal domain, they are often denoted as "absolute rights".[545] For example, a process that leads to an individual to be imprisoned for inability to fulfil a contractual obligation is not only a severe autonomy constraint but also contradicts the widely accepted tenet that no one shall be deprived of their liberty in such situations.[546] (Similarly, the autonomy-relevant prohibition of slavery or torture hold such status.)[547] In the context of autonomy in the technological domain, there is increasing attention to mental integrity as inviolable: For example, the draft European Union's Artificial Intelligence Act's first case of prohibited uses of artificial intelligence systems concerns the use of subliminal techniques to "materially distort a person's behaviour […]".[548] Another example for protections that likely reach this level may be the rights to not be unduly decided upon by automated decision-making without human oversight.[549] Intentionally propagating existing system failures of the societal context the technology is embedded in may be yet another example, e.g. in case of an autonomy constraint knowingly or intentionally imposed on the basis of algorithmic bias.

Holding on to this legal lens, there are also principles that do not fully reach the level of absolute rights but are considered as fundamental nonetheless.[550] As previously discussed, legal sources do not always translate neatly into the field of ethics;[551] however as elaborated on the basis of the "normative ladder", instruments of fundamental rights are one plausible way to determine

[544] Inviolability of certain aspects of an individual autonomy may have further gradients still. In particular certain inviolabilities may actually contradict each other, see e.g. the arguments made by Linda C McClain, 'Inviolability and Privacy: The Castle, the Sanctuary, and the Body Symposium: The Sacred Body in Law and Literature', *Yale Journal of Law & the Humanities*, 7.1 (1995), 195–242 <https://heinonline.org/HOL/P?h=hein.journals/yallh7&i=207>.

[545] See e.g. Natasa Mavronicola, 'What Is an "Absolute Right"? Deciphering Absoluteness in the Context of Article 3 of the European Convention on Human Rights', *Human Rights Law Review*, 12.4 (2012), 723–58.

[546] See e.g. Article 11 ICCPR or Article 1, Protocol 4 to the Convention for the Protection of Human Rights and Fundamental Freedoms securing certain rights and freedoms other than those already included in the Convention and in the First Protocol thereto

[547] Note that both of these also tend to come up in theoretical discussion about autonomy at large, e.g. the example of slavery and the potentiality of consenting to it often serves as the prime example to distinguish between procedural and substantive autonomy, see Section 2.3.1.

[548] Art 5 para. 1 lit a, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, 2021/0106(COD), hereinafter AI Act.

[549] See e.g. Art. 22 GDPR. In a wider sense, this also applies to the military use of autonomous weapons, as the automated decision-making without human oversight with potentially lethal consequences triggers numerous ethical concerns that exceed the ethical concerns connected to weapon use generally, see e.g. Markus Wagner, 'The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems', *Vand. J. Transnat'l L.*, 47 (2014), 1371.

[550] See for a more in-depth discussion of some of these that also served as methodological justification for the pragmatic account for autonomy Section 4.7.1.3.

[551] See Section 5.3.

overarching and most-important ethical concerns. [552] Nonetheless, there may very well be actions that are not covered by a legal umbrella of protection that nonetheless are invariably unethical.[553]

Such principles serve as the most inflexible source of ethical information to evaluate autonomy constraints. Autonomy constraints that violate such fundamental principles are always unethical, and as such no justification is available.[554] As to which principles should inform such analysis, there remains much to be written still.

11.2.3 Justified Restrictions

Below the level of inviolable principles, which denote autonomy constraints as generally unacceptable are situations in which autonomy constraints may be justified. Here we can distinguish between different types of justification that the constraining entity may be entitled to. The first distinction we may take is in which interest the autonomy constraint is attempted: Is an individual's autonomy constrained in their interest or is it in service of another entity's interest? The first of these cases relates mainly to the concept of paternalism, the second relates to the process of balancing competing interests.[555]

11.2.3.1 Paternalistic Autonomy Constraints

One of the most important concepts that modulate the ethics of exercising and constraining autonomy relate to the concept of paternalism. In short, paternalism describes measures taken against the will of the affected individual based on the motivation that these measures are in the interest of said individual.[556] Theorists distinguish between different types of paternalism, e.g. with respect to what type of coercion is used,[557] what weight to give the preferences of the individual affected,[558] or the (non-)identity of the affected individual and the protected

[552] See Section 1.5. Note that the ranking of ethical principles in its detail tends to be contentious, see for a similar discussion in practical contexts e.g. Donna Harrington and Ralph Dolgoff, 'Hierarchies of Ethical Principles for Ethical Decision Making in Social Work', *Ethics and Social Welfare*, 2.2 (2008), 183–96 <https://doi.org/10.1080/17496530802117680>; Thomas Hadjistavropoulos and David Cruise Malloy, 'Ethical Principles of the American Psychological Association: An Argument for Philosophical and Practical Ranking', *Ethics & Behaviour*, 9.2 (1999), 127–40 <https://doi.org/10.1207/s15327019eb0902_4>.

[553] See for example the discussion in Thomas E Hill Jr, 'Autonomy and Benevolent Lies', *J. Value Inquiry*, 18 (1984), 251.

[554] A different question is the matter of exculpation. From a legal standpoint justification privileges a concrete action (or phenomenon) as legal, whereas exculpation only privileges an individual to be shielded from ordinary (legal) consequences, see for some insights into this debate e.g. Andrew Botterell, 'A Primer on the Distinction between Justification and Excuse', *Philosophy Compass*, 4.1 (2009), 172–96; Kent Greenawalt, 'The Perplexing Borders of Justification and Excuse', *Columbia Law Review*, 84.8 (1984), 1897–1927; Albin Eser, 'Justification and Excuse Symposium: The New German Penal Code', *American Journal of Comparative Law*, 24.4 (1976), 621–37 <https://heinonline.org/HOL/P?h=hein.journals/amcomp24&i=629>. Note that not everyone appreciates the extension of this concept from the domain of (criminal) law to the domain of philosophy, see e.g. Marcia Baron, 'Is Justification (Somehow) Prior to Excuse? A Reply to Douglas Husak', *Law and Philosophy*, 24.6 (2005), 595–609 <https://doi.org/10.1007/s10982-005-0842-y>.

[555] Note that paternalistic constraints to may require a balance of interest-test.

[556] See for this and the following distinctions e.g. Gerald Dworkin, 'Paternalism', *The Stanford Encyclopedia of Philosophy*, 2020 <https://plato.stanford.edu/archives/fall2020/entries/paternalism/>.

[557] Broad or Narrow paternalism, whereas narrow paternalists are concerned only with legal coercion.

[558] Weak or strong paternalism, whereas weak paternalists consider interferences legitimate, if the affected individual is merely hindered of choosing actions that would counteract their own preferences. Similar to this is the distinction between hard or soft paternalism, whereas soft paternalism only allows interference to determine that the affected individual is acting with appropriate knowledge and out of their own will but with no requirements to what the action ought to be.

individual(s).[559] The recent years have also brought a specific type of paternalism to public attention and prominence. In the context of so-called nudging,[560] Sunstein and Thaler have proposed a concept of libertarian paternalism, in which measures taken have the potential to alter the affected individual's behaviour but do so without any type of coercion.[561]

Paternalism introduces a layer of ethical evaluation that fits well with the inquiry at hand.[562] In particular, adoption of some paternalistic stance allows to create exemptions to the previously outlined base assumption of autonomy being an ethically valuable state if the autonomy-constraining measure is justified by an overarching interest either in the interest of the constrained individual or other relevant entities. (In practice, some sort of paternalism is immanent in current societies and codified in law.[563]) This evaluation ultimately requires a balancing test between the individual's (competing) interests.

11.2.3.2 Externally Motivated Autonomy Constraints

Whereas paternalistic autonomy constraints are imposed in the interest of the constrained individual, there are also autonomy constraints that are done in the interest of the constraining entity or a third party altogether. Evaluating these cases requires a careful balancing of interests.[564] In cases like this, there are two possibilities: First, the constraining entity may do

[559] Pure or impure paternalism, whereas pure paternalism means that only the individual that ought to be protected is also the individual that is affected by any taken paternalistic measures.
[560] See Section 6.4.2
[561] See e.g. Thaler, Sunstein, and Balz; Sunstein, 'Nudging: A Very Short Guide'. The terminology "libertarian paternalism" comes from the characteristics of the approach of preserving freedom of choice but remaining focused on the interest of the individual affected, see Dworkin, 'Paternalism'; Cass R Sunstein and Richard H Thaler, 'Libertarian Paternalism Is Not an Oxymoron', *The University of Chicago Law Review*, 2003, 1159–1202.
[562] Many of the existing debates around the peculiarities of paternalism also translate into the domain of autonomy. For example, discussion about transparency of paternalistic measures connect well to the concept of autonomy-specific informational constraints outlined in Section [5.4.3.1]. Similarly, the conflict of hard and soft paternalism mirrors the debate around procedural or substantive autonomy in Section […].
[563] A full discussion would exceed the scope of this thesis. Suffice it to say that many widely accepted legal norms (e.g. often quoted seatbelt laws) as well as some controversial ones (e.g. limits to consenting to medical procedures) are encoded in legislation around the (western) world. See e.g. Matthew McCoy, 'Autonomy, Consent, and Medical Paternalism: Legal Issues in Medical Intervention', *The Journal of Alternative and Complementary Medicine*, 14.6 (2008), 785–92; Joel Feinberg, 'Legal Paternalism', *Canadian Journal of Philosophy*, 1.1 (1971), 105–24; Anthony T Kronman, 'Paternalism and the Law of Contracts', *Yale Lj*, 92 (1982), 763.
[564] The need for balancing interests in the context of ethics is particularly developed in the context of bioethics, see e.g. JOSEPH DEMARCO and PAUL FORD, 'Balancing in Ethical Deliberation: Superior to Specification and Casuistry', *The Journal of Medicine and Philosophy*, 31.5 (2006), 483–97 <https://doi.org/10.1080/03605310600912675>; N Hallowell, 'Balancing Autonomy and Responsibility: The Ethics of Generating and Disclosing Genetic Information', *Journal of Medical Ethics*, 29.2 (2003), 74–79 <https://doi.org/10.1136/jme.29.2.74>. At the same time balance tests between different (legitimate) interests is a common legal phenomenon, see e.g. for a US-perspective on such doctrine Peter Bachrach, 'The Supreme Court, Civil Liberties, and the Balance of Interest Doctrine', *The Western Political Quarterly*, 14.2 (1961), 391 <https://doi.org/10.2307/443595>; Charles Fried, 'Two Concepts of Interests: Some Reflections on the Supreme Court's Balancing Test', *Harvard Law Review*, 76.4 (1963), 755 <https://doi.org/10.2307/1338701>. Not entirely congruent but of similar importance are the balancing aspects inherent in solving issues that relate to fundamental rights (e.g. through application of proportionality principles), see e.g. Tor-Inge Harbo, 'The Function of the Proportionality Principle in EU Law', *European Law Journal*, 16.2 (2010), 158–85 <https://doi.org/10.1111/j.1468-0386.2009.00502.x>; Wolf Sauter, 'Proportionality in EU Law: A Balancing Act?', *Cambridge Yearbook of European Legal Studies*, 15 (2013), 439–466 <https://doi.org/10.5235/152888713809813611>.

so in its exercise of its own necessity. In particular, the available space of action[565] may not be large enough for the entity to act on its necessity without constraining an individual's autonomy.[566] Second and in contrast to the above, the constraining entity has no necessity to constrain but does so on the basis of "softer" interests. Both of these conflicts require careful consideration and balancing, but the second case will typically require stronger additional justification compared to the first, all things being equal.

11.2.4 Special Cases

There are other cases which the previous categories do not appropriately describe sufficiently. These are conceptually interesting even if not always as relevant (or challenging) for pragmatic ethical analysis. For completeness' sake, they are shortly described hereafter.

11.2.4.1 Collateral and Maleficent Autonomy Constraints

In cases where an entity constrains an individual in their autonomy without taking any utility out of the constraint, a balancing test of interest is hardly useful. We can distinguish based on the general intent of the constraining entity (if such can be attributed). If the autonomy constraint is a side-effect unbeknownst to an activity of the constraining entity (and an alternative activity is possible without significant disadvantages)[567], the constraint is collateral. In case the autonomy constraint, again without expected utility, is imposed intentionally (or perhaps for the mere purpose of imposing the constraint itself),[568] we may consider the autonomy constraint not as collateral but maleficent.[569] In both of these cases, the autonomy constraint is generally unethical.[570]

---

[565] This term as used here is borrowed from related disciplines such as decision theory or artificial intelligence, but is used in a slightly wider context, as these field tend to use the term with respect to a subjective view on the potential actions to be taken (i.e. based on the perception of the agent with respect to possible actions and possible environments).

[566] Many of these situations are practically unproblematic. For example, a software developer of a social media network that also has a profile on the network for debugging purposes and is assigned with a certain unique identifier (e.g. his name), necessarily restricts the space accessible to another's autonomy: for example, a future user of the same name will be unable to use the same unique identifier (perhaps as a nickname), as the space of potential actions (i.e. choosing and holding the identifier on the social media network) is too scarce to accommodate more than one individual to take a certain action. The question if the action space needs to be designed in such a way may however be worthy of follow-up consideration.

[567] If this condition does not apply, then the constraint would be externally motivated, as outlined in the previous section.

[568] In which case one may argue that there is indeed utility derived from the autonomy constraint, but general ethical analysis would broadly discount such interest as invalid in most cases.

[569] Note that much of ethical literature in the field of medical ethics also uses the term "malevolence". The semantic difference between these two terms is negligible in the current context.

[570] All things being equal, the principle of "do no harm" is widely considered as a central tenet of ethical decision-making in abstract and practical analysis alike, see e.g. for its use in a health context (as the Hippocratic injunction) Cedric M. Smith, 'Origin and Uses of Primum Non Nocere -Above All, Do No Harm!', *The Journal of Clinical Pharmacology*, 45.4 (2005), 371–77 <https://doi.org/10.1177/0091270004273680>, in research adjacent to data Maddalena Favaretto and others, 'First Do No Harm: An Exploration of Researchers' Ethics of Conduct in Big Data Behavioural Studies', ed. by Daniel Jeremiah Hurst, *PLOS ONE*, 15.11 (2020), e0241865 <https://doi.org/10.1371/journal.pone.0241865>, or more generally Mill with the following: "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others", in John Stuart Mill, *On Liberty* (Kitchener, Ontario: Batoche Books Limited), p. 13. Further elaboration on this topic is avoided here due to the highly intuitive nature of this principle. Note that this does not necessarily mean that such behaviour would be illegal under a regime, even if it is informed by such ethical considerations.

11.2.4.2 Non-Attributable Environmental Autonomy Constraints

We may recall that the pragmatic account of autonomy focuses on conceiving autonomy based around the capacity of an individual to exercise self-government and how this capacity may be constrained.[571] This entails that autonomy constraints may not only be imposed by other individuals, groups of individuals, other entities or phenomena previously summarised as informational agent but by factors that cannot be attributed to any of the aforementioned. Instead we may consider such autonomy constraints as *environmental* or *external* autonomy constraint. In the context of information technology, clear delineation of environmental autonomy constraints and attributable autonomy constraints are not always easy. As the "digital" environment is (to this point) mostly man-made still, one may argue that no autonomy constraint originating from a digital "environmental" source is possible. However there may be interesting discussion about how man-made e.g. a digital environment ought to be considered, if it is shaped by automatic personalization by e.g. informational agents which are sufficiently opaque to (any) human observer. (This will also be further touched upon below in Section 11.3.1.2 below). At the same time, as different contributions shaping the digital domain compound, the more difficult it becomes to justify any type of appropriate causal attribution beyond the typical "*sine qua non*".[572] With respect to ethical analysis, these types of "background"-constraints seem to rule out most types of ethical responsibility for such constraints that are based on their own contribution. Insofar ethical obligations persist, these must instead be based on a general duty to protect.[573]

11.2.4.3 Other Non-Attributable Autonomy Constraints

We finally turn to non-attributable autonomy constraints that are not environmental *per se* but may still not be attributed to any aforementioned carrier of ethical obligations. In particular, distributed systems of multiple (human) individuals interacting with each other and informational agents, that may obfuscate or sever responsibility through their opacity represent a real challenge when it comes to attributing *ex ante* ethical obligations but also actual *ex post facto* responsibility. Consequently, ethical meta-analyses have pointed this out as a major unresolved area of contention, with no clear consensus in sight.[574]

11.2.4.4 Self-Imposed Autonomy Constraints

As previously identified e.g. in Section 5.4.1.2.3, autonomy constraints may ultimately be derived from the constrained entity themselves. If we can truly attribute the autonomy constraint to the individual, this means that they constrained themselves as an exercise of their autonomy. However, as previously discussed, there are reasons to be worried about such an

[571] See Section 5.2.

[572] The weaknesses of the *condition sine qua non*-formula has been readily discussed in contexts of civil damages / torts and criminal law, as well as in the field of (legal) logic. While a full description would exceed the appropriate scope of this text, the avid reader is referred to the concept of e.g. alternate causality, hypothetical causality, cumulative causality, "overtaking" causality [*German: überholende Kausalität*], as described in e.g. Apostolos Chelidonis, 'Hypothetische Und Überholende Kausalität', *JURA - Juristische Ausbildung*, 43.3 (2021), 227–37 <https://doi.org/10.1515/jura-2020-2649>; Hans J Kahrs, 'Kausalität Und Überholende Kausalität Im Zivilrecht', in *Kausalität Und Überholende Kausalität Im Zivilrecht* (De Gruyter, 2015); Helmut Satzger, 'Kausalität Und Gremienentscheidungen', *JURA – Juristische Ausbildung*, 36.2 (2014) <https://doi.org/10.1515/jura-2014-0028>. For a non-German source see e.g. Uwe Murmann, 'Problems of Causation with Regard to (Potential) Actions of Multiple Protagonists', *Journal of International Criminal Justice*, 12.2 (2014), 283–94.

[573] Compare e.g. the notion of balancing benefits and obligations outlined in FN 183.

[574] Mittelstadt and others; Tsamados and others.

approach.[575] Ethical evaluation will depend upon the ethicality granted to actions akin to self-harm.[576]

11.2.5 Interim Summary

The above outlined prescriptive prerequisites create a formal system that maps situations in which autonomy is constrained to a type of ethical risk. The aim of the above was to outline the general structure to which ethical assumptions (or values) must be grouped when attempting to apply them to autonomy constraints. In summary and as outlined in the figure below, depending on the type of risk, the autonomy constraint may be justifiable ethically (such as is the case in paternalistically or externally motivated autonomy constraints) or not (such as is the case if the autonomy constraints intersect with the aforementioned inviolable principles or if the autonomy constraint is).[577] Naturally, the outcome of any ethical analysis will depend on the base set of ethical assumptions and values taken, the justified selection of which is the subject of ongoing debate, particularly in the field of AI ethics.[578]

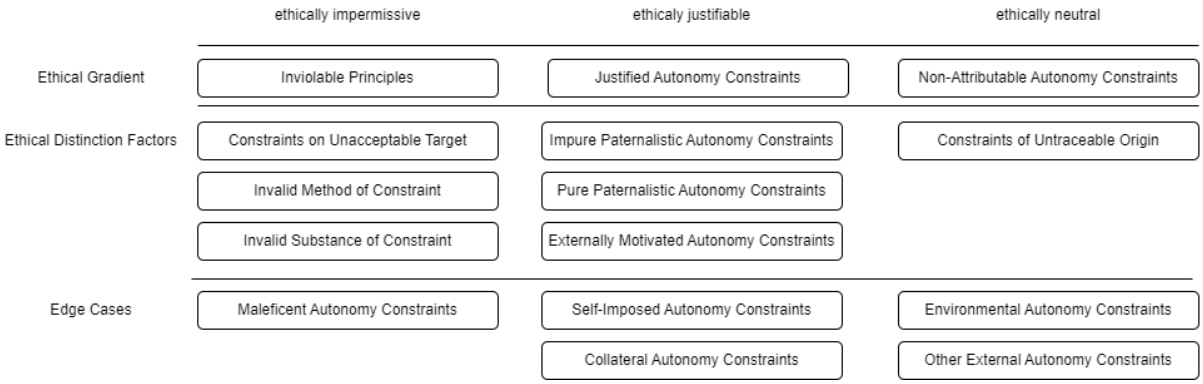

Figure 9 - Prescriptive Prerequisites

11.3 Applying the Ethical Gradient

Having outlined the principles of relevant ethical frameworks in the previous sections, we can now consider how they interact with the previous analysis and conceptions of autonomy and autonomy-relevant phenomena.

11.3.1 The Pragmatic Account of Autonomy Revisited

As outlined in Section 5, within this thesis autonomy is described as freedom of constraints on at least three major axes of observation. Intrinsic autonomy describes the coherence of mental states such as preferences and their reflection in actions and decision taken. Relational autonomy describes the position of the individual in relation to physical, normative and societal phenomena. Informational autonomy describes an individual's actual access to information including how this information is structured. This section considers (briefly and non-

[575] See e.g. Section 2.3.1. The issue here again is to which extent individuals may forego future autonomy.
[576] Joel Feinberg, *The Moral Limits of the Criminal Law Volume 3: Harm to Self* (Oxford University Press, 1989) <https://doi.org/10.1093/0195059239.001.0001>.
[577] We may note here that this type of rough ethical grid merely represents a gradient dependent on ethical base values. While this thesis adopts some of such values (e.g. the "do-no-harm principle") in this section, the methodology generally is flexible with respect to the ethical base values.
[578] As noted in FN 538, this thesis does not take an explicit stance on underlying value sets, as doing so would exceed the scope of inquiry significantly.

exhaustively) some of the peculiarities of ethical analysis of constraints on these dimensions of observation.

11.3.1.1 Some Ethical Considerations on Mental Coherence

We may recall that intrinsic autonomy is constrained by an incongruence of mental states generally, and by decisions and actions made that are in congruence with mental states in particular. Importantly, the pragmatic account of autonomy considers mental states as interdependent and connected with potentially hierarchical connotations.[579] Strongly embedded, higher-order mental states form a sort of personal baseline around which new mental states can be compared and evaluated. Decisions that are made in accordance with a mental state that is not covered by this validating mesh of higher-order mental states suffer from reduced autonomy. However, the reduced autonomy is not merely visible in the decision or action taken but in the incongruence that invalidates the decision or action in the first place. Hence, as outlined, typical autonomy constraining phenomena typically involve the insertion of a foreign mental state or the dislocation of existing mental states.

The ethical quality of such measures will generally depend on the two factors.[580] First, the substantive content of the constraining measure seems relevant. As intrinsic autonomy constraints may not just constrain autonomy by stopping an individual from certain actions but also "persuade" them to take others, the actions (potentially) resulting from the autonomy constraint reflect back on the ethicality of the constraint itself. Consider the following example below.

Example 6.1	*F is exposed to persuasive information, amplified through a social media personalization algorithm, that makes F believe in a conspiracy theory about the workings and crimes of an alleged political cabal. After F is estimated to be appropriately conditioned, members of the conspiracy theory, aware of its shaky epistemic grounds, use the personalization algorithm to imbue a call-to-action in F.*
Example 6.1 (cont. var. a)	*As a result, F becomes convinced that it is necessary for them stage peaceful protests highlighting the perceived issue.*
Example 6.1 (cont. var. b)	*As a result, F becomes convinced that it is necessary for them to use violence against alleged criminals.*
Example 6.1 (cont. var. c)	*As a result, F becomes convinced that it is necessary for them to use violence against innocent bystanders to create public uproar.*
Example 6.1 (cont.)	*As a result, F, generally adverse to any types of violence, truly believes it to be necessary to exercise vigilante justice and acts upon it.*

In this example, the individual in question is not just subject to an informational autonomy constraint (that is the false conspiracy),[581] but also subject to an intrinsic autonomy constraint, which inserts the perceived need for actions that are incongruent with the individual's wider value set. The example variants modulate the ethicality of the persuasion objective. Staging a

[579] See in detail for this Section 5.4.1

[580] Some of the below mirrors existing debate particularly with respect to the ethics of persuasion or advertising. See e.g. the already mentioned Andrew Sneddon, 'Advertising and Deep Autonomy', *Journal of Business Ethics*, 33.1 (2001), 15–28; Spahn; Maclean. However note that this section is meant merely as a cursory overview in connection with previous findings and not as an exhaustive ethical analysis generally.

[581] Inspired by the infamous Pizza Gate conspiracy theory, see e.g. Marc Fisher, John Woodrow Cox, and Peter Hermann, 'Pizzagate: From Rumor, to Hashtag, to Gunfire in DC', *Washington Post*, 6 (2016), 8410–15.

protest is likely ethically unproblematic, but active violent vigilantism is at least problematic.[582] The final variant presents an action or decision that is clearly unethical irrespective of the factuality of the underlying beliefs. It seems clear then that part of the ethical validity of "persuasion", i.e. the mental state insertion", relies on the permissiveness of the intended outcome. Imposing an autonomy constraint to result in an unambiguously ethically invalid action is clearly more significant than trying to effect an ethically neutral action (albeit on false pretences).

Comparing the first and last variants also highlights another factor: subjectivity. In all of the variants, the constrained individual is convinced that their actions are justified (although the final variant makes this an ethically invalid assessment), but this justification is subjective. The constraining entity, in knowledge of the lack of veracity of the motivating factors, can hardly be exculpated in the first variants, as the instigators still used manipulative means (the aforementioned informational autonomy constraint), just because the resulting action was in itself not ethically problematic.

The second determinant for the ethical quality of an autonomy constraint is its method. Consider the following variant of the examples before:

Example 6.2	*F is exposed to persuasive information, amplified through a social media personalization algorithm, that makes F believe in a conspiracy theory about the workings and crimes of an alleged political cabal.*
Example 6.2 (cont. var. a)	*The algorithm presents F with articles of scholarly nature including references supporting the conspiracy theory but allowing F to fact-check its claims.*
Example 6.2 (cont. var. b)	*The algorithm, based on knowledge extracted from profiling, presents F with an assortment of media most likely to elicit approval and subvert active critical engagement.*
Example 6.2 (cont. var. c)	*The algorithm uses highly effective subliminal messages to anchor the respective beliefs in F and completely circumvent any possibility of reflexive conscious reflection by F.*

The exercise of autonomy typically means choosing between different options within a potential space of action. This also includes introspective reflection about the values and factual representation an individual considers adopting explicitly or implicitly. However, intrinsic autonomy constraints can run the gamut from overt attempts of persuasion, allowing critical engagement and subsequent validation of a potentially inserted mental state, to covert or undetectable measures. As the latter counteract the very nature of autonomy exercise, that is to be able to choose (including about the foundations of future choices), it already runs against the basic assumption of autonomy omnirelevance. Extrapolating from this example, it is clear that an autonomy constraint mode becomes increasingly ethically problematic the more it circumvents the potential for critical engagement and validation by the individual.[583]

---

[582] Regina Bateson, 'The Politics of Vigilantism', *Comparative Political Studies*, 54.6 (2020), 923–55 <https://doi.org/10.1177/0010414020957692>.

[583] Though this thesis does not go as far as saying that subliminal information exposure ought always to be considered harmful. Indeed, prior autonomy exercise may very well validate future subliminal autonomy constraints. At the same time, paternalistic autonomy constraints (e.g. to combat addiction) may very well be ethically justified. At the same time, covert autonomy constraints will face a considerably higher bar of justification in situations where a balance of interest is necessary. As mentioned before, a very common use of

Connecting this to the ethical gradient outlined before we outline some rough ethical guidelines. With respect to the substantive nature of an intrinsic autonomy constraint, we can find that intending to effect ethically impermissible behaviour through intrinsic constraints also renders the constraining action impermissible. In other cases, ethical evaluation must be done by considering the mode of the potential constraint and the opportunity for the affected individual to engage with, validate or reject the information that leads to the constraint. Within the range of ethical justifiability, such as paternalistic intrinsic autonomy constraints, the balance test outlined prior must also take the mode into account and heightens the burden of justifiability.

11.3.1.2 Some Ethical Considerations on Relational Capacity

Relational autonomy is characterized by an individual's relation to factual, normative and societal phenomena and their capacity to act as constraints. Out of this conception, multiple attachment points for ethical considerations become available.

First and most obviously, Section 5.4.2 has already outlined a distinctive factor for characterizing such constraints, namely repellency or actual permissiveness. To recall, repellency describes the allowance of an individual within their potential space of action to overcome the constraint. In the (extreme) case of no potential action available to the autonomy, the constraint is absolutely repellent (or respectively impermissive). This translates rather easily into the domain of ethical analysis: In general, when imposing constraints, higher levels of repellency will typically require higher level of justification.

However, the distinction between factual, normative and societal phenomena holds further things to consider. This thesis outlines implications of societal constraints (and how this is inherently subjective and dependent on cultural embeddings) in Section 13. Normative autonomy constraints are arguably comparatively independent from technological phenomena. Hence this section will focuses on the implications of technology as may be understood as factual autonomy constraints.

We may recall from Section 5.4.2.1 that factual constrains, serving as a sort of catch-all, represent manifested circumstances that function as obstacles to the potential of the effectuation of one's (already formulated) mental state. In the context of technology this clearly accounts for physical obstacles imposed within the reasonable scope of causality of technology.[584] More interesting issues arise when considering that information technology does not only serve as a tool to extend one's scope of action within the "real" or "physical" world, but also creates a domain (and a space of action and decision making) in itself. However, as opposed to the "offline" domain, this action space is not merely coincidentally environmental but (at least to a degree) intentionally formed. To elaborate on this, we may consider the following example.

Example 7.1 *G is looking to stay in contact with their friends. Of particular interest to them are live interactions in which G is able to see and hear their friends.*

such (and requiring a balance of interest) is online advertising, see e.g. Meral Elci and Arzu Sert, 'Subliminal Advertising and Its Ethical Dimensions in the Social Media Age', in *Handbook of Research on Effective Advertising Strategies in the Social Media Age* (IGI Global, 2015), pp. 368–87.

[584] See e.g. Example 2.4 in Section 5.4.2.1.

In general, it is easy to see technology in the above example as an extensions of the available action space. The individual in question can stay in contact with their friends (and thus act according to their preferences) through meeting up with them in their offline life, or use technology to facilitate similar conversations.

Example 7.1 (cont.) *To this end, G and their friends sign up to messenger and communication service.*

Looking at the situation from this perspective, technology clearly widens the action space (i.e. gives more potential actions to the individual to act in line with their preferences to communicate with their friends). Of course, the use of this type of service may come with restrictions or not be as enabling as it could be.

Example 7.1 (cont.) *The messenger and communication service allows for video chats of up to 3 people, but no texting or other communication of any sorts.*

Nonetheless, the initial assessment may still stand: Even a smaller than possible extension of one's space of action is still an extension. However, some issues arise in cases where the additionality of the digital domain to the existing "offline" domain is practically displaced.[585]

Example 7.2 *G is looking to resolve a conflict connected to a flight cancellation. The only type of communicate avenue available to them is a web-interface with interdependent multiple layers (questionnaire, chat-bot, service agent).*

In this example, the action space is (realistically) compressed to a solely digital domain. However, this digital interaction environment (i.e. the contact website) has been intentionally shaped and allows only certain input. In other words, because is the action space limited generally, all exercise of autonomy is compressed and mapped to whatever the options under this action space are on the basis of the intention. We can see the differences when varying the above example slightly.

Example 7.2 (cont. var. a) *The questionnaire and chat-bot only allow a limited number of types of input.*

Example 7.2 (cont. var. b) *The questionnaire and chat-bot allow for open-ended input.*

In these variants, the digital environment is still present, and at large communication with the service provider is still compressed. However, in the first example, the compression of action is larger, and the options are more limited. Despite the technology generally serving as an extension of an individual's action space, the fact that their recourse with respect to a specific set of actions is limited to the digital domain and furthermore that set of action being specifically designed as a limited set means that there seems to be a loss of autonomy rather than a gain. This is alleviated to an extent in the second example.

[585] As noted by Floridi, the distinction between online and offline domains is becoming harder to draw; he uses the term "on-life" to describe this phenomenon, see e.g. Luciano Floridi, *The Onlife Manifesto: Being Human in a Hyperconnected Era* (Springer Nature, 2015), passim. Within more comparative studies or in the context of researching attitudes or compliance, this phenomena is also sometimes denoted as mandatory technology adoption, see e.g. Poey Chin Lai, 'The Literature Review of Technology Adoption Models and Theories for the Novelty Technology', *JISTEM-Journal of Information Systems and Technology Management*, 14 (2017), 21–38.

It is noteworthy that there seems to be no circumvention of this autonomy-compression phenomenon when considering information technology. As outlined in Section 8.2.1, one of the main domain characteristics is the (need for) information intermediation. As individuals cannot (yet) interface with digital information directly, some sort of process or artifact must serve as a looking glass into the digital domain, and this looking glass is but a mix of extensions and limitations of an individual's capacity to exercise autonomy. The same holds true about the other layers beyond the first intermediation layer: to access data or processes that are not immediately available (e.g. the cancellation of a flight online), an individual must interact with other processes and do so utilizing a set of action options intentionally constrained.

This observation tracks well with the notion of relational autonomy as outlined previously. Insofar a digital environment's use and presence therein approaches a mandatory nature, its peculiarities, design choices and consequent action space it offers to the individuals therein can themselves be considered as factual autonomy constraints. In contraposition to the previous examples in Section 5.4.2.1, ethically relevant autonomy constraints are not limited to the digital domain directly effecting obstacles in the "offline" domain. Indeed, such autonomy constraints within the digital domain may be at least as ethically relevant insofar as the requirement of displacement is fulfilled.

From an ethical perspective, consideration of these phenomena thus follows the general tenet of autonomy relevance. Insofar as technology is used to compress an action space practically, this is already a "proto"-autonomy constraint, but within this context, the "ethical burden of proof" reverses. To preserve autonomy, more actions are needed by the constraining entity (i.e. the entity to which the space is attributed), as the intentional action of allowing a more widely varied number of inputs is preserving the individual's autonomy after their action space has been compressed. This creates an interesting conundrum. Generally, proactive ethical duties of entities to remove risks or facilitate heightened action potential (as opposed to abstain from harm or constrain) are more difficult to justify. However, given that the entire domain may be considered as flipping from an action potential-extension to a compressed action space that may be attributable to that entity, the administration of this domain creates an unusual requirement that practically may translate into a duty of care and duty of facilitation of sorts, as the digital domain displaces the offline domain for all pragmatic uses.

How may we understand this to connect to the ethical gradient outlined previously? In general, the mandatory and irreversible compression of an individual's capacity for autonomy into a digital domain seems, for many purposes, unacceptable or at least undesirable.[586] The ethicality of effecting autonomy compression therefore requires ethical justification in itself, to which alternative actions or validating factors like consent can contribute. Beyond this, the subsequent balancing of interests must take into account in particular the digital displacement that the respective digital environment has produced, which types of alternative avenues remain for the individual and if the existing limitations within the digital domain allow for a sufficient level

[586] Within the European Union there are many instances in which we can infer a moral intuition from a range of legal instruments ensuring recourse within action-constrained environments in which individual's capacity are compressed. For example, Art. 22 GDPR gives data subjects the rights of not being solely subject to automated decision making (which may be on the basis of data resulting from compressed actions). Similarly, we may see the rise of standardized consumer rights forms (e.g. in the context of flight purchases which are increasingly happening online) as a countermeasure to complaint-interfaces that do not allow for the full range of information necessary to be transmitted for a favourable outcome. Another example is the wide-ranging obligations of website-providers to provide a so called "*impressum*", in which certain contact information is mandated to which (typically, as in the case of e-mails) an inquiry can be expressed within the full range of language.

of action and communication that cover a wide range of typical interactions and consequences for which the digital environment was conceived.

11.3.1.3 Some Ethical Considerations on Informational Capacity

We can finally turn to considering informational autonomy. Again, we may begin by recalling the general outlines of this dimension of autonomy. In general, informational autonomy constraints describe obstacles to the exercise of autonomy imposed through (or by) a certain quality and/or quantity of information or through its presentation.

We may first consider the quality (or veracity) of information. Information that individuals are exposed to may be true, or (partly) untrue, as Examples 1.6 and its variant in Section 5.4.3.4 have shown. When considering this from an ethical perspective, it may also be interesting to control for the knowledge of the information-imparting entity. Between true and (partly) false information and such information being transferred intentionally, knowingly, or unknowingly, there is yet another gradient to consider.[587] We may consider the following examples.

Example 1.7	*A would like to buy a tent. A has been profiled by a customer-analysis algorithm. On the basis of A's profile, a seller aims to advertise a tent to A by emphasizing desired features. A's profile suggests advertising certain features of interest to capture A's attention.*
Example 1.7 (cont. var. a)	*The seller, knowing that the tent in question does not have these features proceed to advertise in such a way regardless.*
Example 1.7 (cont. var. b)	*The seller, not knowing that the tent in question does not have these features because they are not proficient in outdoor equipment, proceeds to advertise in such a way.*
Example 1.7 (cont. var. c)	*The seller, knowing in principal that the tent in question does not have these features but failing to pay attention and hence correct to the advertising-strategy suggestion of the algorithm, proceeds to advertise in such a way.*

In general, using false information to induce or affect an individual's behaviour intentionally, not taking into account any other factors, will typically be difficult to justify, and symbolizes the extreme end of that gradient. Intuitively, the knowing transfer of false information but with no intent for certain autonomy constraints to be imparted by doing so seems to be of lesser severity; ethical implications for unknowingly transferring false information requires a duty to control information veracity in the first place, and as such is of weaker ethical intensity than effecting the other autonomy constraints.[588]

[587] See e.g. for an overview over some of these competing concepts, i.e. misinformation, disinformation, manipulation, deception, missing information, etc. Thomas J Froehlich, 'A Not-So-Brief Account of Current Information Ethics: The Ethics of Ignorance, Missing Information, Misinformation, Disinformation and Other Forms of Deception or Incompetence.', *BiD*, 39, 2017. There is also additional context with respect to mis- and disinformation specifically in one of the Annexes attached at the end of this thesis as well.

[588] This is not to say that failure to control for veracity may not be a breach of an ethical obligation. Indeed, holders of ethical obligations may be required to either possess sufficient knowledge themselves or ensure that their instruments that extend their range of actions (e.g. technology) work within certain substantive or procedural parameters. This obligation should likely depend mostly on the potential risk of possessing, applying or imparting incorrect "knowledge".

With respect to information that is not false, ethical challenges will arise principally in case the information is misleading.[589] Insofar this occurs, the ethical evaluation again depends on the intent as it relates to the potential (and effect) of the misleading property of the information as just outlined above.[590]

Another axis of observation of informational autonomy constraints outlined previously was information prevalence. We may recall that autonomy constraints may be imposed by a lack of information or by information in exceeding amounts, searching and parsing costs of which prevent adequate reception by the individual. Both of these provide different grounds for ethical consideration as opposed to the previously discussed situations in which false (or misleading) information is transferred. Let us consider omission first.[591] With respect to information omission, ethicality depends on the obligation to positively share information. Typically, autonomy constraints by information omission will be accompanied by the transfer of some sort of information.[592] In other words, the autonomy constraint may be attributed not just to omission itself but a set of incomplete information, as in the following example.

Example 1.8	*A would like to buy a tent fit for inclement weather. A has been profiled by a customer-analysis algorithm.*
Example 1.8 (cont. var. a)	*On the basis of A's profile, an algorithm highlights the water-resistant properties of the tents fabric in an advertisement shown to A. However, while the information shown is correct, it omits the fact that the water-resistant fabric is not used all around the tent but merely on selected parts, seriously compromising the usefulness of the tent.*
Example 1.8 (cont. var. b)	*While relevant for A's decision, the algorithm does not provide any information about any type of water-resistant properties of the tent and its fabric.*

This example highlights what is likely a common theme in the ethical consideration of information omission: Omission of counterfactual specific information in the presence of general information is ethically less tenable than complete omission in certain cases. Analogously to the argumentation in the previous sections, the likelihood of ethical obligation to not abstain from autonomy constraints is typically embedded within an initial action (that may not be by itself a constraint).[593]

---

[589] Using truthful information for misleading purposes is also sometimes denoted as "paltering" when discussing ethics of information transfer, see e.g. Todd Rogers and others, 'Artful Paltering: The Risks and Rewards of Using Truthful Statements to Mislead Others.', *Journal of Personality and Social Psychology*, 112.3 (2017), 456–73 <https://doi.org/10.1037/pspi0000081>; Matthew Kopec, 'Deceptive Omissions, Half-Truths, and the Moral Exemplar in Clinical Ethics', *The American Journal of Bioethics*, 21.5 (2021), 33–35 <https://doi.org/10.1080/15265161.2021.1906993>.

[590] Note generally that there is not necessarily consensus when it comes to the weighing of misleading and presenting false information comparatively. Indeed some have suggested that that (intentionally) presenting misleading information presents a more grave ethical failure than than the (intentional) presentation of false information (i.e. lying), see e.g. C. F. Rees, 'Better Lie!', *Analysis*, 74.1 (2014), 59–64 <https://doi.org/10.1093/analys/ant104>; J. Saul, 'Just Go Ahead and Lie', *Analysis*, 72.1 (2012), 3–9 <https://doi.org/10.1093/analys/anr133>.

[591] Compare generally for these concepts e.g. D. Fallis, 'Lying and Omissions', in *The Oxford Handbook of Lying*, ed. by J. Meibuaer (Oxford: Oxford University Press, 2018), pp. 183–92.

[592] In a way this leaves room for classifying a given information transfer either from a perspective of prevalence or veracity. The information that is transferred is rendered misleading by the omission, while the information not transferred may be constraining by itself. A clear separation of these concepts is hence not possible.

[593] Compare to this e.g. to the elaborations about factual autonomy constraints in the context of information technology in the previous section.

On the other end of information prevalence as a characteristic of information exposure is information overload. In these cases, transfer of and exposure to too much information creates a practically diminished level of "informedness" on the side of the recipient.[594] The "overload" may pertain both to contextual information generally, as well as the options that derive out of them, [595] as highlighted in the examples already touched upon in Section 5.4.3.5, some of which are repeated below.

Example 4.4 (rep.)	*D prefers some of his friends not knowing how often he goes out, but one of D's wearables tracks their location and informs D's friends when he is out and travelling.*
Example 4.4 (cont. var. b)	*While there is an option to disable this feature, the documentation is unordered, of excessive length and cannot be parsed via electronic search. As a result, D does not find any information about this option during his limited search because of all the information present and does not disable the feature.*
Example 4.4 (cont. var. c)	*When parsing the documentation, D is hit with a multitude of different options to restrict access to D's personal data presented in parallel by a seemingly never-ending array of privacy-preserving options, whose differences are not immediately apparent in comparison. D is almost paralyzed and has difficulties deciding between their options.*

Insofar as these two different types of information overload ought to be evaluated differently, it seems intuitive that option information overload seems less ethically problematic as opposed to contextual information overload. As in the latter the decision the individual may want to take is not apparent to them, we may compare it (at least in effect) to the previously discussed information omission. On the other hand, a situation with option information overload still presents (valid) decisions points, of which each one may suffice to align with the individual's interest. It seems that only if the paralyzing effect of the option information overload is intentional (or ought to be predicted as part of a duty of risk minimization or similar),[596] that autonomy constraints through option information overload ought to register as an ethically sensitive situation.

Connecting this with the ethical gradient outlined in previous sections, we may thus find that the principal factor determining if an autonomy constraint is ethically impermissible or justifiable depends on the combination of intent, (known) veracity of information and its potential to be perceived in a way that the underlying relevant information becomes accessible to the information recipient.

---

[594] See for this problem e.g. Bawden and Robinson; Persson; Mark R Nelson, 'We Have the Information You Want, but Getting It Will Cost You! Held Hostage by Information Overload.', *XRDS: Crossroads, The ACM Magazine for Students*, 1.1 (1994), 11–15. This is also a particular area of concern when it comes to consumers as users, see e.g Carol Brennan and Martin Coppack, 'Consumer Empowerment: Global Context, UK Strategies and Vulnerable Consumers', *International Journal of Consumer Studies*, 32.4 (2008), 306–13 <https://doi.org/10.1111/j.1470-6431.2007.00640.x>; and in general Omri Ben-Shahar, *More Than You Wanted To Know: The Failure of Mandated Disclosure* (Princeton University Press, 2014) passim.

[595] In other words, we may also speak of a "choice overload" in certain situations, see e.g. Chernev, Böckenholt, and Goodman; Scheibehenne, Greifeneder, and Todd.

[596] This assumes that the options presented are all valid ways to enact the more abstract autonomous decisions made by the individual, as opposed to a multitude of options of which only some are relevant to the individual, in which case we may understand it better as contextual information overload.

11.3.1.4 Interim Summary

When considering the pragmatic account of autonomy outlined in Section 5 in light of prospective ethical analysis and in particular the ethical gradient from impermissive to justifiable autonomy constraints, the distinction between the three types of major autonomy constraints, namely intrinsic, relational and informational, indeed allow for a more granular analysis.[597] When considering intrinsic autonomy constraints, the previous analysis has argued for the importance of the substantive content of the mental state insertion and its method, while relational autonomy constraints call for analysis of their repellency and (in case of factual constraints) the nature of the compression of an individual's action space into the digital domain. Lastly, informational autonomy constraints require close attention to the veracity of information transferred *in combination with* the content-specific intent and knowledge of the constraining entity. Further analytic distinction may be necessary in case of autonomy-relevant prevalence of information (i.e. information overload or omission) particularly connected with if the prevalence affects contextual information or information about an individual's action potential in particular.

As is clearly visible from this cursory overview, ethical analysis of autonomy on the basis of the herein proposed pragmatic framework can build upon the three dimensions of autonomy to produce useful and intuitively satisfying results. At the same time and as stated previously, ethical analysis on this basis requires additional value-laden frameworks to fill in the structure outlined by the prescriptive prerequisites above, and map values and consequent outcomes as impermissible or justifiable, as the pragmatic account of autonomy not an ethical framework itself. Having inserted such a value framework, the peculiarities of each autonomy dimensions outlined in this section ought to be given particular attention and can aid and structure subsequent ethical analysis.

11.3.2 Vectors of Influence Revisited

The second part of this thesis has described the compound phenomena that characterize a given interaction between an informational agent and another entity that is capable of imparting change onto an individual's autonomy.[598] These vectors of influence have been roughly subdivided along a concept denoted here as the informational pipeline, consisting of three somewhat compact and divisible parts: information origin, information processing and information recipient. As a result, this thesis has identified the major characteristics that affect information flow in view of potential autonomy constraints in each of these three areas. Prior to this section, the analysis has been largely descriptive: While the consolidation and systemic appraisal was novel at the time of writing and thus serves one of the goals of autonomy research,[599] it is not yet any operationalization of the full stated axiom that has also motivated this thesis: autonomy is at risk by information technology and is (considered to be) worth protecting. Within this section, the analysis now proceeds towards understanding the ethical implications of the conceptual superstructure that has been introduced through the system of vectors of influence and its distinctive characteristics. Specifically, the notion of vectors of influence add value in two ways in the context of ethical analysis. First, they serve as a first coarse filter for (ethical) attribution of autonomy constraints based on the phenomena that

[597] Note that this was one of the goals of the thesis as outlined as a "secondary" goal of autonomy research in Section 1.6 and as a methodological justification for the conception of the pragmatic account of autonomy in Section 5.2.
[598] See for this definition Section 7.2
[599] See Section 1.2

determine the imparting of influence. Second, they allow us to locate general clusters of (intuitive) concerns connected to certain information processing environments or methods.

11.3.2.1 Presuming Ethical Attribution

We first move to the question of ethical attribution.[600] Considering the distinction of vectors of influence between (1) phenomena mostly arising out of the information domain and its characteristics generally, (2) phenomena mostly arising out of informational agent characteristics and (3) phenomena mostly arising out of the characteristics of information receival, it may be useful to consider if this differentiation allows a first assumption of attribution.

Section 8 has outlined that the vectors of influence arising out of the general domain of information technology can be further distinguished between actual (or physical) domain characteristics (e.g. the changes to device or infrastructure capability) and meta-characteristics that are a side-effect (e.g. intermediation of information) or consequence of increased connectivity enabled by the domain (e.g. immediacy of information transfer, plurality and potential anonymity of information transfer participants or increased personalization). In other words, the vectors of influence associated with the technology domain *per se* are typically fundamental characteristics arising out of the general shape of infrastructure and data processing.[601] The fundamentality of these vectors of influence gives the technology in question and its characteristics an interesting position: The are undoubtedly connected to many autonomy constraints that are imposed through the informational pipeline, as they are necessary building blocks for any technology interaction and serve as enablers for such derived autonomy constraints. On the other side, the chain of causality, while clearly determinable, is stretched to an extreme outcome.[602] This means that any ethical attribution of an autonomy constraint to entities connected to or responsible for these vectors of influence is difficult to establish. Instead, it seems to make more sense to accept the domain characteristics, risk-enabling as they are, as the landscape in which the other vectors are situated and must be considered not just merely by themselves but in connection with the now pre-existing risks of their underlying environment.[603]

---

[600] Note that this thesis uses the term attribution (or responsibility) in a rather colloquial way here. Ultimately, the question that is asked here refers to the general connection and assignation of unethical autonomy constraints to other entities or individuals in a broad way. In particular, this section does not distinguish between any ethical attribution to natural or legal persons at all, but moves on a higher level of abstraction. However, it is worth noting that there is seems to be lively debate about fundamental philosophical questions relating to the identity of or distinction between e.g. attribution, accountability or responsibility, see only Section 3 in Matthew Talbert, 'Moral Responsibility', *The Stanford Encyclopedia of Philosophy*, 2022 <https://plato.stanford.edu/archives/fall2022/entries/moral-responsibility/>.

[601] Considering this in parlance typical for information or network science we may consider that the domain characteristics are generally routed in the lower levels (e.g. the non-application layers in the OSI model, which are physical, data link network, transport, session and (parts of the) presentation layer or network access layer, internet layer and transport layer respectively). See for these classifications e.g. Neil Briscoe, 'Understanding the OSI 7-Layer Model', *PC Network Advisor*, 120.2 (2000), 13–15.

[602] See for the weaknesses of simple approaches for causality also FN 572.

[603] To illustrate this statement, consider the following example. The company maintaining fiber optic cables or the World Wide Web Consortium (W3C) are clearly causally connected to an individual being coaxed into a purchase by misleading advertising to some extent, but it would be indefensible in the author's opinion to make them ethically responsible for such autonomy constraint as any causality is far too removed. Instead, the playing field granted by internet protocols, the increasing processing power of servers on which the advertisement algorithm runs and the improving state-of-the-art catalogue of knowledge regarding personal advertisement targeting is best understood to present the landscape which amplifies certain risks of subsequent technology (e.g.

On the other hand, we have the vectors of influence that relate to the information recipient and the circumstances of their interactions. Here, many of interaction characteristics (e.g. trust, perceived veracity or accuracy, cognitive biases) are primarily located in the sphere of the information recipient. This raises the questions to what extent autonomy constraints utilizing such vectors should be attributed to entities other than the information recipient. In other words, how should ethical obligations be imposed on other entities to consider (and abstain from exploiting) vulnerabilities of the information recipient. Multiple approaches may be considered here. First, it seems that exploiting a vulnerability within the sphere of the information recipient that is generally prevalent (i.e. present in a sufficiently large amount of interaction targets) enough, so that the entity can reasonably expect that exploitation of such vulnerability will occur. For example, using deceptive design to coax potential data subjects to give consent in the context of a cookie banner may very well use a vector of influence that we can locate in the sphere of the information recipient, i.e. their susceptibility for such design.[604] However, if it is clear (or likely) that a sufficiently large amount of information recipients will display such susceptibility, we may consider it to be predictable enough to presume abstaining from its exploitation to be subject to ethical obligations of the entity responsible for the cookie banner instead. The same will hold true in case of exploitation of such vectors of influence that are not highly prevalent, but of which the constraining entity has high confidence in. For example, personalized advertising that pinpoints an individual's lack of self-control at certain times of day, expectations of which are held with high confidence due to extensive profiling, is clearly targeted at a certain vulnerability. As a result, the constraining entity will likely "enjoy" ethical attributability of the consequences of information interaction over this vector of influence.

This leaves the vectors of influence which are not merely presenting as extensions of the underlying domain or the sphere of information recipients: informational agents and their characteristics. We may recall that informational agents are non-human processes within or on the informational pipeline that intersects with the IoE, that are (from a subjective viewpoint) sufficiently compact, can be reasonably ascribed agency, are sufficiently complex and have the potential to impart persuasive qualities on information.[605] As Section 9.2 has outlined, this encompasses processes which, in colloquial parlance, may be addressed as carriers of intuitive concern (and as a result, attachment points for ethical analysis). For example, we may denote a profiling algorithm, a process for personalized advertising or the software fueling adaptive user interface design in real time as informational agents.[606] As opposed to vectors that are domain specific or located mostly in the sphere of the information recipient, affirmation of

certain algorithms) and serves as an embedding of other vectors of influence. This is visible in current developments as well, as outlined previously knowledge about e.g. machine learning was already developed long before it was applicable, as the underlying hardware was simply not up to the task at the time.

[604] For a discussion of such dark patterns see e.g. Section 6.4.3

[605] See Section 9.2. Note that this thesis has eschewed ostensibly similar definitions that were conceived for different purposes. For example, the thesis has explicitly not considered informational agents as moral agents. This is in line with the present analysis in which informational agents and the ethical obligations and their breaches may be attributed to a moral agent (i.e. another entity or individual)

[606] The pragmatic approach taken in this thesis does not necessitate exact delimitations. Indeed, as was already flagged previously, one may very well consider an amalgamation of agents that process information in subsequent order as a singular agent as well. This is particularly true given that the concept of informational agent adopted in this thesis relies on subjective elements as well; depending on the insight of the technology stack and the respective informational pipeline along which information flows to an information observer or recipient, subjective compactness and agency may be assessed quite differently. Ultimately, the utility of the adopted agent conception is not meant to satisfy stringent philosophical and logical requirements but to give yet another useful term in the vocabulary toolset this thesis aims to provide to address intuitive concerns about autonomy constraints imposed by technology.

attributability of such agents and their characteristics is comparatively easier. Compared to the first, any causality between the agent and an autonomy constraint is much stronger, the resulting autonomy constraints are likelier covered by intent or acknowledgement of the entity deploying the agent and the deployment of many agents is typically a choice with real alternatives. As a result, whereas attribution may not be possible at all in many vectors of influence connected to the fundamental technology domain, this problem does not apply here. Compared to the second, there is a similar picture. Autonomy constraints characterized by workings and characteristics of informational agents can be attributed to the deploying entity much cleaner as there is less reason to consider attribution to a third party (i.e. the information recipient). However, separate from the question of attributability is the question of ethical attribution. Even when affirming that a certain type of technological process is attributable, and that its deployment and use raises ethical challenges or breaches ethical obligations of its deployer, connecting this to a concrete deployer may be difficult. As is pointed out in the existing literature on this topic, the interweaving of different interdependent processes and human actions and their opacity create situations in which informational agents are attributable generally but not in the concrete case.[607]

Assuming that attribution is possible both in the abstract and the concrete, the question then remains to which entity responsibility for autonomy constraints ought to be attached. The first agent characteristic to consider is that informational agents must have some intent, goal or target themselves, and such intent is typically externally inherited.[608] The second characteristic to consider is that informational agents are placed upon the informational pipeline and are unlikely to form spontaneously.[609] As a result, informational agents can impose ethical obligations on different entities. This section suggests a few distinctive levels of attribution.

First, the entity that has placed and uses an informational agent in pursuit of their interests and to their benefit, i.e. enjoys benefits from imposing autonomy constraints and to whom the inherited intent may be traced to serves as most likely holder of ethical obligations. Second, the entity that has constructed the informational agent, if similar, may be ethically responsible as well. Here at least two instances of responsibility are conceivable: Either attribution to the first entity fails, and the "manufacturing" entity resumes subsidiary responsibility (e.g. in cases of malfunction), or the nature of the agent is inherently so risky that the constructing entity may never be fully exculpated. Third, we may presume (lesser) ethical responsibility for entities that have facilitated the installation of the informational agent and its consequent autonomy constraints through some means. Fourth and finally, we may attach ethical responsibility akin to an obligation to protect in knowledge of potential autonomy constraints to entities that are adjacent to the information flow potentially carrying persuasive and constraining elements.[610]

---

[607] See e.g. Mittelstadt and others and the follow-up research by Tsamados and others. Specifically for the relevance of attribution in the sense of responsibility under European Data Protection Law see in particular Anton Vedder and Laurens Naudts, 'Accountability for the Use of Algorithms in a Big Data Environment', *International Review of Law, Computers & Technology*, 31.2 (2017), 206–24 <https://doi.org/10.1080/13600869.2017.1298547>.

[608] See Section 9.2.3

[609] While there are edge cases in which this may be the case, e.g. in the case of automated programming and deployment of processes, this seems currently of lesser practical relevance. In such edge cases, attribution of ethical responsibility will likely flow up to the first process to which intent may be attributed.

[610] Consider for an analogy the legal (and arguably ethical) obligation of providing appropriate cybersecurity measures. In case of a data breach or malicious cyberattack, the infrastructure provider is clearly not responsible for the actual matter, but may be held accountable for failing to protect against (illegal and unethical) measures taken by other bad-faith actors. The same may hold true for autonomy constraints. For example, a internet platform may be ethically obligated to take certain measures to prevent misleading advertising or minimize profiling opportunities or information leakage by and to third actors.

The structure of the respective informational agent will likely determine on which entity the main weight of responsibility ought to lie. In particular when the risk of autonomy constraints is a particularity of the design of the agent and not immediately obvious, more responsibility will rest on the entity that has designed it as opposed to the (perhaps unknowing user). At the same time, as knowledge and awareness of the deploying entity increases, the burden of ensuring ethical compliance will likely shift.

We can make plausible this conceptual structure by considering existing approaches to minimize risks in the information technology domain. In particular, the above mirrors existing strategies in the European Data Laws. For example, the European Artificial Intelligence Act (currently in its draft stage) imposes obligations on AI system users, providers and manufacturers: Providers (i.e. developers) (or product manufacturers) are responsible for the general compliance of an AI system, if it inherently presents high risk.[611] Subsequent obligations of the same nature are imposed on users, importers and distributors insofar they place the system on the market, put it into service or modify it.[612] In addition to this, users of such systems have additional responsibilities as relate to their use.[613]

Finally, in all cases discussed above, it is worth keeping in mind that ethical attribution does not yet mean that an autonomy constraint is indeed unethical, but merely that if ethical obligations exist or they are breached, that these connect to the entity to which they are attributed to. The content of such obligations, potential avenues for justification and the like remain unaffected by ethical attribution of a vector of influence to the constraining entity, to the information recipient or perhaps not being able to attribute it at all.

11.3.2.2 Clustering Ethical Concerns

In addition to the above, an added "bonus" result of intersecting the existing elaborations on both how autonomy ought to be conceptualized and how to understand the autonomy-relevant phenomena of information transfer processes (i.e. the respective vectors of influence) with the now outlined general ethical grid, allows us to place autonomy concerns that are both intuitive and present in existing debate (as outlined throughout this thesis). For example, we may consider the below figure in which common areas of ethical concern can be assigned to the ethical grid.

[611] See generally Chapter 3 Draft AI Act
[612] Art 28 Draft AI Act.
[613] Art 29 Draft AI Act. It is noteworthy that in the modus of the draft AI act, the main burden is presumed to lie with the provider, as users may not have the knowledge or skill to fully grasp the working of an AI system. In particular, monitoring and use ought to follow instructions drawn up by the provider, reinforcing the notion that the inherently risky nature of such systems ought to be best met with protection at the level of the designing entity.

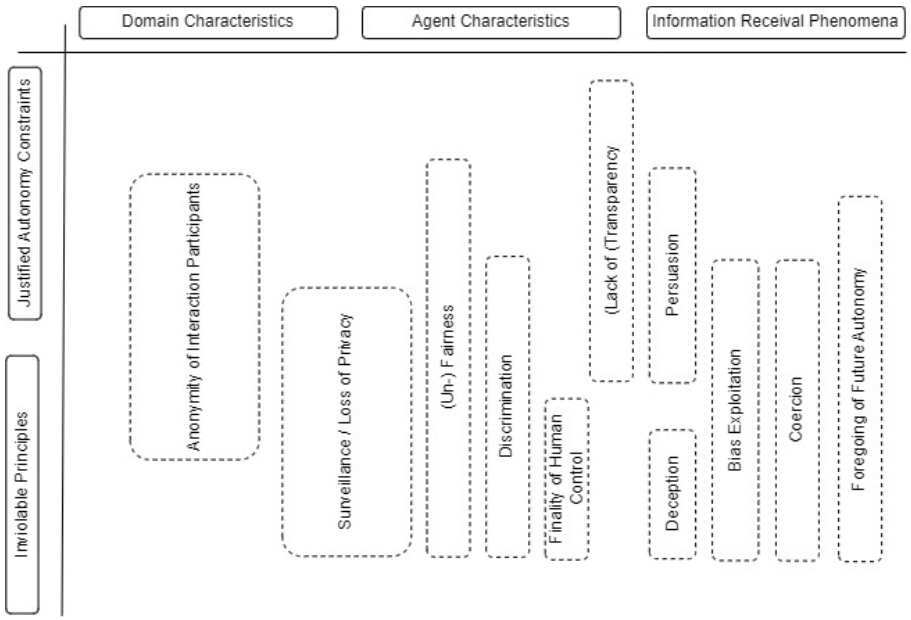

Figure 10 – Visualization of the Intersection of Ethical Gradient and Vectors of Influence (select.)

From an operationalisation standpoint, this may prove useful in visualizing and prioritising different risk of autonomy constraints that a given technology stack may incur on users or other individuals which are affected by the respective technology.

11.3.2.3 Interim Summary

Just as the pragmatic account of autonomy has outlined important factors to consider in an ethical analysis of autonomy constraint, so do the concept of vectors of influence. Not only does the outlined framework allow for a visualization (and subsequent prioritization) of autonomy-constraint-related risks, but also allow for a first approximation of the attribution of ethical responsibility for (unethical) autonomy constraints. The intensity of an autonomy constraint is determined by the respective vectors of influence along the flow of information from origin to recipient. Insofar illegitimate persuasiveness is added by these vectors, the responsibility for such unethical effectuation follows the responsibility for the respective vector. As this section has shown, vectors derived from domain-characteristics will rarely result in attributable ethical responsibility, while vectors derived from the sphere of the information recipient may be attributed to the recipient themselves, potentially precluding responsibility of other entities. The main areas of ethical responsibility incurred will be the vectors that relate to informational agents, in which situation attribution may be assigned to or divided among the user or beneficiary of the agent, its designer and in some situations to facilitating entities as well as to entities that have not fulfilled any remaining obligation to enact protective measures.

11.4 Summary: Towards an Informed Ethical Analysis of Autonomy Constraints

The purpose of this section was to answer research sub-question 3.1, namely how and under consideration of which factors an ethical analysis of technology-imposed autonomy constraints should be conducted, and to what extent the pragmatic account of autonomy is useful in this context. Consequently, this chapter has accumulated the most important components to inform an ethical analysis of autonomy constraints under consideration of the concepts of the

pragmatic account of autonomy and vectors of influence. The guidance given by these concepts and the general structure of the analysis is visualized in the below figure.

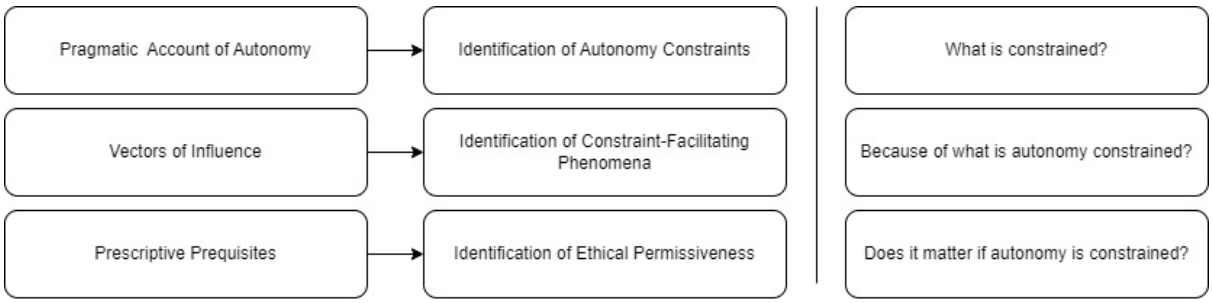

Figure 11 - Components of Ethical Analysis

We can thus formulate a strategy for analysis of potential autonomy constraints of a certain phenomena. As a "zeroth" step, we may consider if a certain situation triggers intuitive concerns or if it belongs to a nexus of phenomena that is generally considered ethically problematic. This may serve as an initial motivation and selection criteria for subsequent ethical analyses. The grid outlined between previously identified vectors of influence and a relevant ethical gradient may aid with this purpose, as it can help cluster situations of intuitive concern. The analysis can then proceed as follows:

First is the subsumption of phenomena under the pragmatic account of autonomy. This step of the analysis investigates if certain aspects of a situation meet the requirement of an intrinsic, relational or informational autonomy constraint as outlined by the pragmatic account. In the affirmative case, whatever the applicable ethical framework in the relevant situation[614], particular consideration will have to be given to factors derived from the pragmatic account. Irrespective of applying a framework informed e.g. by deontological or teleological considerations, the difference in quality of autonomy constraints highlighted by the pragmatic account require careful consideration and weighing. As this section has shown, the factors of particular importance will likely be (apart from harm caused or risked and intent) the type and method of mental state insertion undermining intrinsic autonomy, the repellency and action potential compression undermining relational autonomy and the (lack of) veracity of information transferred as well as the (lack of) prevalence of information both contextual and about the constrained individual's action potential undermining informational autonomy.
The second step in an analysis may consist of mapping the phenomena triggering concern and subsumed under the pragmatic account onto an ethical gradient and considering if the resulting autonomy constraint (or the process leading to it) is completely morally impermissive, may be justified, or raises no concerns after closer inspection.

Having established the ethical "weight" of the autonomy constraints, a logical third step is attribution. Here again, previous analysis provides useful heuristics. Connecting autonomy constraints to their dominant vectors of influence allows for assumption of a first likely attribution. This section has outlined how domain-specific vectors only allow attribution in limited cases, while vectors within the sphere of the information recipient will likely mandate a close weighing of responsibility between the constraining entity and the information recipient. Finally, agent-specific vectors will typically be attributable in abstract; in concrete cases an autonomy constraint may touch upon ethical obligations of different classes of entities, such as the user, constructor or facilitator of the agent. Further analysis can build upon this

[614] See FN 538 and FN 578 for a closer elaboration of this neutral approach.

preliminary assumption, map the phenomena in question and consider the existence of other, exigent factors,[615] which may change the logic of attribution.

Utilizing both the ethical weight of an autonomy constraint and any additional factors derived from a quality or status of an entity to which the autonomy constraint was attributed, we finally have collected all necessary information to reach a verdict about the ethical implication of the autonomy constraint in question.[616]

12 Regulatory Acknowledgment of Individual Autonomy in European Digital Legislation

Section 4 has already analysed the reflection of autonomy generally in the legal domain and has found it to be both a natural inherent prerequisite to the legal domain and reflected in many of its aspects. This section provides another layer of consideration and aims to answer research sub-question of autonomy which asked how the concept of autonomy as exercised and constraint in presence of technology is reflected in existing and upcoming European legislative instruments. The analysis proceeds as follows: Section 12.1 introduces the fact that newly proposed European digital legislation has introduced the term "autonomy" as an explicit object worthy of protection. Section 12.2 gives a short overview over the common meaning of the term, as the basis for subsequent interpretation. Similarly, Section 12.3 provides context by highlighting how the link between data protection, privacy and autonomy has been recognized by a now famous German court case pertaining to informational self-determination. Both of these two latter sections retreat some ground previously discussed in this thesis and contextualize it for this current chapter.[617] Section 12.4 argues that the mental aspect of autonomy and decision-making have enjoyed increased attention as more recent legislative proposal aim to strengthen protection for decisional privacy. Section 12.5 investigates the scope of Article 6 under the DA proposal under this context and considers if the provision is aimed purely at adversarial design. Section 12.6 connects these threads and highlights that recognition and protection of individual autonomy is no longer limited to a status of meta-principle but has surfaced as an explicit value. Section 12.7 concludes and highlights potential next steps in the European Data Laws.

[615] Of course these factors will depend upon the general ethical framework that is assumed. An example for such factors may be e.g. a duty to protect or minimize environmental or infrastructural risks, or to abstain from exploiting biases the constraining entity is not responsible for.

[616] Consider the example of the cluster related to surveillance and loss of privacy. First, we can validate that these are indeed an ethical concern. This seems mostly uncontentious: surveillance is widely identified as ethically risky, and these concerns have already been operationalized to a certain extent through international consensus on establishing (or recognizing) the fundamental right to privacy. Second, using the above grid we can place the cluster of ethical concerns. From the perspective of the informational pipeline and its respective vectors of influence we can situate this cluster of ethical concern to be predominantly connected to the domain characteristics and agent characteristics. Recalling the vectors of influence pertaining to information technology domain characteristics, we find that surveillance and potential loss of privacy is affected e.g. through the immediacy of information transfer (removing a shielding temporal buffer against observation), and the already identified increasing erosion of privacy, as well as the easy implementation of scalable, non-human actors (such as observation and profiling algorithms). Similarly, when considering informational agent characteristics, we can recall them to exhibit e.g. placement and effectuation obscurity as well as opacity of their decision making (meaning that privacy protective measures may be undermined).

[617] See e.g. Sections 1.1, 3 and 4.3.

12.1 Introduction

Under the umbrella of its digital strategy, and in particular its European Strategy for Data (ESD),[618] the European Commission has intensified its regulatory presence in the digital domain in recent years. One of the most recent proposals in this space is the Data Act (DA)[619], an instrument that primarily (but not exclusively) regulates data access and use between businesses themselves (B2B) and the public sector (B2G). Curiously, with this proposal, the European Commission has now introduced the term 'autonomy', into the (prospective) legislative body that regulates the European digital landscape. This section suggests that this marks a milestone in a turn towards protection of decisional privacy, surfacing a hidden trend that has been growing in recent years.

How did we get here? One of the main goals of the DA (which is still in its draft stage at the time of writing) is to 'increase legal certainty for companies and consumers' in relation to the generation and use of data. To this end, the DA proposal will mandate that data-collecting devices (i.e IoT-devices)[620] must be designed in a way that the data their use generates is accessible by the user either directly[621] or through request to the respective data holder.[622] This requirement seemingly mirrors the General Data Protection Regulation's (GDPR) right of access, but now applies to data beyond personal data. The DA offers a shortcut of sorts to the user, who now has discretion over this data and may want to make it available to a third party. Upon request of either the user or another party acting for the user, the data holder must make the aforementioned data available to a third party directly.[623] These third parties, upon obtainment of the data by the data holder, are subject to a number of requirements. Interestingly the first of these requirements reads as follows:

'*The third party shall not […] coerce, deceive or manipulate the user in any way, by subverting or impairing their autonomy, decision-making or choices of the user, including by means of a digital interface with the user.*'[624]

This provision marks the first time that the term autonomy is used in the main text of data-related European legislation as proposed by the European Commission, and the first time it is used to refer to human autonomy. However, the draft does not provide a definition of the term, so the scope of the obligation to not impair a user's autonomy appears somewhat opaque.

While the explicit wording is novel, it is by no means a new idea. This text aims to trace the legal recognition of self-determination in recent European data-related legislation and provide the context in which this regulatory approach is grounded. Through a short survey of legal sources, this section shows that individual autonomy has always been held as a meta-principle informing regulatory measures, in particular in the field of privacy and data protection. The explicit inclusion of individual autonomy in the DA must be seen as a consequential next step following increased focus on decisional privacy and the mental aspects of self-determination.

---

[618] European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions A European Strategy for Data COM (2020) 66 Final' <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066>.
[619] Proposal for a Regulation of the European Parliament and of the Council on harmonized rules on fair access to and use of data (Data Act) [2022] COM/2022/68 final.
[620] See e.g. ibid Rec. 14, 19.
[621] Ibid Art 3
[622] Ibid Art 4
[623] This as long as the third party is not considered a gatekeeper under the DMA, see ibid Art 5 DA
[624] Ibid Art 6 para 2 (a)

12.2The Ordinary Meaning of Autonomy

Proper legal methodology suggests that a good place to start interpreting this newly introduced term is to consider its ordinary meaning.[625] We can recall from the beginning of this thesis that by translating literally from its Greek origin words *αὐτο* (meaning 'self') and *νόμος (*meaning rule or law), we derive at a meaning of a state or capacity to govern oneself, or more liberally, self-determination. While this can be a characteristic of many entities (e.g. states, universities, artificial agents), the use within Article 6 DA makes clear that we must consider individuals (ie the users of a product or service collecting data) as the subject of this self-determination. Autonomy in the meaning of Article 6 DA hence clearly refers to individual (or personal) autonomy. The concept of individual autonomy itself is used colloquially and is in need of further qualification. Two avenues allow further insight. In the domain of philosophy, where individual autonomy is best situated as a technical term,[626] autonomy is often considered as a congruence between an individual' choices, the motivations (or mental states) that inform those choices, and the motivations behind those motivations[627], although this view is not uncontested.[628] Simplified, an individual is autonomous if they make autonomous choices, and they make autonomous choices if their choices are aligned with their preferences. But clearly, the term autonomy is already used in the legal domain as well, as outlined in Section 4. For example, the concept of self-determination is found e.g. in the principles of contractual autonomy where it describes the capacity of self-governance through freedom of contract.[629] And more broadly still, an individual's capacity to govern oneself has been described by the legal philosopher Hans Kelsen as one of the basic presuppositions of the legal domain itself.[630] Considering this, it is clear that the term in question needs to be understood in context, lest it be ambiguous in its meaning. The first contextual clue comes from the centrality of the tenet of privacy within the digital strategy of the European Union. The second clue here comes from the domain, which is targeted by the regulation, namely the domain of data processing.

12.3 Informational Privacy and Autonomy

Much of the current and prospective legislation surrounding the ESD is firmly aimed at measures to safeguard privacy, and its derivative, but now mostly emancipated concept of data protection.[631] (Presently, the right to data protection is understood to extend beyond the protection of the right privacy.)[632] This is highly relevant, as some sort of privacy is often seen

[625] Koen Lenaerts and José A Gutiérrez-Fons, 'To Say What the Law of the EU Is: Methods of Interpretation and the European Court of Justice', *Colum. J. Eur. L.*, 20 (2013), 3.
[626] See eg Reath; Christen.
[627] Frankfurt; Sneddon, *Autonomy*; Dworkin, 'The Nature of Autonomy'.
[628] Stoljar; Brison; Mackenzie.
[629] See for the nuances in terminology eg Patti.
[630] Kelsen. Hans Kelsen describes the concept as freedom of will.
[631] See eg M Tzanou, 'Data Protection as a Fundamental Right next to Privacy? 'Reconstructing' a Not so New Right' (2013) 3 International Data Privacy Law 88 <https://academic.oup.com/idpl/article-lookup/doi/10.1093/idpl/ipt004>; J Kokott and C Sobotta, 'The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR' (2013) 3 International Data Privacy Law 222 <https://academic.oup.com/idpl/article-lookup/doi/10.1093/idpl/ipt017>; Raphaël Gellert and Serge Gutwirth, 'The Legal Construction of Privacy and Data Protection' (2013) 29 Computer Law & Security Review 522 <https://www.sciencedirect.com/science/article/pii/S0267364913001325>.
[632] Alexander Roßnagel and Christian Geminn, ''Privatheit' Und 'Privatsphäre' Aus Der Perspektive Des Rechts - Ein Überblick' [2015] Juristenzeitung 703.

to serve as a prerequisite to autonomy[633], or at least as an enabling state hereto.[634] As a result, due to their entangled nature, measures that safeguard autonomy tend to do so informed by privacy concerns as well.[635] And at the same time, apprehensions about autonomy inform the consideration of privacy concerns.[636] This is not only of theoretical relevance but relevant to understanding European data-related legislation as well. The prospective legislation under the ESD continuously emphasizes its grounding in the catalogue of fundamental rights. And naturally, two of these fundamental rights, as codified in the European Charter of Fundamental Rights are the right to privacy and data protection.

This is only further substantiated when considering the substantively similar Article 8 of its 'precursor', the European Convention on Human Rights and its attached case-law. Under this provision, protection of one's psychological integrity, personal beliefs and personal development, as well as one's physical and moral integrity sphere of relations to and between other people (and oneself) was found to be protected.[637] This aligns well with the ordinary meaning of autonomy explored previously, in which mental coherence and integrity are paramount.[638]

As outlined in Section 3.3, it is common to distinguish different types of privacy. (To reiterate just one example, Tavani and Floridi differentiates between physical, mental (or psychological), decisional and informational privacy, understanding them as freedom from sensory, psychological, procedural and informational interferences and intrusions.[639]) For the purpose of this analysis, informational privacy, mental privacy and decisional privacy (the latter two of which will both be denoted by the term decisional privacy in this text) of are of most interest here.

In the domain of technology regulation, informational privacy (or information privacy), i.e. privacy pertaining to the collection and use of information about oneself, seemed to be historically established as a primary motivator. Informational privacy is an intuitive prerequisite to autonomy to the extent that interference with an individual's autonomy becomes more effective as more information about the individual is known and can be acted upon. Connecting the concepts of privacy and autonomy is then the notion that the right to privacy encompasses (partly) the concept of individual autonomy and concerns itself specifically with the human being as an autonomous subject. Somewhat prescient to the issues of emerging

---

[633] See eg Stanley Benn, 'Privacy, Freedom, and Respect for Persons' in Ferdinand Shoeman (ed), *Philosophical Dimensions of Privacy: An Anthology* (Cambridge University Press 1984) pp 241ff.
[634] See eg Kupfer, p. 83.
[635] Indeed, there seems to be considerable overlap in concerns of privacy and autonomy. Holvast, surveying existing literature notes that the terms freedom, control and self-determination are used in almost all publications relating to privacy, from which the relevance to autonomy is self-evident; see Jan Holvast, 'History of Privacy' in Vashek Matyáš and others (eds) (Springer Berlin Heidelberg 2009) p 16.
[636] For example, the European Court of Human Rights has found that Article 8 of the ECHR (ie Right to Privacy) is based on the concept of 'human autonomy'; see eg Pretty v. The United Kingdom (2002) ECHR para 61; (2002) Christine Goodwin v. The United Kingdom (2002) ECHR para 90; Evans v. The United Kingdom (2007) para 71.
[637] Cf. Nicolae Virgiliu Tănase v. Romania [GC] (2019) ECHR para 128 and Bensaid v. the United Kingdom (2001) ECHR para 47.
[638] For a closer analysis on how the case law aligns with the philosophical concept of autonomy compare also Maximilian Gartner, 'Fit for the Future: A Pragmatic Account of Human Autonomy to Understand Emerging Issues in The Internet of Everything'..
[639] See H. Tavani, 'Informational Privacy: Concepts, Theories and Controversies', in *Handbook of Information and Computer Ethics*, ed. by K.E. Himma and H.T. Tavani (Haboken: John Wiley, 2008), pp. 131–64; Luciano Floridi, *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*, p. 208f.

technologies, the German Constitutional Court (*Bundesverfassungsgericht*) (as already mentioned in Section 4.3, recognized the concept of informational self-determination as encompassed in the general personality rights of German Citizens, as imbued by the German Constitution as early as 1983.[640] The decision noted that automated information gathering and processing allowed for an exceedingly complete profile of the personality of individuals and, consequentially, ever more effective means of influence.[641] Consequently, the court recognized the right of an individual to determine use and disclosure of their personal data.[642] Naturally, this decision, affirming the arguments that German legal scholars had been bringing forward for a while,[643] had significant impact beyond the jurisdiction it was reached in.[644] Rephrasing this in the context of this text, the court recognized the increased leverage over individual self-determination that technology enables and as a result strengthened their informational privacy (and it did so in a way that is largely compatible with concept of autonomy outlined in Section I). Since then, informational privacy is generally still safeguarded by current legislation. One of the more prominent matters of discussion with respect to informational privacy and autonomy, even preceding European legislation, was the so-called *right to be forgotten*.[645] For example, the General Data Protection Regulation gives data subjects extensive control and information over how their personal data is collected and processed. And even the draft DA entitles users of services to information about non-personal data they produce. We take from this that if autonomy in some shape is inherent to (informational) privacy and data protection, the existing safeguards for those values means that the protection of autonomy has been a non-explicit goal of legislation already, even if left unnamed.

12.4 The Turn Towards Explicit Decisional Privacy Protection

Current and upcoming regulations gravitate toward more explicit protection of decisional privacy as well. As mentioned above informational privacy is seen as a prerequisite safeguard to an individual's exercise of autonomy. Decisional privacy is even closer entrenched with an individual's autonomy, if not identical to the concept[646], as it describes the very ability of an individual to make decisions.[647] Current and upcoming regulation seems to have taken a turn towards more explicit protection of this aspect of privacy and autonomy *per se* as well, as evidenced by the following data points.[648]

12.4.1 Autonomy Considerations in the GDPR

Following in the footsteps of the Data Protection Directive 95/46/EC, the GDPR represented the first comprehensive European regulation of the digital domain. While much of the measures

[640] See eg Rouvroy and Poullet.
[641] *1 BvR 209/83 (Volkszählung)* [1983] German BVerfG BVerfGE 65, 1 – 71 [93].
[642] Ibid para 1.
[643] Cf in particular Steinmüller and others, p. 93,96,120.
[644] With some scholars going as far as calling it an '*avant-garde* decision', see Rouvroy and Poullet (n 23) 45.
[645] See Section 4.1f in Cécile de Terwangne, 'The Right to Be Forgotten and Informational Autonomy in the Digital Environment', in *The Ethics of Memory in a Digital Age* (Springer, 2014), pp. 82–101.
[646] Finding a distinction between decisional privacy and autonomy *per se* is difficult and beyond the scope of this text. The interested reader is referred to Marjolein Lanzing, ''Strongly Recommended' Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies' (2019) 32 Philosophy & Technology 549 <http://link.springer.com/10.1007/s13347-018-0316-4>.
[647] Koops and others, p. 50ff. Note the distinction between decisional and intellectual privacy (freedom of mental intrusions), in which the authors suggest that decisional privacy is the active exercise of intellectual privacy. This tracks well with the conception of autonomy as a product of congruent mental states.
[648] The following represents but a short survey of the legal instruments referenced therein. A full analysis of each instrument through the lens of autonomy would be beyond the scope of this section.

of data protection generally relate closely to informational privacy as mentioned in the previous section, there are similarities to the above-mentioned Article 6. The most salient protection of decisional (and mental) privacy comes with the provisions concerning a data subject's consent. As is well known, consent represents one of the main justifications for data processing, and this consent must be given freely. As highlighted by the European Data Protection Board (EDPB) and its predecessor, the Article 29 Data Protection Working Party, deception, intimidation, coercion, compulsion pressure or 'inability to exercise free will' can preclude the validity of consent given.[649] The wording of Article 6 DA is strongly reminiscent of this language and can be considered as translating the assessment of the EDPB into code (albeit in a different overall context).[650] Another autonomy-safeguarding measure is the protection awarded by the principle of lawful, fair and transparent processing outlined in Article 5 para 1 (a) of the GDPR. For example, the EDPB has interpreted the fairness principle encoded in the GDPR as being incompatible with autonomy-constraining nudges and dark patterns (see for this below) even prior to their guidelines on adversarial design.[651]

12.4.2 Autonomy Considerations in the European Approach to Artificial Intelligence

Another important testament to the increasingly explicit focus on autonomy is the report by the Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission. In it the expert group lists respect for human autonomy, including mental autonomy, as a core ethical principle (and equates the concept with decision-making).[652] The report identifies practices of coercion, deceiving, manipulating conditioning or herding individuals as particular threats to this principle.[653] This is also in line with international trends; a 2019 survey of ethical guidelines for AI found that 'freedom and autonomy' was considered an explicit core principle in almost half of the frameworks investigated.[654] Following this, a draft of the Artificial Intelligence Act, leaked by the online journalism company Politico, included the prohibition of an AI system if it was designed or used in a manner that 'manipulates human behaviour, opinions or decisions through choice architectures or other elements of user interfaces, causing a person to behave, form an opinion or take a decision to their detriment' or 'exploits information or prediction about a person or group of persons in order to target their vulnerabilities or special circumstances, causing a person to behave, form

[649] European Data Protection Board, 'Guidelines 5/2020 on consent under Regulation 2016/679, Version 1.1 (2020) https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf, para 24, 47.

[650] The fact that the scope of the legislation is different between the instruments is not as relevant here. While the GDPR deals first and foremost with personal data, and the data act complements this by also dealing with non-personal data, the notion of giving justification to processing by a data controller (in the GDPR) or a third party (in the Data Act proposal) remains congruent. At this point it is noteworthy that the Data Act proposal does not verbatim use the concept of consent in its provisions dealing with third party data sharing, but instead describes a situation in which a request by the user or by the third party acting on behalf of the user issues a request to the data holder.

[651] European Data Protection Board, 'Guidelines 3/2022 on Dark Patterns in Social Media Platform Interfaces: How to Recognise and Avoid Them' (2022) <https://edpb.europa.eu/system/files/2022-03/edpb_03-2022_guidelines_on_dark_patterns_in_social_media_platform_interfaces_en.pdf>

[652] High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 2019, p. 10,26 <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419>.

[653] High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence (n 35) p 12.

[654] Jobin, Ienca, and Vayena.

an opinion or take a decision to their detriment'.[655] While both of these provisions have been changed (and arguably weakened) by the final proposal by the European Commission, the adapted prohibitions still target the mental integrity of individuals under threat. Under Article 5 of the proposed AI Act, prohibition now applies to certain 'subliminal techniques' which 'distort a person's behaviour', or to certain systems that 'exploit any of the vulnerabilities' such as 'age, physical or mental disability' of persons. These formulations may very well be subject to further changes, as the LIBE and IMCO committees have tabled possible amendments to the language of the article.[656] Nonetheless, even with the narrower scope, the act clearly recognizes mental integrity of individuals as a matter of concern. Unlike the report of the High-Level Expert Group, the Act does not use the term autonomy in connections with individuals (and instead describes artificial intelligence as autonomous, somewhat exhausting the term for the purposes of the regulation).[657] However, a current tabled amendment by Axel Voss, Deirde Clune and Eva Maydell would introduce a reference to (human) personal autonomy as well.[658]

12.4.3 Autonomy Considerations in the European Data Strategy

While the European Approach to Artificial Intelligence was focused on the eponymous intelligent systems, the European Commission has introduced a swath of legislation that covers the European market for data as a whole under the umbrella of the European Data Strategy. The DA, which first explicitly introduced the notion of individual autonomy, is the latest of a few instruments under this strategy. Crucially, due to the long timeframe of European legislation, these instruments have been developed in parallel.

During this development phase, one phenomenon of technology design that is considered highly problematic for individual autonomy has demanded increased attention.[659] Increasingly, user interface design characteristics are recognized for their coercive or manipulative power and their ability to lead users to take actions against their interest. This phenomenon is typically called a 'dark pattern'. Examples of dark patterns include protracted procedures to withdraw from paid services or visual and procedural asymmetry in offering choices about data collection

[655] Art 6, leaked AI Act. An archived version is accessible at the following link at the time of writing: https://drive.google.com/file/d/1ZaBPsfor_aHKNeeyXxk9uJfTru747EOn/view

[656] Committee on the Internal Market and Consumer Protection and Committee on Civil Liberties Justice and Home-Affairs, 'Amendments - Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' <https://bit.ly/3Mboami>. It is notable that the IMCO committee particularly has scheduled hearings on the risks of dark patterns by external experts.

[657] See eg Rec. 6 of the AI Act as proposed by the European Commission, see Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised rules on artificial intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts (2021) COM/2021/206 final.

[658] See the proposed Article 4 a in Committee on the Internal Market and Consumer Protection and Committee on Civil Liberties Justice and Home-Affairs.

[659] See eg Gregory Day and Abbey Stemler, 'Are Dark Patterns Anticompetitive? ' 72 Alabama Law Review 1 <https://heinonline.org/HOL/P?h=hein.journals/bamalr72&i=11>; Davide Maria Parrilli and Rodrigo Hernandez-Ramirez, 'Re-Designing Dark Patterns to Improve Privacy', *2020 IEEE International Symposium on Technology and Society (ISTAS)* (IEEE 2020); Diana MacDonald, 'Anti-Patterns and Dark Patterns', *Practical UI Patterns for Design Systems* (Apress 2019) <http://link.springer.com/10.1007/978-1-4842-4938-3_5>; Ari Ezra Waldman, 'Cognitive Biases, Dark Patterns, and the 'Privacy Paradox'' (2020) 31 Current Opinion in Psychology 105; Than Htut Soe and others, 'Circumvention by Design - Dark Patterns in Cookie Consent for Online News Outlets', *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (ACM 2020); Saul Greenberg and others, 'Dark Patterns in Proxemic Interactions', *Proceedings of the 2014 conference on Designing interactive systems* (ACM 2014); Colin M Gray and others, 'Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective', *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (ACM 2021).

on websites through cookie-banners.[660] As a result of this attention, dark patterns also have received increased scrutiny from data protection watchdogs such as the European Data Protection Board and are subject to fines by regulatory authorities.[661]

At the time of writing, none of the instruments of the EDS have been formally finalized, however the Digital Service Act (DSA) has initially reached agreement in the trilogue procedure.[662] While the original proposal focused on manipulative activities in the context of more systemic negative consequences to society and democracy, both the Council and the European Parliament have added specific provisions to prohibit dark patterns, making their decisional privacy considerations much more explicit and concrete in turn.[663]

12.5 Explicit Autonomy Protection in the Data Act

The drafting of the DA has run in parallel to the previously mentioned instruments and intensive discussion of coercive or manipulative design on a European level has preceded its release. As a result, the proposal of the European Commission continues the trend towards a more explicit protection of decisional privacy. For the first time, the commission has used the term autonomy as a descriptor of a subject's status that ought to be protected. But does the use of new terminology signal a more expansive protective umbrella or is the protection of one's autonomy limited under the proposal?

At first glance, the structure and context of Article 6 seems to narrow its scope of application somewhat.[664] The language of the proposal prohibits the use of measures to subvert or impair user autonomy 'including by means of a digital interface with the user', a provision that focuses on hostile design that was previously also the subject of DSA amendments.[665] Clearly, the wording of Article 6 and its corresponding recital of the DA proposal see dark patterns as the main risk from which to guard users from prospective data recipients. After all, Article 6 references digital interfaces but no other examples.[666] Digital interfaces are clearly already a meaningful way to subvert or impair someone's autonomy, and do not need an explicit reference to be considered as such. Their explicit (and sole) reference suggests instead that these digital interfaces (i.e. dark patterns) are considered to be of unique relevance to this

[660] See eg Bösch and others; Nouwens and others.
[661] See European Data Protection Board (n 34). As a concrete example, in a recent case, the French CNIL has fined Google LLC and Google Ireland Ltd a combined 150 million euros for adversarial cookie banner design, see *Délibération de la formation restreinte n°SAN-2021-023 du 31 décembre 2021 concernant les sociétés GOOGLE LLC et GOOGLE IRELAND LIMITED*.
[662] However, even more recently, a majority of European MEPs have rejected the latest version drafted by rapporteur Schaldemose. Most of this disagreement concerns the Recitals 28 and 29, both of no concern to this text.
[663] Council of the EU, 'Press Release: Digital Services Act: Council and European Parliament Provisional Agreement for Making the Internet a Safer Space for European Citizens'. Prior to that the proposal did mention manipulative advertising techniques in Rec. 63 and 68 which can be further seen as a safeguard against infringements of decisional privacy and autonomy. The European Parliament also suggested an amendment which used the term autonomy explicitly, see the proposal for a regulation Recital 39 a (new), see European Parliament, *Amendments Adopted by the European Parliament on 20 January 2022 on the Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and Amending Directive 2000/31/EC (COM(2020)0825*, 2022. This technically preceded the use of the term by the European Commission in the Data Act proposal
[664] Art 6 Data Act Proposal (n 2).
[665] Further limiting the scope, not every nudge that affects an individual's decision-making will be outlawed by these provisions. Recital 34 of the Data Act proposal lays out that 'common and legitimate commercial practices that are in compliance with Union law' are not 'in themselves […] constituting dark patterns'.
[666] Ibid.

provision. Similarly, the wording of Recital 34 suggests a strong focus on regulating dark patterns and does not mention any other coercive or manipulative measures. In this sense, the language of the text may be read as more than illustrative application, and instead as concrete target-setting.[667] In a wider context, this also accords with the terminology used in the amendments suggested to the DSA by the European parliament which preceded the DA proposal by a brief margin. There, the term autonomy is again used exclusively in conjunction with dark patterns.[668]

On the other hand, grammatical and syntactical interpretation yield that the perceived risk to an individual's autonomy is not exhausted by adversarial design. The provision's wording leaves room for coercion, deceiving, manipulation subversion and impairment in 'any [other] way', beyond the aforementioned adversarial digital interfaces. Under this reading, the concepts of 'autonomy, decision-making or choices of the user' seem to suggest one broad protected element, and the prohibition on prospective data recipients to negatively affect their users' exercise of these capacities is extensive as a result. This is reinforced by the enumeration of protected elements quoted above and their relationship to each other: Clearly 'decision-making' of the user precedes their 'choice' in time and abstraction but is inextricably linked. In a similar way, 'autonomy' is clearly connected (and not separate to) the element of decision-making but describes it merely at another preceding level of abstraction. Including these concepts in the prohibitive provisions of legislation is a strong indicator that the protection users enjoy in relation to their self-determinative capacities should be interpreted holistically. Here, we may draw parallels to the regime of consent under the GDPR. As mentioned above, the EDPB has found coercion, deception or similar practices to be incompatible with valid consent. Despite the slightly different wording, it is clear that practices deemed to be problematic under Article 7 of the GDPR are also prohibited under Article 6 DA, as they would invariably subvert or impair the autonomy, decision-making or choices of the relevant users.

As a result, it seems likely that the unprecedented and equivocal wording present in Article 6 would be fertile ground for expansive interpretation by regulatory authorities and the ECJ alike. While the motivation behind Article 6 DA is clearly derived from increasing public attention on the matter of dark patterns, the current text of the proposal leaves the door wide open for holistic (and correspondingly hard to predict) decisions about user-data recipient interactions.[669]

12.6 Analysis: From Meta-Principle to Explicit Protection

As is often the case with European legislation covering the digital sector, much of the ambiguity of certain provisions stems from the lack of explicit definitions. Article 6 DA is no exception. But this opacity and the many potential amendments to the proposal that will surely be brought up in the ongoing legislative process notwithstanding, the inclusion of individual autonomy in the DA signals a paradigm shift.

---

[667] Ibid Rec. 34

[668] Report on the proposal for a regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM (2020) 0825 – C9-0418/2020 2020/0361(COD)) < https://www.europarl.europa.eu/doceo/document/A-9-2021-0356_EN.pdf>.

[669] This would create a situation not dissimilar from the impact of principles relating to processing of personal data enshrined in Art 5 of the GDPR as mentioned above, see European Data Protection Board, *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default*, 2020 <https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf>.

Beyond the digital domain, individual autonomy, as a concept situated somewhere between the domains of ethics, law and cognitive science, has long held the position of a meta-principle in the legal domain. Its status as an underlying foundation of legal systems generally and fundamental rights specifically is well established.[670] There is also considerable reflection of different aspects of individual autonomy in specific fundamental rights. Arguably, more physical aspects of an individual's autonomy have commanded the majority of attention for a long time. For example, the right to liberty or the prohibition of forced labour interdict palpable and direct constraints of an individual's factual capacity to make choices. But as the digital domain has grown in relevance and our understanding of its potential to interfere with an individual's non-physical autonomy has improved, new tools within the catalogue of fundamental rights have emerged as appropriate safeguards.[671] As Diggelmann and Cleis have shown, the modern conception of a right to privacy has only emerged relatively recently, and this has been at least correlated with the rise of (information) technology.[672] Clearly, this right to privacy (and by an extension, the right to protection of personal data) is closely connected to an individual's mental capacity of self-determination, just as the German *Bundesverfassungsgericht* found. Similarly, the right to mental integrity, freedom of thought, conscience and religion, and the freedom of holding opinions without interference similarly capture concerns for an individual's mental (or decisional) autonomy.[673] And while the original legal instruments have eschewed the term autonomy, secondary literature such as legal commentaries have recognized the concept when discussing the underlying motivating principles of fundamental rights.[674] Even more granularly, many legal provisions can be traced to the purpose of shielding some aspects of an individual's mental autonomy, and they often do so in light of emerging technologies. For example, legal scepticism towards subliminal or misleading advertising, non-validity of legal consent under duress or exposure to false information or limits to data collection all connect to a collective concern over individual autonomy. The current focus on adversarial design or 'dark patterns' in European legislation is the most recent testament to this. Finally, some scholars have noted a perceived disconnect between the value placed on the mental aspects of individual autonomy and the protection fundamental rights award in light of neuroscientific advancements and the manipulative impact they may provide,[675] and the matter has subsequently received attention in a parliamentary question to the European Commission.[676]

---

[670] Kelsen (n 13); Jaunius Gumbis, Vytaute Bacianskaite and Jurgita Randakeviciute, 'Do Human Rights Guarantee Autonomy?' [2008] Cuadernos Constitucionales de la Cátedra Fadrique Furió Ceriol 77. It is also worth noting that the original use of the term autonomy was not in the context of individuals but for politic collectives.

[671] In a similar notion, the European Court of Human Rights maintains a collection of decision related to new technologies, where the application of the human rights catalogue that entered into force in 1953 is summarized, see European Court of Human Rights - Press Unit, 'Factsheet: New Technologies', 2022 <https://www.echr.coe.int/documents/fs_new_technologies_eng.pdf>.

[672] Diggelmann and Cleis. In the United States, the concept of modern privacy is often thought to have been heralded by Brandeis and Warren and their seminal article, see Warren and Brandeis. Interestingly, Warren and Brandeis too discuss emerging technology in detail as a threat to privacy, in their case the advent of photography and the logistic of newspaper circulation. However, the notion of privacy in the context of United States legislation is somewhat idiosyncratic and not fully congruent with European understanding of the same. See also Dorothy J Glancy, 'The Invention of the Right to Privacy', *Ariz. L. Rev.*, 21 (1979), 1.

[673] See Maximilian Gartner (n 21).

[674] See eg Nowak.

[675] See eg Marcello Ienca and Roberto Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotechnology', *Life Sciences, Society and Policy*, 13.1 (2017), 1–27.

[676] Emmanouil Fragkos, 'Question for Written Answer E-004810/2021 to the Commission (Legislation against the Manipulation of the Human Brain through Neuroscience)', 2019 <https://www.europarl.europa.eu/doceo/document/E-9-2021-004810_EN.pdf>.

Thus, to repeat the obvious, the concept of autonomy is not foreign to the legal domain. What is new to a certain extent is first, the increased consideration of non-physical aspects of autonomy and second, its recognition not only as a meta-principle but as explicit protected value. As outlined above, European legislation in the digital domain has inched closer to explicit recognition. Both within the European Approach for Artificial Intelligence and the European Strategy for Data, the constraining effects technology can have on an individual's autonomy have been targeted with increasing precision while maintaining technology-neutral language. With its pending explicit recognition in the DA (and to a certain extent in the Digital Service Act), individual autonomy is now on the cusp of being explicitly recognized by a European legal instrument regulating the digital domain; hence transferring theoretical and ethical concerns highlighted in advisory bodies and scholarship into more durable code. Should the regulation pass in its current or similar form, individual autonomy will have shed its status as meta-principle and take the position of explicitly protected characteristic. But even if the wording is not adapted in the future legislative process, its inclusion in the European Commission's proposal is already indicative of the trend outlined in this section.

In any case, actors in the digital domains that collect and process data will likely be faced with another opaque, but wide-ranging limit on how to structure interactions with their users. While the ambiguity of the language lends itself to test the limits of this new instrument, the DA proposal is outfitted with the now typical GDPR-esque penalties in case of infringements, resulting in significant financial risks for doing so. As the increasingly hostile approach of legislation and regulatory enforcement towards autonomy-subversion and impairment is continuing, careful considerations will be necessary for data collecting and processing entities.

12.7 Conclusion and Outlook

With the DA proposal, the regulatory grasp of the European Union in the digital domain has further caught up with the need voiced by technology ethicists and stakeholders[677] to protect the mental aspects of an individual's capacity to self-determination. While the immediate attention of regulatory authorities applying the provisions in its current form would likely focus on combating instances of adversarial design such as dark patterns, the trend towards more holistic protection of individual autonomy and decisional and mental privacy in the digital domain will continue.

This section has investigated research sub-question 3.2, namely how the concept of autonomy as exercised and constraint in presence of technology is reflected in existing and upcoming European legislative instruments.
This analysis has shown that (individual) autonomy has already served as a meta-principle informing relevant legislation. Nevertheless, its 'emancipation' as an explicitly protected concept raises questions to what extent this newly adopted posture will affect the envelope of acceptable interactions in the digital space going forward. The provisions of the DA proposal prohibit impairment or subversion of an individual's autonomy only in the context of user-data recipient relationships. However, autonomy is not constrained solely in these contexts. The recognition of autonomy may very well spill over into a more confident posture of regulatory authorities and courts when considering autonomy constraints under other existing regimes.

[677] See eg Zuboff, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization'; Council of Europe; Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression; Mittelstadt and others.

There, the application of general principles (and extrapolation of the underlying fundamental rights), such as the fairness principle in the GDPR, may herald a more aggressive stance against technology-assisted autonomy impairment or subversion without the need for further legislative change. Considering the function of the DA proposal, this may even be necessary to maintain regulatory consistency. For example, Article 6 DA functionally complements the GDPR's right to data portability, enshrined in its Article 20. But of course, this provision does not mention coercion or subversion of the data subject's autonomy.[678] It seems questionable that if users under the DA proposal are meant to enjoy more thorough protection from autonomy-undermining measures pertaining to potentially non-personal data than data subjects under the GDPR pertaining to personal data. As a result, revisitation of these existing regimes under the now explicit recognition of autonomy may be needed, if not through legislative means, then through enforcement based on 'updated' interpretation of the existing instruments.

[678] Originally, this may have been due to the fact that the GDPR does not explicitly mention the right of the future data controller to issue its own request for data transfer but considers this right to lay with exclusively with the data subject. Here, the Data Act proposal uses different language with an explicit inclusion of a third party 'acting on behalf' the user. But in both cases, the justification for transferring the data originates from the original data subject or user. It seems unlikely, that the wording of Art 20 of the GDPR ought to be interpreted in a way that precludes potential receiving data controllers equipped with power of attorney and valid consent of data subjects from issuing a request on behalf of the data subject. This is true in particular, as the GDPR foresees controller-to-controller transfers in the context of data portability. Instead, this is likely a sign of an evolved understanding of the power dynamics between data subjects and users on the one hand and data service providers on the other handle. See also Article 29 Data Protection Working Party, 'Guidelines on the right to data portability' (2016) <http://ec.europa.eu/newsroom/document.cfm?doc_id=44099> pp 6f, as endorsed by the EDPB. On the other hand, recent communication of the EDPB and the EDPS seem to suggest a different view. In one of their joint decisions they proclaim that the Data Act would 'in practice likely extend to entities other than the data subject'; ostensibly different than under the GDPR, see European Data Protection Board and European Data Protection Supervisor, 'DPB-EDPS Joint Opinion 2/2022 on the Proposal of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)' (2022) <https://edpb.europa.eu/system/files/2022-05/edpb-edps_joint_opinion_22022_on_data_act_proposal_en.pdf> para 14. A more thorough investigation into receiving data controller's standing and legitimation to make requests on the data subject's behalf is left for another time. In the meantime, the interested reader may consider Teodora-Lavola-Spinks and Daniela Spajic, 'The broadening of the right to data portability for Internet-of-Things products in the Data Act: who does the act actually empower (Part II) (2022), <https://www.law.kuleuven.be/citip/blog/the-broadening-of-the-right-to-data-portability-for-internet-of-things-products-in-the-data-act-part-ii/>.

13 Intercultural Differences in Autonomy Perception and Persuasion Envelopes

13.1 Introduction

To be autonomous is to be self-governed. What it means to be a self-governed human is subject to debate, as amply outlined in Part I of this thesis. Both within a cultural sphere and beyond, different (and perhaps implicit) preconceptions on what ought to be the necessary and constitutive elements of human autonomy may persist. Consequently, when designing and deploying technology that has implications for human autonomy, difficult questions may arise when there is a mismatch between the autonomy-considerations of the deployer of one cultural sphere and the target audience of another.

This section addresses research sub-question 3.3, namely what the impact of cultural embeddings of individuals on efforts to safeguard autonomy is and how this affects intercultural deployment of technology. The subsequent analysis hence extends the previous arguments within this thesis and outlines how intercultural differences in the ways autonomy is conceptualized and subjectively perceived affect the ethics of deployment of algorithmic agents beyond single cultural spheres. This text proceeds as follows*:* In Section 13.2 the text highlights that intercultural differences in autonomy perception are relevant when deploying technology across cultural spheres and introduce two dilemmas that may occur when deploying technology transculturally. In Section 13.3, I argue that human autonomy (in the context of transculturally deployed technology), irrespective of its cultural roots, ought to be understood as self-referential, and propose that harm incurred by actions taken against human autonomy can not only be considered on an objective ethical scale. In Section 13.4, this analysis is then briefly applied to the domain of technology under the use of the theoretical framework of vectors of influence outlined previously already. In Section 13.5, the relationship between cultural factors and persuasion envelopes, i.e., aspects of cultural spheres accessible to persuasive technology is explored. Two main factors of intercultural differences in persuasion envelopes are identified: the timeline of technology introduction and the vulnerability of individuals due to cultural factors. In section 13.6, I suggest ethical limits when deploying algorithmic agents beyond the borders of a cultural sphere and suggest multiple strategies to resolve conflicts in case of a value collision. I conclude in Section 13.7.

13.2 Autonomy as a Subject of Concern in Matters of Interculturality

To contextualize the issues discussed in this section, it is worthwhile to briefly survey the notion (and relevance) of individual autonomy in the context of interculturality as well as recapping[679] general notions of autonomy, and this section is dedicated to this task. Previous analysis has already outlined that the importance of individual autonomy (in some shape) seems often intuitively apparent, if not clearly delimited; likewise, an infringement to the same is noticeable in similar ways. Part I has already outlined that the term tends to be used liberally and often goes undefined, in particular when denoting values worth protecting or risks thereto,

[679] To make this section read easily without too many references to previous analysis and to allow easy tracing of ideas for the reader, some of the arguments and, where appropriate, footnotes from Part I are repeated here in similar form.

and this holds true for matters of technology deployment as well.[680] Ultimately, human autonomy reflects a concept that is, to an extent, valued throughout different cultures but relies on separate connotations and philosophical assumptions in distinct cultural contexts.

13.2.1 Missing Consensus on Conceptions of Autonomy and Possible Implications for Countermeasures

To start, it is again worth restating that there is no true consensus on what the concept of autonomy ought to describe when considering the academic or political discourse accessible in English (wherein this text falls as well), and naturally even less so on a more global scale. This is not to say that current debate does not rely on some shared tenets of autonomy theory. Indeed, there seems to be more of a slant towards some theories than others: contemporary understanding of human autonomy as reflected in current academic discourse in the English language seems deeply rooted in western-centric philosophical theory, as was outlined earlier.[681] However, this slant may become problematic when applying its findings and frameworks to shape technology, which is then deployed in a cultural sphere in which this understanding of autonomy does not track as well. For example, many internet platforms widely used around the globe have their commercial origin within the western cultural sphere. The messenger service WhatsApp and the social media network platforms Facebook and Instagram (all offered by Meta), the multi-service company Alphabet (*inter alia* the provider of Google Search and Google Maps) and the online-retailer and web-service provider Amazon are all mainly situated and embedded in the western cultural sphere, in particular within the United States, but reach into other cultural spheres with their products as well. It is at least plausible to assume, that the products of these technology deployers has been shaped somewhat around the existing discussions and concerns that are characteristic to their cultural spheres, e.g. with respect to their notions of autonomy, privacy, human dignity, etc.[682] In turn, the regulatory framework these products persist in will likely be influenced by the underlying conceptions and values the cultural spheres exhibit. This is of course not unique to the western cultural sphere. Similarly, the e-commerce company Jingdon, the entertainment conglomerate Tencent, the technology company Alibaba and ByteDance (known in particular for the software application TikTok) are all embedded in a certain (different) cultural sphere (and there specifically within the PRC), deploy their technology beyond their origin cultural spheres and consequently may plausibly incur similar effects.

We may conjecture that this has the potential to obfuscate effective analysis of autonomy-related issues and hinder their addressal as it is at least plausible that countermeasures such as legislation, fuelled by intuitive concerns for individual autonomy, are based in part on the existing, visible debate of experts within this field. Recognizing and addressing autonomy risks imposed by technology will likely be influenced (and perhaps guided) by these discussions. (Again, the same would hold true in cases of reverse nature: autonomy-conceptions from other cultural spheres and measures that build upon those may not be all that useful when applied to the western cultural sphere.) Here we may consider as an example the primacy of the individual in "mainstream"-autonomy concepts. Autonomy according to western philosophical doctrine

[680] See only Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression 2018; Council of Europe 2019; Zuboff 2019

[681] See e.g. Sneddon 2001; André et al. 2018; Owens and Cribb 2019

[682] Note that even within the "western cultural sphere", differences may exist. For example, legal privacy and data protection frameworks differ sometimes substantially between e.g. the European Union and the United States. The same then is true for other cultural spheres, e.g. as the western cultural sphere as a whole and e.g. the Japanese cultural sphere, see e.g. Rafael Capurro, 'Privacy. An Intercultural Perspective', *Ethics and Information Technology*, 7.1 (2005), 37–47.

is usually a qualifier that describes the capacity or status of a single individual.[683] Being determined by some phenomenon or entity outside of oneself will usually be considered as a sign of diminished autonomy. However, these theories presuppose the importance of the individual in decision-making *per definitionem.* This may be overly individualistic for some that are embedded in a culture that places less emphasis on individualism in favour of viewing decision making through a collectivistic lens.[684] Social research has suggested that different cultural spheres do differ on these dimensions considerably.[685] It bears pointing out that this would not make the individualistic, "western" accounts of autonomy *wrong* but may merely inhibits their relevance with respect to the analysis of technology in different cultural spheres.[686]

---

[683] As outlined in Section I, current academic discourse (at least within the western cultural sphere) about individual autonomy seems largely based upon or in reference to the work of Henry Frankfurt (1971), and subsequently Gerald Dworkin (1988). Both conceptualized individual autonomy as being constituted by coherence between an individual's mental system to each other and with respect to the actions Sneddon, *Autonomy*; James Stacey Taylor, 'Introduction'., creating highly abstract and intrinsically focused autonomy concepts that have been denominated as *structural* theories. As a response, some theorists (see e.g. Diana Meyers; Baier 1985; Brison 2000; Christman 2004; Oshana 2005; Buss 2005) have developed theories of *relational* autonomy that emphasizes or necessitates stronger consideration of an individual's relationships to their environment and other entities. This may manifest itself in the implementation of factors pertaining to individuals' status as "emotional, embodied, desiring, creative and feeling" entities as well as tracking external structural factors such as oppressive socialization and social relationships Stoljar. This diversification of thought was explicitly motivated to imbue or contrast the existing "masculine" debate with a broader horizon beyond, deeming the existing frameworks as insufficient or holding implicit values that ought to be rejected for not being considerate (see e.g. Mackenzie and Stoljar 2000; Friedman 2003). Irrespective of the theoretical underpinnings, autonomy is generally considered to be of intrinsic value Dworkin, *The Theory and Practice of Autonomy*; Sneddon, *Autonomy*.. It often follows that enabling or exercising one's autonomy is generally regarded as a morally good act, while constraining or failing to respect one's autonomy is generally considered to be an immoral act, other factors notwithstanding. This is visible in fields which have elevated individual autonomy to a central tenet, such as bioethics (see e.g. Rendtorff 2002; Taylor 2010; Jennings 2016; Reis-Dennis 2020). Affirming the relevance of this domain is empirical evidence that suggests individuals' perceiving themselves to exercise their autonomy incur (mental) benefits André and others; Steven Joffe and others, 'What Do Patients Value in Their Hospital Care? An Empirical Perspective on Autonomy Centred Bioethics', *Journal of Medical Ethics*, 29.2 (2003), 103–8; Shi Yu, Chantal Levesque-Bristol, and Yukiko Maeda, 'General Need for Autonomy and Subjective Well-Being: A Meta-Analysis of Studies in the US and East Asia', *Journal of Happiness Studies*, 19.6 (2018), 1863–82 <https://doi.org/10.1007/s10902-017-9898-2>.. Instruments of law and policy too tend to recognize and protect (aspects of) individual autonomy valued by this research, as manifested in the concepts of e.g., contractual autonomy, informational self-determination or privacy protection. Most explicitly, many international ethical guidelines or principles aiming to govern artificial intelligence prominently elevate autonomy as something worth protecting Jobin, Ienca, and Vayena; Stefan Larsson, 'On the Governance of Artificial Intelligence through Ethics Guidelines', *Asian Journal of Law and Society*, 7.3 (2020), 437–51; Thilo Hagendorff, 'The Ethics of AI Ethics: An Evaluation of Guidelines', *Minds and Machines*, 30.1 (2020), 99–120 <https://doi.org/10.1007/s11023-020-09517-8>.. Naturally, these fields also grapple with the issue that despite autonomy being of vital importance there is no consensus on how exactly the concept ought to be understood (see e.g. Jennings 2016; Dive and Newson 2018). Note also the concerns within these fields about issues pertaining to intercultural autonomy perception, see e.g. Roy Gilbar and José Miola, 'ONE SIZE FITS ALL? ON PATIENT AUTONOMY, MEDICAL DECISION-MAKING, AND THE IMPACT OF CULTURE', *Medical Law Review*, 23.3 (2015), 375–99 <https://doi.org/10.1093/medlaw/fwu032>..

[684] See e.g. with respect to informational autonomy Capurro 2005

[685] See e.g. Schimmack et al. 2005; Hofstede 2011

[686] Naturally, this thesis suffers from the very phenomena just outlined. Academic publishing has long been more prevalent in the western cultural sphere. Authors from non-western cultural spheres may face linguistic obstacles to publish in the existing venues Miguel Clavero, '"'Awkward Wording. Rephrase'": Linguistic Injustice in Ecological Journals', 2010. If other publication venues (such as more "local" journals) are used, non-English publications may address important issues of existing discourse but be less likely to find traction. Geographical location seems to play a role in the prevalence / visibility gap within academic discourse as well Ken Hyland, 'Academic Publishing and the Myth of Linguistic Injustice', *Journal of Second Language Writing*, 31 (2016), 58–

13.2.2 Contextuality of Autonomy Considerations and Relevance of Autonomy Discourse

Indeed, the relevance of autonomy research seems highly contextual. We may recall that autonomy research is arguably an exercise of stipulative definition.[687] There is little epistemic recourse to assess the validity of one account of autonomy over another; indeed, it seems difficult to imagine that an account of autonomy can be correct or valid in the meaning of it being true. Beyond pure academic novelty, the relevance of autonomy research comes then in developing frameworks, toolsets and vocabulary to understand intuitive concerns pertaining to autonomy caused by specific phenomena (such as emerging technology). Here then lie the main issues with a mono-cultural focus: it threatens the relevance of the discourse at large. It is at least conceivable that the relevance of autonomy and autonomy as a concept is subject to cultural differences (and this assumption is supported by a rich body of empirical research into autonomy-derivative concepts). If the implicit cultural assumptions that may underlie current autonomy discourse do not apply to people in major cultural spheres, and the respective theories are not adequately scoped, they risk becoming less relevant and precise. This holds true for the derivative analysis of ever more prevalent autonomy-undermining technology as well.

13.2.3 Examples of Intercultural Flashpoints of Autonomy Differences in the Context of Technology Deployment

How may such a conflict between different autonomy conceptions embedded in different cultural spheres materialize? Within this text I mainly highlight two (separate) problematic situations. First, there may be a conflict between different (implicit) autonomy conceptions that are either merely substantially different or even conflicting. We may call this the ethical foundation-dilemma. Second, the inherent risk of technology to (the same or similar concepts of) autonomy may be different within different cultural spheres. We may call this the actual vulnerability-dilemma. To illustrate these issues, we may imagine hypothetical situations in which an entity, embedded and originating from one cultural sphere, aims to deploy technology in a different cultural sphere with corresponding differences in values (in particular but not limited pertaining to individual self-government) and technology exposure.

To encompass the first situation we may conceive a company using a software application to inform its users, and only them, (accurately) about their health for the purposes of ensuring informed consent for a later surgical procedure, and shielding this information from any other individuals, including the user's relatives etc. for the purposes of privacy. This clearly tracks well with individual autonomy pertaining to decisions regarding their own health, as it is understood in modern (western) bio-ethical frameworks. Indeed, from this position, sharing this type of information without an individual's consent would be considered a breach of their autonomy (and privacy). However, it is entirely conceivable that when deploying this technology into a different cultural sphere, the safeguards of the technology, i.e. it's privacy features, may not track well with or even run opposed to how an individual's autonomy is perceived. Within a different cultural sphere, there may be no expectation of excluding certain individuals from obtaining this information, or even an expectation that other individuals ought

69 <https://doi.org/10.1016/j.jslw.2016.01.005>. This phenomenon works both ways, as the author of this section, just as all other scholars in this field, faces similar constraints to access potentially existing discourse in other cultural spheres (see for a discussion and recommendations for this issue e.g. ÓhÉigeartaigh et al. 2020).

[687] This here takes the meaning of theorists assigning the terminology of autonomy and its derivative concepts to phenomena on the basis of intuition and (internal) logic of their constructed framework but without reference to ground-truth.

to obtain this information first. Confronting (only) the intended user with this health-related information may be seen as problematic generally or even autonomy-undermining. At best, these safeguards are not needed, but may perhaps even be detrimental to the self-understanding of one's own autonomy.[688] Should the deploying entity then adhere to the ethical standards set within its (original) cultural sphere, and assume a posture of individual informed consent? Or should the deploying entity adhere to the (culturally derived) different (ethical) expectations of the target cultural sphere? Further difficulties arise, when under the values of one cultural sphere, the presumptions of the other cultural spheres are considered unethical to begin with. In the present case, this may be applicable in case this type of privacy is only awarded to individuals of certain features, such as sex, gender or ethnic background; a distinction which may be in breach of equality principles of the respective other cultural sphere. Put bluntly, western bio-ethic frameworks may see egregious violations if their notion of informed consent is only denied to e.g. women, foreigners or individuals of lower socio-economic classes.

To encompass the second situation, we may imagine a messaging platform software application that allows quick resharing of information, including misinformation. Assuming that the veracity of information is considered to be relevant for the exercise of individual autonomy in the deploying entity's cultural sphere and the target cultural sphere, problems may arise if misinformation becomes (more) viral, persuasive and hence problematic in the target cultural sphere. Indeed, the deployer may have not taken safeguards that may be considered to unnecessarily substitute the individual responsibility in engaging with online information such as moderating and filtering information or applying contextual notifications that hint at the assumed lack of veracity (perhaps because it was not necessary in its cultural sphere). However, the same safeguards may for some reasons be necessary to achieve the same (acceptable) results of curbing the spread of misinformation in the target cultural sphere. Does the technology deployer ought to alter its safeguards for the target cultural sphere, even though the ethical frameworks it operates in may allow for shifting the responsibility onto the individual in both cultural spheres?[689]

These dilemmas highlight issues that even benevolent technology deploying entities may encounter. However, they also highlight a different issue. Malignant deployers may, insofar this would be of benefit to them, arbitrage different cultural and ethical frameworks and justify potential harmful behaviour by reference to local cultural value-sets.[690] This risk heightens the relevance of the present discussion.

---

[688] Ingrid Hanssen, 'From Human Ability to Ethical Principle: An Intercultural Perspective on Autonomy', *Medicine, Health Care and Philosophy*, 7.3 (2005), 269–79 <https://doi.org/10.1007/s11019-004-9032-1>; M. A. Kara, 'Applicability of the Principle of Respect for Autonomy: The Perspective of Turkey', *Journal of Medical Ethics*, 33.11 (2007), 627–30 <https://doi.org/10.1136/jme.2006.017400>.

[689] The same issue is also exemplified within the nexus of persuasive digital advertising. It is entirely conceivable that certain advertising strategies (and the underlying personalization and data collection) are more effective within one cultural sphere than another, and that the intuitive concerns that arise out of highly effective advertising are triggered by the success rate in one cultural sphere but not the other. Multiple situations are possible. The safeguards around the deployment of this highly persuasive technology may have been shaped by the (lower) rate of success within one cultural sphere. Alternatively, the underlying data used for profiling and personalization may be deemed unproblematic in both cultural spheres, but the effectiveness that is derived from this unproblematic dataset in the target cultural sphere may be considered problematic in the origin cultural sphere based on its higher success. In case the underlying data collection is considered differently in both cultural spheres, the example of advertisement also connects to the first issue highlighted, i.e. the conflict between autonomy-conceptions.

[690] (see e.g. Wong 2020)

13.3 Advantages and Disadvantages of Conceptualizing Self-Referential Autonomy

Within this section I suggest that the individual, and culturally derived, self-perception of one's autonomy is a key factor when considering ethical obligations of technology deploying entities that operate across cultural boundaries, but this assertion is not trivial and requires justification. To conclude the theoretical underpinning of the ethical obligations outlined in the final sections of this section below, a pragmatic argument is made based on harm incurred by autonomy constraints.

13.3.1 Autonomy Constraints and Harm

Analysis of individual autonomy is useful in part because it may allow deeper understanding of autonomy constraints and harm done as a consequence of such constraints.[691] There are two ways in which one can understand harm in this context. Firstly, we may consider any illegitimate violation of a moral principle as harm. If we consider autonomy as having intrinsic moral value, then constraining autonomy may incur the harm that follows from breaching the respective moral obligation that arises out of autonomy's intrinsic value. The harm is the violation of the moral principle itself and arises not to any individual but abstractly as a function of the moral obligation to the moral framework assumed. This type of harm tracks well with the idea of autonomy as an objective, globally valid, or omnirelevant concept.[692] Assuming for example that in the informed-consent case laid out above, an individual's autonomy is violated if they are not given sufficient information by the technology application to exercise their consent, then harm would be incurred in cultural spheres by users that do not perceive this as necessary to their capacity for self-government. The second type of harm can be understood as harm that arises to an individual as a consequence of an autonomy constraint imposed on them. Harm then is not constituted by breach of a moral obligation but by the effect an autonomy constraint has on an individual. This effect can be considered from an objective or subjective viewpoint. When assuming an objective viewpoint, i.e., classifying the harm from an external position, the resulting analysis suffers from the same epistemic challenges to find a "true" concept of individual autonomy (and to ascribe moral value to it) as the previous conception of harm. This approach can also create the intuitively unsatisfying situation in which an individual may be considered to be constrained in their autonomy objectively (and perhaps morally), but based on their reflective and informed assessment, does not consider their autonomy to be violated themselves.[693] To continue the informed consent-example above, we may still consider the exposure to pertinent health information through the software application as vital for exercising autonomy, and the same may be true for the default state of exclusion of others of this information. This would again hold true even if this is against the explicit or internalized preferences or autonomy-conceptions of the user themselves. Pragmatically however, it seems more fruitful to assume a subjective viewpoint in which harm caused by an autonomy constraint is precisely that harm which the individual perceives as harm or would perceive as harm would they have had sufficient information about the constraint.[694] In this

[691] This was outlined as the primary and secondary goal of autonomy research in Section 1.6

[692] as is widely propagated, see e.g. Young 1982; Dworkin 1988; Darwall 2006; Sneddon 2013

[693] This issue ties into the distinction between procedural and substantive autonomy theories noted in a previous footnote. One may consider the scoping of one's autonomy as an exercise of autonomy which should be fully at an individual's disposal, while others may want to limit such autonomy-foregoing measures. Under this viewpoint, the posture taken within this text would be considered procedural.

[694] This does not necessitate a moral relativistic standpoint but is merely neutral towards any question of morality. At the same time, this results in the approach here propagated leaning towards some form of utilitarian framework. However, the mitigation strategies discussed in the final section relativize this statement somewhat.

case, autonomy-specific harm would not be incurred by the technology user, if the sensitive information is shared to other persons, in case the user does not or would not consider this potential breach of privacy as a breach of constitutive element of autonomy at all.

13.3.2 Self-Referential Autonomy

If this premise is accepted, an individual's autonomy and their constraints become self-referential. By an extension of an individual's exercise of autonomy, an individual shapes the subjective scope of their autonomy and characterizes certain phenomena as constraints against the same.[695] If an individual does not perceive specific types of external influence as constraints to their autonomy, then they may do so in accordance with any potentially wider autonomy at their disposal. The disadvantage of this viewpoint is that the absence of a more objective standard does not account for intuitive concerns pertaining to hard structural autonomy constraints such as oppression or indoctrination [696]. In the informed-consent case, there may intuitive uneasiness about denying what are considered to be constitutive elements of autonomy within the deployer's cultural spheres such as privacy related to health data only to some. While a cultural sphere may have intrinsic justifications for differentiation between genders, ethnic backgrounds or socio-economic classes, the same justification may not hold up in the deployer's cultural sphere. A deployer may be in full compliance with a target cultural sphere when denying informational privacy e.g. to women or to children of a certain age, but this may be in breach of the expectations of the cultural sphere the deployer is originally embedded in. Reversely, in the context of interculturality, a strict adherence to such a subjective view may thus allow agents of one cultural sphere to impose measures that they would consider to be harmful onto individuals of another cultural sphere under the argument that it would not constitute an autonomy constraint within the second cultural sphere. I return to the implications of this and how this consequence of strict adherence can be defused later in this section when discussing ethical obligations of technology deploying entities.

13.3.3 Implications of Subjective Autonomy Accounts

The subjective view is useful however because it is practically relevant both from a structural and individual perspective. As mentioned previously, when political entities with legislative powers attempt to safeguard autonomy, they are likely to be influenced by a pseudo-objective framework that is in close interplay with the subjective scopes of autonomy of their stakeholders. These frameworks may be intellectually motivated as are the examples above, but they may also be motivated by persisting cultural or religious norms.[697] The same can hold true for private entities acting in compliance with such legislation or yearning for a competitive advantage by adhering to the (collective) subjective autonomy framework that apply to the individuals they interact with. For example, foregoing data collection may increase the attractivity of a product in a market in which privacy (as a prerequisite of autonomy) is valued highly.[698] With respect to the individual perspective, individuals will often benefit from exercising what they perceive to be autonomy.[699] At the same time, being led to exercise a

[695] (see e.g. Beauchamp and Childress 2001, p. 60)
[696] Hanssen, 'From Human Ability to Ethical Principle: An Intercultural Perspective on Autonomy'.
[697] Reviglio and Alunge 2020, p. 14f)
[698] Sören Preibusch, Dorothea Kübler, and Alastair R Beresford, 'Price versus Privacy: An Experiment into the Competitive Advantage of Collecting Less Personal Information', *Electronic Commerce Research*, 13.4 (2013), 423–55.
[699] André and others; Dhar and Wertenbroch; N. T. Feather and J. G. Simon, 'Causal Attributions for Success and Failure in Relation to Expectations of Success Based upon Selective or Manipulative Control', *Journal of Personality*, 39.4 (1971), 527–41 <https://doi.org/10.1111/j.1467-6494.1971.tb00060.x>.

foreign notion of autonomy and to adhere to its prerequisites can create distress [700]. Considering this, maintaining a purely objective view of autonomy risks overlooking consequences of perceived loss of autonomy while emphasizing aspects of autonomy that the respective individual may have consciously yielded.

13.4 Recap: Autonomy and Technology

The implications of intercultural differences in autonomy conceptions, as outlined in the previous sections, and the salience of a self-referential view of autonomy is not unique to the intersection of autonomy and (information) technology, but technology highlights these conflicts in perhaps unique ways. Indeed, information technology, insofar that it is part of the informational pipeline (i.e. information origin, information transfer infrastructure and processes, and information recipient) leading to autonomy constraints, becomes an important subject for analysis when considering an individual's autonomy. We can again use the concept of a vector of influence for the current analysis: The capability of an informational agent to engage with individuals through (personalized) information aligned with their cognitive biases and vulnerabilities and along the goals inherent in the algorithm characterizes the vector of influence along which the information is transferred. In the previously elaborated informed-consent and viral information - examples, the vectors of influence would encompass all processes from the software application, its information processing (e.g. determining the health outlook or ranking information for engagement), the transfer of information, the process of intermediation into human-readable form (e.g. on a smartphone screen or personal computer) and the reception of this information by the user.

How technology is effecting an autonomy constraint is hence both dependent on the technology on one side and the receiving individual on the other side. This serves to highlight that there are cultural differences in the way individuals perceive, interact with and reflect on information transferred via technology, and as a result cultural differences that are displayed within the relevant vector of influence. There may also be a difference in the effect this has on the individual, and how this effect will be classified with respect to its (subjective) moral acceptability. This seems especially relevant as certain technological services reach far beyond the cultural sphere they were initially focused on.[701]

13.5 Potential Cultural Impact on Technological Persuasion Envelope

Having established the abstract relevance of intercultural considerations when it comes to safeguarding (or constraining) autonomy and having highlighted that this seems especially pertinent when considering information technology as the constraining factor, the issue of culture and its respective spheres has so far been treated as a black box in this text. We may recall the two main dilemmas identified by the reference to the informed consent- and viral information- examples earlier: the ethical foundation-dilemma (conflicts may arise between different culturally embedded autonomy conceptions that are either different or even conflicting) on the one hand, and the actual vulnerability-dilemma (if the inherent risk of the deployed technology is more intense within different cultural spheres) on the other hand. Clearly both of these are dependent on some cultural factors. It becomes a necessity then to

[700] Hanssen, 'From Human Ability to Ethical Principle: An Intercultural Perspective on Autonomy'.
[701] Daniele Archibugi and Simona Iammarino, 'The Globalization of Technological Innovation: Definition and Evidence', *Review of International Political Economy*, 9.1 (2002), 98–122.

analyse the cultural factor closer, and consequentially unpack the aforementioned black box to analyse these dilemmas.

13.5.1 Culture and Persuasion Envelopes

Culture is the compound shared or learned characteristics of a group of individuals including their traits, norms, laws, values, patterns of behaviour, experiences, knowledge, and emotional responses.[702] The connection to the dilemma posed by the informed-consent example is obvious. If different cultural spheres have different value-sets and superimpose those value-sets on their realities, different ethical preferences are the result. Individuals in one cultural sphere not considering informational privacy as a prerequisite to their autonomy (i.e. their autonomy not constrained if they are not afforded informational privacy) clearly differ by some part of their value-set from individuals of another cultural sphere that disagree with their assessment. The main ethical question within the ethical foundation-dilemma then remains on how to weigh and consider their different conceptions and value-sets, in particular in cases of conflict.

But the viral misinformation-example highlights a different type of dilemma. Indeed, we may consider individuals of two cultural spheres in which their ethical distain for misinformation is fully congruent, but who have different levels of vulnerability towards the misinformation-disseminating technology. This too may be assigned to cultural factors, namely either peculiarities that are intrinsic to the cultural sphere itself (i.e. cultural factors in the narrow sense of the word such as scepticism towards technology) or factual circumstances that are attached to the cultural sphere (such as factual access to technology). Different cultural contexts and their lived realities offer a different set of phenomena, hence different vectors of influence, that may be more or less conducive to their persuasion-*enveloping* through technology (i.e. the accessibility through persuasive technologies[703].[704] In other words, technology may find different attachment points, venues of communication and opportunities to persuade based on the lived realities of individuals within a cultural sphere. As a consequence, insofar technology

[702] Aliaksandr Birukou and others, 'A Formal Definition of Culture', 2013, pp. 1–26 <https://doi.org/10.1007/978-94-007-5574-1_1>; R Boyd and P J Richerson, 'Culture and the Evolutionary Process. Chicago (The University of Chicago Press) 1985.', 1985; M Mead and R Métraux, 'The Study of Culture at a Distance. Chicago: Univ' (Chicago Press, 1953).

[703] see e.g. Floridi 2011; Robbins 2020

[704] In this Section, the term cultural sphere is generally used to describe rather large cultural boundaries (e.g. the western cultural sphere, etc.) for the purpose of poignant examples. However, the term (here) ought to be understood in a more encompassing way: The borders of a cultural sphere may exceed, match or not even reach the boundaries of a single country. For example, the uncontacted people of the Toromona, living in Bolivia, span between them a cultural sphere that is significantly smaller than the state they reside in geographically. Another simplification for the purpose of readability is the term of cultural sphere as a pseudo-atomic concept. This is naturally not the case: different layers of culture may permeate between different groups of individuals in a single geographical location. Individuals living within European Union member states may display idiosyncratic lived realities with respect to some aspects in their lives that are determined by their country (e.g. in matters of rite and tradition, denominating a separate cultural sphere *with respect to these phenomena*) but may fit in nicely into a larger cultural sphere with respect to other aspects (e.g. use of technology). To this end the term cultural sphere used in this section may be understood to point at the cultural sphere *with respect to the use of and resilience against technology*. Lastly, while the term cultural sphere may invoke a calcification of living realities, it is not static but subject to constant change. However, changes to the lived realities of individuals do not immediately have to invalidate the existing persuasion-*envelope* / change substantially the vectors of influence at interest here. Ultimately, the definition of cultural spheres at a granular level may become necessary for the purposes of implementation of autonomy safeguards against intercultural deployment of technology, but for the understanding of the concepts propagated in this section, the aforementioned simplifications suffice.

exerts shaping force onto this lived reality (i.e. envelopes it)[705], the different lived realities of different cultural spheres may exhibit varying levels of resistance to such attempts of enveloping. Inherent incompatibility of a culture's exercise of autonomy with existing autonomy-undermining technology might lead to an increased resistance, for a time, due to a narrower envelope and consequently weaker vector of influence. In contrast, increasingly persuasive technology, fuelled by faster computing devices, increased sophistication of underlying algorithms and extended reach into the physical domain via an increase in connected interfacing devises to input and output information broadens the persuasion envelope accessible by algorithmic agents. The respective envelope is hence not uniform but subject to cultural characteristics.

13.5.2 Digital Literacy

Understanding a vector of influence encompasses understanding the characteristics, including any informational processing flaws and biases, of the respective affected individual. A term that has been developed to cover most characteristics pertaining to knowledge, competence and understanding of technology ensuring effective and critical interaction is *digital literacy* [706]. As a result of some sort of experience, in particular education and interaction, individuals ought to build up knowledge and awareness that aid in interacting with information technology. Digital literacy has recently come into particular focus due to the impact of online misinformation [707] and the varied susceptibility to that phenomenon of different groups [708]. Because digital literacy is part of a given vector of influence, intercultural differences in digital literacy factor into the effectiveness of a given autonomy constraint.

There are at least two factors that can play into differences in digital literacy among different cultural spheres. First, differences may be present due to external processes. Individuals within a cultural sphere may have been exposed to a certain type of information technology over a period of time, during which individuals may gain the ability to track and internalize the iterative changes to the interaction characteristics while building up literacy pertaining to the progressively evolving technology. Second, we can conceive differences that stem from intrinsic peculiarities within a cultural sphere. A cultural sphere may foster or inhibit mental processes that are required to further digital literacy with respect to certain types of information technology for certain contexts.

13.5.3 Gradual and Temporal Factors in Technology Introduction

Information technology did not emerge at the same time and with the same intensity at different regions and cultural spheres, and at the time of writing has not been established equally around

[705] With the words of Milano et al. 2020: "Any recommendation is a nudging, and any nudging embeds values."

[706] David Buckingham, 'Defining Digital Literacy', in *Medienbildung in Neuen Kulturräumen* (Wiesbaden: VS Verlag für Sozialwissenschaften, 2010), pp. 59–71 <https://doi.org/10.1007/978-3-531-92133-4_4>; Ivor F. Goodson and J. Marshall Mangan, 'Computer Literacy as Ideology', *British Journal of Sociology of Education*, 17.1 (1996), 65–79 <https://doi.org/10.1080/0142569960170105>.

[707] S. Mo Jones-Jang, Tara Mortensen, and Jingjing Liu, 'Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't', *American Behavioural Scientist*, 65.2 (2021), 371–88 <https://doi.org/10.1177/0002764219869406>; André Calero Valdez, 'Human and Algorithmic Contributions to Misinformation Online - Identifying the Culprit', 2020, pp. 3–15 <https://doi.org/10.1007/978-3-030-39627-5_1>; Emily K. Vraga and Leticia Bode, 'Defining Misinformation and Understanding Its Bounded Nature: Using Expertise and Evidence for Describing Misinformation', *Political Communication*, 37.1 (2020), 136–44 <https://doi.org/10.1080/10584609.2020.1716500>.

[708] see e.g. Nygren et al. 2020; Brashier and Schacter 2020

the world. For example, the usual underlying requisite condition of internet access varies widely in its fulfilment between different regions of the world, a phenomenon dubbed the digital divide.[709] This means that individuals within these regions and their respective cultural spheres experience different timelines with respect to technology exposure. An adult individual e.g. in a *western* cultural sphere is much more likely to have experienced a gradual development of many of the algorithmic agents and their interfaces constraining autonomy today than an individual in a different cultural sphere. The first would have had a comparatively large period of time to observe the iterative improvement of internet infrastructure and services, to hear the developing discourse around this phenomenon and interact with previous, and arguably less persuasive precursors of current technology. As a result, this individual is more likely to have a comparatively high level of reflection with respect to their interaction with information technology. The second, however, would have been faced with fully developed and highly persuasive technology at the time of information technology introduction. There is less historical context and less opportunity to negotiate the relationship between themselves and the now quickly emerging technology. This inexperience, combined with the persuasiveness of the now fully developed infrastructure and services can shape individual's perception, and consequently actions sometimes with severe consequences such as in the case of violence induced by viral rumours propagated via Internet services.[710]

If defensive mental mechanisms are built up by individuals in certain cultural spheres due to their gradual exposure to persuasive technology, this can lead to certain interferences with an individual's autonomy to be seen as irrelevant due to a lack of consequences. A populace that has sufficient tools to engage with such technology to their benefit may have internalized certain trade-offs inherent in modern, data-hungry technology. To give an extreme example, an individual may accept their email-address being accessible online, risking exposure to semi-automatized, fraudulent spam emails. This may be because they have confidence that they can filter out any such attempts to influence them, e.g., by misinformation, to take actions such as wiring money under false pretences. Due to their digital literacy, autonomy constraints along this vector of influence become less likely. At the same time, fully developed persuasive technology can be spontaneously introduced into a different cultural or demographical setting. If individuals do not have the chance to develop appropriate defensive mechanisms, their autonomy is consequently under a more intense threat, not unsimilar to the introduction of foreign pathogens into an unprepared biosphere. In addition to potential unpreparedness of an individual level, the informational infrastructure at the disposal of an individual is also dependent on the timing of introduction of information technology. For example, companies that have been established and grown during the rise of the Internet in certain cultural spheres, may have become powerful entities when accessible information technology is introduced in a different sphere. Consequently, these companies may now be in a position to shape the market for Internet access proactively, such as by subsidizing internet access to their own services, a practice known as *zero-rating* [711]. These practices alter the informational landscape practically accessible to individuals in the late-adopting cultural sphere, as other sources of information (e.g., websites of newspapers) may be more difficult to access than the services privileged by the market shaping behaviour (e.g. messenger or social network services).

[709] Rogers; Chinn and Fairlie; International Telecommunication Union, 'Measuring Digital Developments: Facts and Figures 2020', 2020.
[710] see e.g. Whitten-Woodring et al. 2020
[711] Linnet Taylor, 'From Zero to Hero: How Zero-Rating Became a Debate about Human Rights', *IEEE Internet Computing*, 20.4 (2016), 79–83 <https://doi.org/10.1109/MIC.2016.88>; Guy Thurston Hoskins, 'Beyond "Zero Sum": The Case for Context in Regulating Zero Rating in the Global South', *Internet Policy Review*, 8.1 (2019) <https://doi.org/10.14763/2019.1.1392>.

13.5.4 Vulnerability and Resilience due to Cultural Factors

Just as external factors pertaining to the development of technology, cultural factors may change the scope of the persuasion envelope and affect the vector of influence an autonomy constraint can be imposed with. A distinction can be made between external and intrinsic consequences of cultural influence. First, culturally determined characteristics of individuals and their group can install, through actions of the group members, an environment that affects the interaction with information technology.[712] To give an extreme example, a society such as the traditional Christian Amish people that has access to but rejects most information technology outright gives little room to have their individual's autonomy constrained through such technology. Considering the viral misinformation example, it seems unlikely that this technology (and the corresponding algorithmic agents) will be highly effective in exposing individuals in this sphere to misinformation and as a result constrain their autonomy, as they may either not use it or interact with information derived from it with heightened distrust and scepticism. Furthermore, different cultural spheres imbue different societal entities, such as civic or religious leaders or collective communities as a whole, with importance and consequent trust; as a result these entities and their positioning towards autonomy preservation and technology are highly relevant to contextualize potential autonomy-constraining technology from an individual's viewpoint.[713] Second, culture may be reflected in the intrinsic psychology of an individual, creating resilience or vulnerability towards certain persuasion strategies but also via recontextualizing (subjectively) what autonomy there is to constrain. Culture can affect the trust individuals have towards information technology.[714] Naturally, lack of trust can make individuals more sensitive to otherwise effective, potentially autonomy-undermining practices such as presenting information in a highly personalized way.[715] Acceptance of such technology is also dependent on perceived usefulness, i.e. the utility towards certain objectives; whereas such objectives are likely to be determined at least in part by cultural influences as well [716]. Different cultures may affect certain types of information reception capabilities of individuals including at least certain cognitive biases that may be exploited as part of a vector of influence.[717] Similarly, different cultures may also change the vulnerability or resilience of average sub-cohorts of their populace; with the differences between different cohorts being more pronounced in some cultural spheres than in others. [718] Finally, as previously already highlighted (and again relevant to the informed consent-example), cultural influence may affect an individual's perception of what their autonomy consists of and what ought to be classified

---

[712] Sudhir Rama Murthy and Monto Mani, 'Discerning Rejection of Technology', *SAGE Open*, 3.2 (2013), 215824401348524 <https://doi.org/10.1177/2158244013485248>; Detmar Straub, Mark Keil, and Walter Brenner, 'Testing the Technology Acceptance Model across Cultures: A Three Country Study', *Information \& Management*, 33.1 (1997), 1–11.

[713] see e.g. Barua et al. 2020

[714] see e.g. Cyr et al. 2005; Vance et al. 2008

[715] Bleier and Eisenbeiss, 'The Importance of Trust for Personalized Online Advertising'.

[716] Fred D Davis, 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology', *MIS Quarterly*, 1989, 319–40.

[717] Christine Ma-Kellams, 'Cultural Variation and Similarities in Cognitive Thinking Styles Versus Judgment Biases: A Review of Environmental Factors and Evolutionary Forces', *Review of General Psychology*, 24.3 (2020), 238–53 <https://doi.org/10.1177/1089268019901270>; Hong Chen and Todd Jackson, 'Are Cognitive Biases Associated with Body Image Concerns Similar between Cultures?', *Body Image*, 2.2 (2005), 177–86 <https://doi.org/10.1016/j.bodyim.2005.03.005>; Gerd Gigerenzer, Klaus Fiedler, and Henrik Olsson, 'Rethinking Cognitive Biases as Environmental Consequences', *Ecological Rationality: Intelligence in the World*, 2012, 80–110.

[718] Kiemute Oyibo, Rita Orji, and Julita Vassileva, 'The Influence of Culture in the Effect of Age and Gender on Social Influence in Persuasive Technology', in *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (New York, NY, USA: ACM, 2017), pp. 47–52 <https://doi.org/10.1145/3099023.3099071>.

as an autonomy constraint. For example, cultural spheres that exhibit heightened acceptance for hierarchy (i.e. *power distance* in the parlance of Hofstede)[719] may also be more tolerant towards some interpersonal controlling behaviours, and individuals may concede these as in line with their autonomy, while this may trigger autonomy concerns in individuals from different cultural spheres.[720] Similarly, cultural spheres may impose restrictions on some of their members (e.g., women or members of disenfranchised minorities) with respect to the extent of the range of decisions or actions that they may take. These restrictions may be seen by the restricted individuals as in line with their autonomy, while attempting to lift the restriction may be perceived as going against their autonomy and cause concerns.[721] More generally, within some cultural contexts, an individual's independence of other individual's influences (if feasible) may be considered as a constitutive element of their autonomy, while an endorsement of another's influence by the affected individual may be sufficient in other cultural spheres to retain autonomy.[722] Acceptance of certain phenomena as being compatible with an individual's autonomy, and the resulting (re-)shaping of the concept of autonomy instead of general devaluing or rejection of autonomy by the individual in these cases is likely due to the fact that individuals in most cultural sphere (consciously) value some sort of autonomy [723].

13.6 Ethical Deployment of Persuasive Technology

The previous analysis has highlighted some of the most relevant cultural underpinnings of the ethical foundation- and actual vulnerability-dilemmas through the application of the concept of vectors of influence. These dilemmas are relevant, as the underlying value at risk, i.e. autonomy in whatever relevant permutation, is generally relevant. At the same time, this concept of autonomy too is subject to intercultural differences with respect to its scope, value and utility. Together, this creates issues both academic and practical.[724] Within this final section, the text suggests potential ethical risk mitigation strategies for technology deploying entities when encountering these dilemmas.

Section 13.3 has argued that a self-referential conception of autonomy is generally useful and appropriate in addressing these dilemmas. As outlined, there also seem to be real implications in using *foreign* conceptions of autonomy to guide policy decisions and to structure interactions between individuals within a cultural sphere. In edge cases, enforcing autonomous behaviour under one theoretical or societal account may be considered to impede an individual's autonomy under a different framework. From a pragmatic point of view, much is therefore in favour of considering an individual's autonomy, at least partly, from their subjective point of view, as many of the mental benefits attached to exercising one's autonomy seem to correspond to the approximate conception of autonomy that an individual and its surrounding cultural

[719] Hofstede 1984)
[720] see e.g. Chua et al. 2014
[721] Ingrid Hanssen, 'An Intercultural Nursing Perspective on Autonomy', *Nursing Ethics*, 11.1 (2004), 28–41 <https://doi.org/10.1191/0969733004ne664oa>.
[722] see e.g. Chen et al. 2013
[723] Yu et al.o 2018
[724] For an example of the first, we may consider that western autonomy accounts are typically descriptive and objective, often considering independence as a constitutive factor. Coupled with the assertion that autonomy is something valuable and worthy of preservation, and the often-implicit moral privilege that is awarded to individuals and their actions insofar they are considered to be autonomous, this creates an intuitively unsatisfactory situation of autonomy primacy *per definitionem* of certain cultural spheres (see e.g. Shweder et al. 1997 for the distinction of ethical systems that build on different values than pure autonomy). As mentioned before, the academic obstacles are not focus of this text and are left for later research.

sphere has. At the same time, such an approach may entrench dependencies between individuals that some may find unacceptable.

13.6.1 General Observations

These above tension between subjective (potentially limiting) and objective autonomy accounts is not new. Autonomy research has long recognized the inherent tension between favouring an individual's autonomy *per se* and the consequence that the individual may exercise their autonomy to potentially forgo their autonomy.[725] This dilemma holds true in intercultural contexts all the same,[726] and all of this retains relevance when transferring the analysis to the domain of technology.

Cultural context therefore ought to be recognized when deploying algorithmic agents. Use of a software or similar process that imposes effects on individuals that they would consider an autonomy constraint based on their own perspective and their cultural experiences cannot be exculpated by reference to a different, more lenient cultural sphere in which the relevant measure would not be considered inappropriate, *because* the harm incurred is due to the subjective autonomy experience of the individual. This holds true even when not recognizing separate perceptions of autonomy as valid *per se*. Indeed, respect for an individual's exercise of autonomy (in an objective meaning external to them) by shaping the scope of their autonomy going forward mandates that persuasive technology must be deployed under consideration of cultural context, hence yielding the same outcome.

Another important ethical consideration when deploying algorithmic agents is their effectiveness. Generally, most cultural spheres will recognize susceptibility to external influences, be they technological in nature or otherwise. For example, children and individuals of unsound mind are embedded in legal and practical measures that are meant to protect them from external entities influencing them to act against their interest. At the same time, regulation exists in some cultural spheres to regulate such influence practices and to limit the exposure of individuals to these persuasive measures, as exhibited e.g., by regulation governing advertisements specifically targeted at young children. This concept holds true even for individuals that are not considered by their cultural sphere to be compromised in their ability

[725] In the academic literature, these diverging theories are denoted as procedural and substantive theories respectively.

[726] However, many factors considered as most severe autonomy constraints (in particular by western cultural spheres, such as structural oppression and discrimination against subgroups within a cultural sphere) are somewhat inconsistent with the accession of countries belonging to those spheres to instruments aiming to shape global moral consensus such as the Universal Declaration of Human Rights or the International Covenant on Civil and Political Rights. It is important to note that drawing from existing international legal instruments tends to have limitations. The UDHR for example was accepted in a quorum of 58 states, of which 10 did not vote in favour of it (although of course no state voted against it). However, the rights enumerated within seem to have been translated into many cultural spheres, and recent scholarship has recognized the impact of states of non-western cultural spheres in drafting the instrument (see e.g. Waltz 2001, 2002; Ramcharan and Ramcharan 2019). Ultimately, while one may not ascertain completely that international legal instruments that have been accepted, acceded to or implemented by states from a wide variety of cultural spheres represent moral consensus, it is at least a plausible approximation on some value sets that are widely considered as relevant. To this end, some structural autonomy constraints recognized by e.g. western theories of autonomy also do not properly conform to a set of values recognized by other cultural spheres, in which case the issue lies not in conflicting autonomy perception but in consistency of implementation. The reverse may also very well be true. However, even accepting this, intercultural differences remain.

to safeguard their autonomy. Indeed, the prohibition of subliminal advertising[727], or the prohibition of advertising for some addictive substances present in many cultural spheres highlight that measures thought to be of exceeding persuasive quality may have to be restricted even when considering only adult individuals of sound mind. These considerations translate well into an intercultural context. A measure of persuasive technology may not be limited in its deployment in one cultural sphere because it is not effective enough there. Transferring the same technology into a different cultural sphere, where for some reason effectiveness is increased, ought to incur the obligation of the deploying entity to place limitations on the technology accordingly, at least to the extent that the effectiveness exceeds what would be acceptable in the cultural sphere of the deploying entity.

In the light of the above and in application of the subjective, self-referential autonomy view, this section argues that deploying entities are faced with two ethical limits when deploying algorithmic agents beyond their initial cultural sphere: respect for the subjective, culturally influenced scope and value of an individual's autonomy *and* adherence to the standards present in the deploying entities' cultural sphere.

13.6.2 Defusing Vulnerability Differences through Extension of Protective Considerations

Application to the actual vulnerability-dilemma and the viral misinformation-example is relatively straightforward and yields that technology providers ought to apply the most protective framework called for by either cultural sphere. The main complication in this dilemma stems from the fact, that the misinformation-spreading platform may undermine individual's autonomy within one cultural sphere more effectively (perhaps due to insufficient digital literacy and built-up resilience at the point in time of interest). This may mean that both cultural spheres generally consider certain interactions along the respective vector of influence not as autonomy constraints, but the interaction is considerably more harmful in one cultural sphere. However, here one may recognize that this may very well be a manifestation of differences in technology introduction as outlined in the previous section. Indeed, the more resilient cultural sphere may very well consider technology with the effectiveness in imposing harm as occurring in the more vulnerable cultural sphere as autonomy-constraining. At the same time, the more vulnerable cultural sphere may not consider the effect has harm due to the relative novelty of the intrusion. This may mean that one cultural sphere does not consider a certain type of misinformation presentation (e.g. through non-chronological ranking or social-proofing) as autonomy constraint, as it expects its members to be aware and hence not be unduly affected by it. The other cultural sphere may not (yet) consider this presentation as an autonomy constraint because the vector of influence is not yet fully understood and internalized by its members. In this case, the recourse to the most protective framework must also encompass the translation of actual harm for consideration *as if* it would have occurred within the respective other cultural sphere. The deploying entity must then consider the more resilient cultural sphere would consider the harmful effects of presenting misinformation on an internet platform as an autonomy constraint, and avoid these in the more vulnerable sphere as well.

[727] Nb that subliminal advertising is generally not seen as effective as public concerns would indicate Charles Trappey, 'A Meta-Analysis of Consumer Choice and Subliminal Advertising', *Psychology \& Marketing*, 13.5 (1996), 517–30..

13.6.3 Addressal of Different Ethical Baselines

More difficulties arrive when considering the ethical foundation dilemma and hence onto the informed-consent example. Application of these limits yields that the technology provider ought to accede both to the level of protection granted by ethical frameworks and rulesets of their origin cultural sphere as well as those of the cultural sphere of the application user (and data subject). This results in two possible situations. Perhaps more uncontroversial, if the technology provider's cultural sphere does not impose the importance and confidentiality of health information (in particular as a basis of the recipient's autonomy), but the cultural sphere the user is embedded in does, the provider ought to accede to the higher standard of autonomy protection. This will likely be of lesser issue, as the deployment of technology would likely be regulated in the target cultural sphere in approximation to the (stricter) ethical frameworks. However, insofar the technology provider's cultural sphere requires (again out of concern for the user's autonomy) that health information be disclosed to them and only to them, and the user's cultural sphere does not, the provider still ought to accede to the higher standard, so as to not engage in any conduct that may be characterized as transcultural exploitation.

Insofar these two limits are in conflict with one another, the deploying entity is faced with a difficult choice of which cultural sphere's limit to prioritize. In case of the informed consent-example, this would be the case in which the user (or data subject) would consider the "burden" of being in power of their health information to be contrary to their autonomy and would exhibit ethical preferences for others (e.g. their relatives or spouses) to have full discretion in these matters.[728] This deference may be contrary to the values of the deployer's cultural sphere, e.g. if this is based on unacceptable distinctions between protected classes (e.g. on the basis of sex, gender or socio-economic class), and if the ethical preferences are suspected to be a result of e.g. oppression or indoctrination. In this case, it seems intuitively difficult to recommend adhering strictly to the subjective, self-referential view of autonomy outlined previously and to prioritize solely the "poisoned" ethical preferences of the technology user, lest one runs the risk of reinforcing and acceding to unacceptable structures of oppression.

13.6.3.1 The Consensus-Approximation Test

At least three (more or less) defensible strategies are at the disposal of the deploying entity in this situation of conflict. One option is the prioritization of the cultural sphere's limit that adheres closest to the generally established consensus of values as codified in instruments of international law. In the previous example, this would mean that in cases of conflict the technology deployer ought to act in accordance with the cultural sphere that most closely reflects what they consider to be the best approximation of global moral consensus. In cases in which the conflict is suspected to be due to systematic oppression e.g. of individuals of one gender or sex, reference to the generally upheld principles of gender equality would require the prioritization of the limits of the more protective cultural sphere; this even if this is in contrast to the technology user's explicit wishes. As a result, the technology user may (subjectively) perceive her autonomy to be constrained, while the deploying entity may see this (subjective)

[728] Note that there is a difference between a technology user and data subject voluntarily deferring their exercise of autonomy to someone else after having grasped the extent of such deference, as opposed to a systematic and perhaps less reflected deference on the basis of existing structures of power. To illustrate this, we may consider within the western cultural sphere there is little inherent problem in e.g. women being allowed to delegate administration and decision-making of certain matters of their life to their spouses (with the understanding that such delegation may be revoked), but there is indeed an inherent problem in these matters being assigned to the full discretion of the spouse by default (and without an understanding that it may be revoked).

constraint as a necessary consequence of counter-acting the harmful embedding of the individual.

13.6.3.2 The Observation-Based Corrective Test

A second option is the prioritization of the affected individual's perspective but to correct for potential mental constraints through some sort of corrective tests. Again, this may be useful in case the conflict is suspected to be rooted in some type of unacceptable phenomenon. Here, the deploying entity may attempt to approximate a hypothetical version of the individual in a state free of the its original state and compare it to the individual's preferences being transposed into the deploying entity's cultural sphere. For this to be the case, the informed consent-application, the technology deployer must suspect that the individual suffers from some sort of (culturally-derived) indoctrination, e.g. that the user and data subject feels an overwhelming obligation to exclude from her exercise of autonomy the ability to obtain and decide upon their own health information. The deployer must further have reason to believe, that this overwhelming urge must be imposed by some peculiarity of the cultural sphere the individual is embedded in, e.g. that they may believe to have no right and hence are obliged to defer their respective autonomy to their close relatives or spouses. In the case of the affirmative, the deploying entity must then apply some sort of corrective on top of the individuals stated (or implied) ethical preferences. For example, the deployer may conclude based on empirical data that a majority of certain data subjects within a cultural sphere that suffer from disenfranchisement and later leave that sphere's influence, tend to regret excluding sovereignty over their health information, while other distinguishable individuals do not. When considering the primacy of values to apply, the deployer can then apply a corrective assessment. The individuals most likely to regret conceiving their autonomy and its awarded protection in a more limited way ought to be empowered with sovereignty over their health information even if they (initially) consider this as a constraint to their autonomy. The individuals unlikely to regret their previous stance in cases where they leave the cultural sphere's influence may instead be subject to autonomy protection measures in correspondence with their cultural sphere's values.

13.6.3.3 The Constraint-Elimination Test

The third option again requires the deployer to prioritize the individual's perspective but screen for culturally derived constraints. Here the deployer must investigate if the individual would (or at least would be likely to) freely choose and affirm this set of beliefs, if the suspected indoctrinated element would be removed. By that, the deploying entity may consider which preferences about foregoing future autonomy are held by the individual only or mostly because of the "mental constraints" of concern and which would be present even when removing the potential constraints from the equation. Instead of broadly determining (perhaps on the basis of empirical data) the constraints certain groups of individuals suffer to their autonomy but are likely to regret, the deployer instead conducts its assessment on an individual level on a forward-looking basis. In doing so, one may attempt to model the hypothetical individual without the specific "mental constraints" in question and ask if the now "unconstrained" individual would consider them subjectively valid. As a result, the deploying entity may still choose to adhere to the individual's subjective scope and value of autonomy but with the exception of the mental constraints that the individual would consider unacceptable if they would be able to make that analysis freely. As a result, this type of analysis is both much more complex and much more granular, and if conducted successfully much more adherent to the individual's ethical preferences and self-conception of autonomy, both actual and potential.

13.6.3.4 Epistemic and Practical Challenges

As mentioned before, the mere task of defining autonomy is a thorny one. The same applies to determining individual's subjective view on autonomy at large (if such explicit view exist) and on the implications of those concepts for their interactions with technology. This poses challenges for all of the strategies outlined, but the opacity introduced by this issue affects them differently. Addressing the actual vulnerability-dilemma is somewhat straightforward. The deploying entity must ultimately "only" ascertain (or defer to legal, ethical or other normative frameworks for this purpose) what the actual outcomes are, that ought to be achieved. This outcome of this reverse-engineering of the ethical rules to be followed, i.e. the situations to be avoided (such as individuals being manipulated or coerced) must then be re-applied to the more vulnerable cultural sphere. To achieve this in detail is likely to be a complex undertaking nonetheless. Addressing the ethical foundation-dilemma gives a more differentiated field of challenges. The consensus-approximation test suffers only from the difficulty of determining whatever (global) moral consensus persists with respect to the autonomy risks their technology incurs. The observation-based corrective test instead is handicapped by the difficulties of obtaining empirical data and is at best reactive. It also risks painting with too wide a brush and assigning autonomy safeguards based on group characteristics. The constraint-elimination test seems most convincing in result, but most difficult and idealistic in its implementation. While there are certainly particularly egregious situations in which we may predict an individual's posture on the extent their autonomy should be constrained[729], this may not be possible to achieve with high confidence in many cases. The selection of risk mitigation strategies when deploying technology across cultural spheres to properly safeguard autonomy then requires an active choice and a meta-assessment of the viability of the strategies best suited for the technology deployed, the autonomy risks incurred and the cultural peculiarities the deployer and the affected individuals are embedded in. Ultimately, a technology deploying entity will not be able to avoid to take an active and non-neutral posture to balance these practical, epistemic challenges to ensure that the autonomy of individuals interacting with or affected by their technology is duly considered and protected.

13.7 Conclusion

This section has outlined how intercultural differences in how autonomy is conceptualized and subjectively perceived affects the ethics of deployment of algorithmic agents and thus answered research sub-question 3.3. This text has elaborated on why intercultural differences are of importance when considering the impact of technology on autonomy. To this end, this text has suggested to adopt (mostly) a self-referential view when considering ethical issues arising in this domain. It has also identified two main factors with respect to how these differences may manifest, and how this changes their respective persuasion envelopes. Both intrinsic cultural factors as well as temporal and gradual factors of technology introduction have been found to be relevant in this context. Lastly, two dilemmas that stem from potential cultural differences within cultural spheres and potential strategies to address them have been introduced and possible addressal strategies identified.

Considerations of matters of interculturality, specifically with respect to autonomy implications of technology, are sparse at the time of writing. Future research both conceptually into the interplay between competing autonomy theories and empirically into the differences in implication of autonomy-undermining technology deployed in different cultural spheres and

[729] For example, we may safely assume that an individual that has been indoctrinated to the result that they prefer a life of incarceration will choose not to be incarcerated if this indoctrination and its effects would be removed.

their defusal strategies may further illustrate the framework that has been proposed here. However, I predict that many intercultural differences identified in this section will generally erode with time due to the following reasons: First, some differences in autonomy conceptions are based on structural discrimination of certain subgroups of individuals within a cultural sphere. Autonomy conceptions that use a more egalitarian approach are likely to be more persuasive and be adopted in the long term. Second, in an interconnected world, some sort of cultural convergence is bound to happen; part of this convergence of ideas and preferences will invariably be individuals' subjective perception of autonomy. Third, technology and in particular persuasive algorithmic agents exert force on the individual interacting with them. As algorithmic agents interact with more and more individuals, more and more individuals are pressed into the technology's interaction envelope. This standardization of behaviour may radiate outwards into the preferences, concerns and conceptions these individuals and their respective cultural sphere have about autonomy and its interplay with technology.

With this section, the thesis has addressed the final remaining research sub-questions and has concluded in its holistic and interdisciplinary evaluation of the development and exercise of autonomy and its constraints.

14 Conclusion

This thesis has investigated the exercise and constraint of autonomy that individuals experience in connection the domain of information technology. In line with parlance adopted by the research group the author was a part of, this domain was denoted as the Internet of Everything (or IoE in short). The primary motivation for this research was the steadily intensifying concerns voiced by political stakeholders, academics, civil society organisations and watchdogs of varying standing. There seems to be a consensus that information technology, as it becomes more advanced and prevalent, poses an increasing risk to the individual's capacity and exercise of self-government. A rigorous analysis into some aspects of these matters was therefore appropriate.

14.1 Research Questions Revisited

The overarching research objective was to contribute to answering the question of what freedom and autonomy an individual can develop and maintain in presence of systems of the IoE. More specifically, this thesis has dealt with an inversion of this question and has focused on identifying and classifying obstacles to the development and maintaining of autonomy in presence of systems of the IoE. Owed to the broad research mandate this objective provides, and due to the fact, the investigation, if it ought to be useful as well as novel, benefits from an interdisciplinary approach consolidating and connecting different fields of inquiry, the thesis touched upon many aspects of the interplay between individual autonomy and technology.

Very broadly, the thesis was divided into three parts. The first part identified a problem with current autonomy discourse: There is no agreed upon object of reference when bemoaning loss of or risk to an individual's autonomy. Believing this to be a detriment both to engage in useful discourse and to develop appropriate countermeasures against autonomy constraints that may be deemed inappropriate or unacceptable, the first part of this thesis has introduced a pragmatic conceptual framework to classify autonomy constraints.

The first part of this thesis was guided by four research sub-questions:

- Question 1.1 asked if there is a current consensus on how to conceptualize autonomy in general and in the domain of technology specifically. In pursuit of this question this thesis, drawing from the field of philosophy, has outlined existing hierarchical (or mesh-) theories of autonomy in Section 2, which conceive autonomy as a congruence of mental states, and contraposed them with so-called relational or feminist critiques. It has found no truly established consensus on how to conceptualize autonomy either generally or with respect to technology.
- Question 1.2 considered how the concepts of privacy and autonomy are connected. Section 3 has highlighted the highly interdependent nature of both concepts, noting the prerequisite nature of privacy to many practical autonomy exercise situations, and has argued that this relationship is becoming more pronounced as technology becomes more advanced and prevalent.
- In a similar vein, Question 1.3 asked about the representation of individual autonomy in the legal domain. Section 4 has shown that autonomy permeates the legal domain on all levels of application; serving as a prerequisite fundamental assumption but also reflects on many subfields and regulatory regimes such as concepts of responsibility and liability in criminal and civil law or contractual autonomy. Particular focus was given to the mirroring of many of the concepts outlined in Section 2 in the fundamental

protection awarded by human rights regimes such as the European Convention of Human Rights.

- Finally, Question 1.4 was concerned on how to classify obstacles to the development and maintaining of autonomy. In answer to this and to the overarching Question 1, Section 5 has outlined a pragmatic account of autonomy in which autonomy can be constrained on three different dimensions: intrinsic, relational and informational. The utility and robustness of the pragmatic account was made plausible by reference to the previously outlined philosophical-theoretical frameworks, the autonomy-protection awarded by human rights frameworks and its conformity with the intuitive concerns about autonomy-constraining technology.

The second part of the thesis followed Research Question 2, namely what role information technology has in enabling and facilitating autonomy constraints. For this purpose, the thesis introduced to theoretical concepts that have structured the rest of the inquiry. First, it has grouped all processes from information origin to information recipient into a compound denoted as the informational pipeline. Second, it has introduced the term vector of influence to denote the compound phenomena that characterize a given interaction between an information agent and another entity that is capable of imparting change onto an individual's autonomy. These two concepts in tandem structure the factors that enable and facilitate autonomy constraint into three parts: domain-specific vectors, agent-specific vectors, and information recipient-specific vectors. The rest of Part II followed this structure in answering four research sub-questions:

- Question 2.1 asked about the relevance of existing research on persuasive technology to understanding autonomy constraints. Consequently, Section 6 has given a brief overview over the field of persuasive technology, its qualifications, ethical implications and matters of disambiguation, and has shown it to be, in many ways, valuable precursory research to this thesis. By outlining this field of research, the thesis has provided important context to the inquiry at hand.
- Question 2.2 considers the first of three types of vectors of influence and asked how the IoE-domain generally and structurally affects the imposing of autonomy constraints. To this end, Section 8 has found there to be two main clusters of domain characteristics, physical and meta characteristics, of which both have the potential to facilitate autonomy constraints. The physical domain characteristics describe the general development of technology: devices and infrastructure become more powerful, capable and prevalent. In addition, the thesis has outlined eight meta characteristics, which are the intermediation of information, the immediacy of information transfer, low barriers to impart information, plurality of information recipients, anonymity or pseudonymity of information transfer participants, personalization, fragmentation and erosion of privacy, and easy implementation of scalable, non-human actors.
- Moving to the second of three types of vectors of influence, Question 2.3 asked, what the characteristics of information transfer processes (i.e. informational agents) are, which would be potentially relevant to the thesis' inquiry. To answer this question, Section 9 has first provided a definition for the concept of an informational agent based on the inquiry at hand and existing research on agents as carriers of agency. Particularly, this thesis has proposed defining agents as informational pipeline-bound non-human processes, that display perceived compactness, ascribable agency, sufficient complexity and impart persuasion. Further analysis then found such agents to display at least seven relevant characteristics which are derived both from the underlying domain and its characteristics, and the context, use and function of the agent itself. The

agent characteristics identified are the (relative) immediacy of information processing, selective superior reasoning capabilities, opacity of algorithmic decision-making, data reliance and susceptibility to bias, their complex status between mistrust and deference, their potential to facilitate and amplify information transfer and their placement, function and effectuation obscurity.

- Finally, Question 2.4 asked what the typical patterns and phenomena that characterize risk to an individual's autonomy in the context of information receival are, thereby concerning itself with the third of three types of vectors of influence. Section 10 identified four additional factors to consider: the medium of information transfer, the (perceived) veracity, accuracy, trustworthiness and assertiveness of the information process interacted with, transparency and completeness of information received and, as a special case, the assignation of actionable data.

The third part of this thesis aimed to answer research question 3 which asked which ethical and legal boundaries pertaining to autonomy constraints imposed by technology (and in particular informational agents) exist? While the main part of this thesis was not of ethical focus, the motivation for this research was in part fuelled by intuitive concerns and the proposed utility of the frameworks presented herein lie in their facilitation of operationalizing ethical viewpoints concerning autonomy constraints. As a result, this third part rounds out the existing, stipulative research with considerations from the legal and ethical domain, guided by the three following research sub-questions.

- Question 3.1 asked how and under consideration of which factors an ethical analysis of technology-imposed autonomy constraints should be conducted and to what extent the pragmatic account of autonomy and the concept of vectors of influence would be useful in this context. Section 11 outlined a rough ethical gradient and highlighted potential junctures in the analysis of autonomy constraints, depending on the ethical viewpoint taken. Connecting both the concept of the pragmatic account of autonomy and the concept of vectors of influence, Section 11 also highlighted how these frameworks are useful to highlight main areas of concern and identify important fulcrums of justifications and attribution.
- Question 3.2 called for an investigation in how the concept of autonomy as exercised and constraint in presence of technology is reflected in existing and upcoming European legislative instruments. On the basis of careful analysis of European legal sources, Section 12 has traced the concept of autonomy and its protection from its implicit, privacy-focused addressal to the increasingly aggressive posture taken by the European legislator.
- Finally, and in recognition of the euro-centric approach of this thesis, question 3.3 asked, what the impact of cultural embeddings of individuals on efforts to safeguard autonomy and how does this affect intercultural deployment of technology is? Expanding on the existing conceptualization, Section 13 showed that autonomy and harm to it is dependent on its cultural embedding and ought to be seen as subjectively determined and highlighted intercultural flashpoints of such autonomy differences in the context of technology development. Further analysis suggested that the persuasion envelope accessible to technology is particularly dependent on digital literacy, the differences in technology introduction from a temporal standpoint and the vulnerability and resilience awarded by the cultural embedding. Finally, Section 13 has highlighted ways to diffuse concerns about exploitation of vulnerability differences and has suggested three ethical tests to determine the ethicality of intercultural technology deployment.

With the addressal of these three main research questions and their respective eleven research-sub question, this thesis has aimed to address the question of how individuals can develop and maintain autonomy (or conversely how they can be constrained in such efforts) in presence of systems of the IoE from multiple directions. It has provided a conceptual superstructure both to understand autonomy and autonomy constraints and to classify and compare factors which facilitate autonomy constraints. It has also shown how this conceptual superstructure is useful in outlining important factors to consider when attempting ethical analysis and traced the evermore important concept of autonomy in the legal domain. As a result, the nature of autonomy exercised and constrained in the digital domain has hopefully become clearer, and progress has thus been made for both of the two goals of autonomy research outlined in the beginning of this thesis.

14.2 Outlook

Individual autonomy will remain a concern in the future. As technology continues to become inextricably linked with modern life, and as many individual interactions become adjacent to, dependent on or intermediated by the digital transfer of information, the risk of autonomy constraints imposed by technology only rises. At the same time, the context in which such constraints are imposed is constantly subject to change: As technology adoption and digital literacy moves with every "big next thing" announced and put in practice, and every generation experiences a different embedding of technology within their lives, the impact remains difficult to scope. Ongoing quantitative and qualitative research will be needed to paint an accurate picture of how persuasion, manipulation or coercion is happening as these circumstances are evolving, and strong conceptual frameworks will be needed to identify, classify, compare and evaluate such instances. To this, the present thesis has aimed to contribute, but there seems pressing need for more research.

In addition, due to the ethical implications of autonomy constraints, it also follows also that the regulation of such technology will continue to be worth watching. The increasingly aggressive posture taken by the European legislature suggests that regulators want to take a proactive role in preventing technology from unduly influencing human behaviour. As the financial rewards for predicting and influencing behaviour remain attractive, the economic incentives point towards stakeholders attempting to skirt regulation as long as possible. In this, they are aided by the considerable opacity and blurredness of the intuitions behind many of the regulation efforts; the very same which this thesis has tried to crystallize out. Time will tell, to what extent these efforts will be successful.

Bibliography

Abdel Monem, Hisham, 'The Effectiveness of Advertising Personalization', *Journal of Design Sciences and Applied Arts*, 2.1 (2021), 114–21

Abo-Tabik, Maryam, Yael Benn, and Nicholas Costen, 'Are Machine Learning Methods the Future for Smoking Cessation Apps?', *Sensors*, 21.13 (2021), 4254 <https://doi.org/10.3390/s21134254>

Acquisti, Alessandro, 'Privacy in Electronic Commerce and the Economics of Immediate Gratification', in *Proceedings of the 5th ACM Conference on Electronic Commerce - EC '04* (New York, New York, USA: ACM Press, 2004), p. 21 <https://doi.org/10.1145/988772.988777>

Adebiyi, Marion O, Oluwasemilore O Adeoye, ROSELINE O OGUNDOKUN, Julius O Okesola, and Ayodele A Adebiyi, 'Secured Loan Prediction Systens Using Artificial Neural Network', *Journal of Engineering Science and Technology*, 17.2 (2022), 854–73

Adedinsewo, Demilade, Rickey E. Carter, Zachi Attia, Patrick Johnson, Anthony H. Kashou, Jennifer L. Dugan, and others, 'Artificial Intelligence-Enabled ECG Algorithm to Identify Patients With Left Ventricular Systolic Dysfunction Presenting to the Emergency Department With Dyspnea', *Circulation: Arrhythmia and Electrophysiology*, 13.8 (2020) <https://doi.org/10.1161/CIRCEP.120.008437>

Adler, Julian, Sarah Picard, and Caitlin Flood, 'Arguing the Algorithm: Pretrial Risk Assessment and the Zealous Defender Jed D. Melnick Symposium: Innovations in Justice: Experiments in Restorative Justice', *Cardozo Journal of Conflict Resolution*, 21.3, 581–96 <https://heinonline.org/HOL/P?h=hein.journals/cardcore21&i=601>

Aguirre, Elizabeth, Dominik Mahr, Dhruv Grewal, Ko de Ruyter, and Martin Wetzels, 'Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness', *Journal of Retailing*, 91.1 (2015), 34–49 <https://doi.org/10.1016/j.jretai.2014.09.005>

Akour, Mohammed, and Bouchaib Falah, 'Application Domain and Programming Language Readability Yardsticks', in *2016 7th International Conference on Computer Science and Information Technology (CSIT)* (IEEE, 2016), pp. 1–6 <https://doi.org/10.1109/CSIT.2016.7549476>

'Algorithm', *Merriam-Webster Dictionary*, 2022

Allen, Anita, 'Coercing Privacy', *William & Mary Law Review*, 40.723–757 (1999)

Allred, Sarah R., L. Elizabeth Crawford, Sean Duffy, and John Smith, 'Working Memory and Spatial Judgments: Cognitive Load Increases the Central Tendency Bias', *Psychonomic Bulletin & Review*, 23.6 (2016), 1825–31 <https://doi.org/10.3758/s13423-016-1039-0>

Amazeen, Michelle A., and Erik P. Bucy, 'Conferring Resistance to Digital Disinformation: The Inoculating Influence of Procedural News Knowledge', *Journal of Broadcasting & Electronic Media*, 63.3 (2019), 415–32 <https://doi.org/10.1080/08838151.2019.1653101>

Amin, Rafik Khairul, Yuliant Sibaroni, and others, 'Implementation of Decision Tree Using C4. 5 Algorithm in Decision Making of Loan Application by Debtor (Case Study: Bank Pasar of Yogyakarta Special Region)', in *2015 3rd International Conference on Information and Communication Technology (ICoICT)*, 2015, pp. 75–80

André, Quentin, Ziv Carmon, Klaus Wertenbroch, Alia Crum, Douglas Frank, William Goldstein, and others, 'Consumer Choice and Autonomy in the Age of Artificial Intelligence and Big Data', *Customer Needs and Solutions*, 5.1–2 (2018), 28–37 <https://doi.org/10.1007/s40547-017-0085-8>

Andrew, Adrienne, Gaetano Borriello, and James Fogarty, 'Toward a Systematic Understanding of Suggestion Tactics in Persuasive Technologies', 2007, pp. 259–70

<https://doi.org/10.1007/978-3-540-77006-0_32>
Ang, Nobel, 'Positive Freedom as Exercise of Rational Ability: A Kantian Defense of Positive Liberty', *The Journal of Value Inquiry*, 48.1 (2014), 1–16 <https://doi.org/10.1007/s10790-013-9399-4>
Archibugi, Daniele, and Simona Iammarino, 'The Globalization of Technological Innovation: Definition and Evidence', *Review of International Political Economy*, 9.1 (2002), 98–122
Arcimowicz, Beata, Katarzyna Cantarero, and Emilia Soroko, 'Motivation and Consequences of Lying. A Qualitative Analysis of Everyday Lying', *Forum: Qualitative Social Research*, 16.3 (2015) <https://doi.org/https://doi.org/10.17169/fqs-16.3.2311>
Armstrong, David, 'Truth', in *Belief, Truth and Knowledge* (Cambridge University Press, 1973), pp. 113–34 <https://doi.org/10.1017/CBO9780511570827.009>
Arpaly, Nomy, 'Autonomy by Andrew Sneddon', *Notre Dame Philosophical Reviews*, 2014
Atkinson, Bernardine M. C., 'Captology: A Critical Review', 2006, pp. 171–82 <https://doi.org/10.1007/11755494_25>
Attia, Zachi I, Peter A Noseworthy, Francisco Lopez-Jimenez, Samuel J Asirvatham, Abhishek J Deshmukh, Bernard J Gersh, and others, 'An Artificial Intelligence-Enabled ECG Algorithm for the Identification of Patients with Atrial Fibrillation during Sinus Rhythm: A Retrospective Analysis of Outcome Prediction', *The Lancet*, 394.10201 (2019), 861–67 <https://doi.org/10.1016/S0140-6736(19)31721-0>
Aylsworth, Timothy, 'Autonomy and Manipulation: Refining the Argument Against Persuasive Advertising', *Journal of Business Ethics*, 175.4 (2022), 689–99 <https://doi.org/10.1007/s10551-020-04590-6>
Bacchelli, Alberto, and Christian Bird, 'Expectations, Outcomes, and Challenges of Modern Code Review', in *2013 35th International Conference on Software Engineering (ICSE)* (IEEE, 2013), pp. 712–21 <https://doi.org/10.1109/ICSE.2013.6606617>
Bachrach, Peter, 'The Supreme Court, Civil Liberties, and the Balance of Interest Doctrine', *The Western Political Quarterly*, 14.2 (1961), 391 <https://doi.org/10.2307/443595>
Baier, Annette, *Postures of the Mind: Essays on Mind and Morals* (Minneapolis: University of Minnesota Press, 1985)
Barandiaran, Xabier E., Ezequiel Di Paolo, and Marieke Rohde, 'Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action', *Adaptive Behavior*, 17.5 (2009), 367–86 <https://doi.org/10.1177/1059712309343819>
Baron, Marcia, 'Is Justification (Somehow) Prior to Excuse? A Reply to Douglas Husak', *Law and Philosophy*, 24.6 (2005), 595–609 <https://doi.org/10.1007/s10982-005-0842-y>
Barragan-Jason, G., M. Cauchoix, and E.J. Barbeau, 'The Neural Speed of Familiar Face Recognition', *Neuropsychologia*, 75 (2015), 390–401 <https://doi.org/10.1016/j.neuropsychologia.2015.06.017>
Barth, Susanne, and Menno D.T. de Jong, 'The Privacy Paradox – Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior – A Systematic Literature Review', *Telematics and Informatics*, 34.7 (2017), 1038–58 <https://doi.org/10.1016/j.tele.2017.04.013>
Barua, Zapan, Sajib Barua, Salma Aktar, Najma Kabir, and Mingze Li, 'Effects of Misinformation on COVID-19 Individual Responses and Recommendations for Resilience of Disastrous Consequences of Misinformation', *Progress in Disaster Science*, 8 (2020), 100119 <https://doi.org/10.1016/j.pdisas.2020.100119>
Bateson, Regina, 'The Politics of Vigilantism', *Comparative Political Studies*, 54.6 (2020), 923–55 <https://doi.org/10.1177/0010414020957692>
Baum, Manfred, 'Positive Und Negative Freiheit Bei Kant', *Jahrbuch Für Recht Und Ethik / Annual Review of Law and Ethics*, 16 (2008), 43–56 <http://www.jstor.org/stable/43579350>

Bawden, David, and Lyn Robinson, 'Information Overload: An Introduction', in *Oxford Research Encyclopedia of Politics* (Oxford University Press, 2020) <https://doi.org/10.1093/acrefore/9780190228637.013.1360>
Baynes, Kenneth, 'Freedom as Autonomy', in *The Oxford Handbook of Continental Philosophy* (Oxford University Press, 2007), pp. 551–87 <https://doi.org/10.1093/oxfordhb/9780199234097.003.0017>
Beauchamp, Tom, and James Childress, *Principles of Biomedical Ethics*, 5th edn (New York: Oxford University Press, 2001)
Ben-Shahar, Omri, *More Than You Wanted To Know: The Failure of Mandated Disclosure* (Princeton University Press, 2014)
Ben-Shahar, Omri, and Ariel Porat, 'Personalizing Negligence Law ', *New York University Law Review*, 91.3 (2016), [i]-688 OP-688 <https://heinonline.org/HOL/P?h=hein.journals/nylr91&i=650>
Benn, Stanley, 'Privacy, Freedom, and Respect for Persons', in *Philosophical Dimensions of Privacy: An Anthology*, ed. by Ferdinand Shoeman (Cambridge: Cambridge University Press, 1984)
Benson, Paul, 'Feminist Intuitions and the Normative Substance of Autonomy', in *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, ed. by James Taylor (Cambridge: Cambridge University Press, 2005), pp. 124–42
Berdichevsky, Daniel, and Erik Neuenschwander, 'Toward an Ethics of Persuasive Technology', *Communications of the ACM*, 42.5 (1999), 51–58 <https://doi.org/10.1145/301353.301410>
Bergstein, Brian, 'What AI Still Can't Do', *MIT Technology Review*, 2020 <https://www.technologyreview.com/2020/02/19/868178/what-ai-still-cant-do/>
Berka, Walter, 'Aktuelle Bedrohungen Des Grundrechts Auf Privatsphäre', *Österreichische Juristenzeitung*, 17 (2018), 755
Berlin, Isaiah, 'Two Concepts of Liberty', in *The Liberty Reader* (Routledge, 2017), pp. 33–57
Bilal, Muhammad, Abdullah Gani, Muhammad Ikram Ullah Lali, Mohsen Marjani, and Nadia Malik, 'Social Profiling: A Review, Taxonomy, and Challenges', *Cyberpsychology, Behavior, and Social Networking*, 22.7 (2019), 433–50 <https://doi.org/10.1089/cyber.2018.0670>
Birukou, Aliaksandr, Enrico Blanzieri, Paolo Giorgini, and Fausto Giunchiglia, 'A Formal Definition of Culture', 2013, pp. 1–26 <https://doi.org/10.1007/978-94-007-5574-1_1>
Bjorge, Eirik, *The Evolutionary Interpretation of Treaties*, 2014 <https://doi.org/10.1093/acprof:oso/9780198716143.001.0001>
Blair, J. Anthony, 'Argumentation as Rational Persuasion', *Argumentation*, 26.1 (2012), 71–81 <https://doi.org/10.1007/s10503-011-9235-6>
Bleier, Alexander, and Maik Eisenbeiss, 'Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where', *Marketing Science*, 34.5 (2015), 669–88 <https://doi.org/10.1287/mksc.2015.0930>
———, 'The Importance of Trust for Personalized Online Advertising', *Journal of Retailing*, 91.3 (2015), 390–409 <https://doi.org/10.1016/j.jretai.2015.04.001>
Bodle, Robert, 'Predictive Algorithms and Personalization Services on Social Network Sites: Implications for Users and Society', in *The Ubiquitous Internet* (London/New York: Routledge, 2014)
Bogen, Miranda, and Aaron Rieke, 'Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias', 2018
Bogert, Eric, Aaron Schecter, and Richard T. Watson, 'Humans Rely More on Algorithms than

Social Influence as a Task Becomes More Difficult', *Scientific Reports*, 11.1 (2021), 8028 <https://doi.org/10.1038/s41598-021-87480-9>
Bösch, Christoph, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher, 'Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns', *Proceedings on Privacy Enhancing Technologies*, 2016.4 (2016), 237–54
Botterell, Andrew, 'A Primer on the Distinction between Justification and Excuse', *Philosophy Compass*, 4.1 (2009), 172–96
Boyd, R, and P J Richerson, 'Culture and the Evolutionary Process. Chicago (The University of Chicago Press) 1985.', 1985
Bradshaw, Amy C., 'Internet Users Worldwide', *Educational Technology Research and Development*, 49.4 (2001), 112–17 <https://doi.org/10.1007/BF02504952>
Brashier, Nadia M., and Daniel L. Schacter, 'Aging in an Era of Fake News', *Current Directions in Psychological Science*, 29.3 (2020), 316–23 <https://doi.org/10.1177/0963721420915872>
Breen, Stephen, Karim Ouazzane, and Preeti Patel, 'GDPR: Is Your Consent Valid?', *Business Information Review*, 37.1 (2020), 19–24 <https://doi.org/10.1177/0266382120903254>
Brennan, Carol, and Martin Coppack, 'Consumer Empowerment: Global Context, UK Strategies and Vulnerable Consumers', *International Journal of Consumer Studies*, 32.4 (2008), 306–13 <https://doi.org/10.1111/j.1470-6431.2007.00640.x>
Briscoe, Neil, 'Understanding the OSI 7-Layer Model', *PC Network Advisor*, 120.2 (2000), 13–15
Brison, Susan, 'Relational Autonomy and Freedom of Expression', in *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*, ed. by Catriona Mackenzie and Natalie Stoljar (Oxford University Press, 2000), pp. 280–300
Brustoloni, Jose C, *Autonomous Agents: Characterization and Requirements* (Citeseer, 1991)
Buckingham, David, 'Defining Digital Literacy', in *Medienbildung in Neuen Kulturräumen* (Wiesbaden: VS Verlag für Sozialwissenschaften, 2010), pp. 59–71 <https://doi.org/10.1007/978-3-531-92133-4_4>
Bucy, Eric, Walter Gantz, and Zheng Wang, 'Media Technology and the 24-Hour News Cycle', in *Communication Technology and Social Change* (London/New York: Routledge, 2007)
Bumke, Christian, and Anne Röthel, 'Das Paradox Der Autonomy Und Seine Entfaltungen', in *Karsten Fischer*, ed. by Christian Bumke and Anne Röthel (Tübingen: Mohr Siebeck, 2017), pp. 411–35
Burns, Kelly, and Antoine Bechara, 'Decision Making and Free Will: A Neuroscience Perspective', *Behavioral Sciences \& the Law*, 25.2 (2007), 263–80
Burr, Christopher, Nello Cristianini, and James Ladyman, 'An Analysis of the Interaction Between Intelligent Software Agents and Human Users', *Minds and Machines*, 28.4 (2018), 735–74 <https://doi.org/10.1007/s11023-018-9479-0>
Burrell, Jenna, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms', *Big Data & Society*, 3.1 (2016), 205395171562251 <https://doi.org/10.1177/2053951715622512>
Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova, *Social Media and Xenophobia: Evidence from Russia* (Cambridge, MA, December 2019) <https://doi.org/10.3386/w26567>
Buse, Raymond P.L., and Westley R. Weimer, 'A Metric for Software Readability', in *Proceedings of the 2008 International Symposium on Software Testing and Analysis - ISSTA '08* (New York, New York, USA: ACM Press, 2008), p. 121 <https://doi.org/10.1145/1390630.1390647>
Buss, Sarah, 'Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints', *Ethics*, 115.2 (2005), 195–235

<https://doi.org/10.1086/426304>
BVerfG, 1 BvR 209/83 (Volkszählung), 1983
Cahn, Aaron, Scott Alfeld, Paul Barford, and Shanmugavelayutham Muthukrishnan, 'An Empirical Study of Web Cookies', in *Proceedings of the 25th International Conference on World Wide Web*, 2016, pp. 891–901
Calero Valdez, André, 'Human and Algorithmic Contributions to Misinformation Online - Identifying the Culprit', 2020, pp. 3–15 <https://doi.org/10.1007/978-3-030-39627-5_1>
———, 'Human and Algorithmic Contributions to Misinformation Online - Identifying the Culprit', in *Disinformation in Open Online Media*, ed. by Christian Grimme, Mike Preuss, Frank W Takes, and Annie Waldherr (Cham: Springer International Publishing, 2020), pp. 3–15
Cambridge Dictionary, 'Disinformation', 2021
Candogan, Ozan, and Kimon Drakopoulos, 'Optimal Signaling of Content Accuracy: Engagement vs. Misinformation', *Operations Research*, 2020, opre.2019.1897 <https://doi.org/10.1287/opre.2019.1897>
Capurro, Rafael, 'Privacy. An Intercultural Perspective', *Ethics and Information Technology*, 7.1 (2005), 37–47
Carmon, Ziv, Klaus Wertenbroch, and Marcel Zeelenberg, 'Option Attachment: When Deliberating Makes Choosing Feel like Losing', *Journal of Consumer Research*, 30.1 (2003), 15–29 <https://doi.org/10.1086/374701>
Carter, Ian, 'Positive and Negative Liberty', 2003
Cetnarowicz, Krzysztof, 'From Algorithm to Agent', 2009, pp. 825–34 <https://doi.org/10.1007/978-3-642-01973-9_92>
Cevolini, Alberto, and Elena Esposito, 'From Pool to Profile: Social Consequences of Algorithmic Prediction in Insurance', *Big Data & Society*, 7.2 (2020), 205395172093922 <https://doi.org/10.1177/2053951720939228>
Chandler, Daniel, and Rod Munday, 'Disinformation', in *A Dictionary of Media and Communication* (Oxford: Oxford University Press, 2011)
Chelidonis, Apostolos, 'Hypothetische Und Überholende Kausalität', *JURA - Juristische Ausbildung*, 43.3 (2021), 227–37 <https://doi.org/10.1515/jura-2020-2649>
Chen, Beiwen, Maarten Vansteenkiste, Wim Beyers, Bart Soenens, and Stijn Van Petegem, 'Autonomy in Family Decision Making for Chinese Adolescents', *Journal of Cross-Cultural Psychology*, 44.7 (2013), 1184–1209 <https://doi.org/10.1177/0022022113480038>
Chen, Hong, and Todd Jackson, 'Are Cognitive Biases Associated with Body Image Concerns Similar between Cultures?', *Body Image*, 2.2 (2005), 177–86 <https://doi.org/10.1016/j.bodyim.2005.03.005>
Chen, Qi, Yuqiang Feng, Luning Liu, and Xianyun Tian, 'Understanding Consumers' Reactance of Online Personalized Advertising: A New Scheme of Rational Choice from a Perspective of Negative Effects', *International Journal of Information Management*, 44 (2019), 53–64 <https://doi.org/10.1016/j.ijinfomgt.2018.09.001>
Chernev, Alexander, Ulf Böckenholt, and Joseph Goodman, 'Choice Overload: A Conceptual Review and Meta-Analysis', *Journal of Consumer Psychology*, 25.2 (2015), 333–58 <https://doi.org/10.1016/j.jcps.2014.08.002>
Cherry, S., 'Edholm's Law of Bandwidth', *IEEE Spectrum*, 41.7 (2004), 58–60 <https://doi.org/10.1109/MSPEC.2004.1309810>
Cheung, Anne S.Y., 'Location Privacy: The Challenges of Mobile Service Devices', *Computer Law & Security Review*, 30.1 (2014), 41–54 <https://doi.org/10.1016/j.clsr.2013.11.005>
Chinn, M. D., and R. W. Fairlie, 'The Determinants of the Global Digital Divide: A Cross-Country Analysis of Computer and Internet Penetration', *Oxford Economic Papers*, 59.1

(2006), 16–44 <https://doi.org/10.1093/oep/gpl024>
Christen, Markus, 'Autonomie – Eine Aufgabe Für Die Philosophie', *Schweizerische Zeitschrift Für Philosophie*, 66 (2007) <https://doi.org/10.24894/StPh-de.2007.66012>
Christman, John, 'Autonomy and Personal History', *Canadian Journal of Philosophy*, 21.1 (1991), 1–24
———, 'Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves', *Philosophical Studies*, 117.1/2 (2004), 142–64
Christman, John, and Joel Anderson, 'Introduction', in *Autonomy and the Challenges to Liberalism*, ed. by John Christman and Joel Anderson (Cambridge University Press, 2005) <https://doi.org/10.1017/CBO9780511610325>
Christopherson, Kimberly M., 'The Positive and Negative Implications of Anonymity in Internet Social Interactions: "On the Internet, Nobody Knows You're a Dog"', *Computers in Human Behavior*, 23.6 (2007), 3038–56 <https://doi.org/10.1016/j.chb.2006.09.001>
Chua, Sook Ning, Nathan Wong, and Richard Koestner, 'Autonomy and Controlling Support Are Two Sides of the Same Coin', *Personality and Individual Differences*, 68 (2014), 48–52 <https://doi.org/10.1016/j.paid.2014.04.008>
Church, Karen, and Rodrigo de Oliveira, 'What's up with Whatsapp?', in *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '13* (New York, New York, USA: ACM Press, 2013), p. 352 <https://doi.org/10.1145/2493190.2493225>
Cialdini, R., *Influence, Science and Practice* (Boston: Allyn and Bacon, 2001)
CISCO, *The Internet of Everything - Global Public Sector Economic Analysis*, 2013
Clavero, Miguel, '"'Awkward Wording. Rephrase'": Linguistic Injustice in Ecological Journals', 2010
Cohen, Andrew Jason, 'Autonomy, Written by Andrew Sneddon', *Journal of Moral Philosophy*, 13.6 (2016), 764–67 <https://doi.org/10.1163/17455243-01306007>
Colomina, Carme, Héctor SÁNCHEZ Margalef, and Richard Young, *The Impact of Disinformation on Democratic Processes and Human Rights in the World (Study Requested by the EU-DROI Subcommittee)*, 2021
Commission, European, 'COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS A European Strategy for Data COM (2020) 66 Final', 2020 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066>
Cook, David A., So-Young Oh, and Martin V. Pusic, 'Accuracy of Physicians' Electrocardiogram Interpretations', *JAMA Internal Medicine*, 180.11 (2020), 1461 <https://doi.org/10.1001/jamainternmed.2020.3989>
Cooper, John, 'Stoic Autonomy', in *Autonomy*, ed. by Ellen Frankel Paul, Fred Miller, and Jeffrey Paul (Cambridge: Cambridge University Press, 2003), pp. 1–29
Coppolino Perfumi, Serena, Franco Bagnoli, Corrado Caudek, and Andrea Guazzini, 'Deindividuation Effects on Normative and Informational Social Influence within Computer-Mediated-Communication', *Computers in Human Behavior*, 92 (2019), 230–37 <https://doi.org/10.1016/j.chb.2018.11.017>
Costa, M Victoria, 'Neo-Republicanism, Freedom as Non-Domination, and Citizen Virtue', *Politics, Philosophy \& Economics*, 8.4 (2009), 401–19
Council of Europe, *Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes*, 2019
Crisp, Roger, 'Persuasive Advertising, Autonomy, and the Creation of Desire', *Journal of Business Ethics*, 6.5 (1987), 413–18 <https://doi.org/10.1007/BF00382898>
Cvrcek, Dan, Marek Kumpost, Vashek Matyas, and George Danezis, 'A Study on the Value of

Location Privacy', in *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society - WPES '06* (New York, New York, USA: ACM Press, 2006), p. 109 <https://doi.org/10.1145/1179601.1179621>
Cyr, Dianne, Carole Bonanni, John Bowes, and Joe Ilsever, 'Beyond Trust: Web Site Design Preferences across Cultures', *Journal of Global Information Management (JGIM)*, 13.4 (2005), 25–54
Danaher, John, and Henrik Skaug Sætra, 'Technology and Moral Change: The Transformation of Truth and Trust', *Ethics and Information Technology*, 24.3 (2022), 35 <https://doi.org/10.1007/s10676-022-09661-y>
Darwall, Stephen, 'The Value of Autonomy and Autonomy of the Will', *Ethics*, 116.2 (2006), 263–84
Davis, Fred D, 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology', *MIS Quarterly*, 1989, 319–40
Davis, Janet, 'Design Methods for Ethical Persuasive Computing', in *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09* (New York, New York, USA: ACM Press, 2009), p. 1 <https://doi.org/10.1145/1541948.1541957>
Day, Gregory, and Abbey Stemler, 'Are Dark Patterns Anticompetitive? ', *Alabama Law Review*, 72.1, 1–46 <https://heinonline.org/HOL/P?h=hein.journals/bamalr72&i=11>
Deigh, John, 'Ethics', *The Cambridge Dictionary of Philosophy* (Cambridge University Press, 2015)
DEMARCO, JOSEPH, and PAUL FORD, 'Balancing in Ethical Deliberation: Superior to Specification and Casuistry', *The Journal of Medicine and Philosophy*, 31.5 (2006), 483–97 <https://doi.org/10.1080/03605310600912675>
Dennett, Danniel, *The Intentional Stance, Cambridge* (Cambridge: MIT Press, 1987)
Deutsch, M, and H Gerard, 'A Study of Normative and Informational Social Influences upon Individual Judgment', *J Abnorm Soc Psychol*, 51 (1955), 629–636
Dhar, Ravi, and Klaus Wertenbroch, 'Self-Signaling and the Costs and Benefits of Temptation in Consumer Choice', *Journal of Marketing Research*, 49.1 (2012), 15–25 <https://doi.org/10.1509/jmr.10.0490>
Diana Meyers, *Self, Society and Personal Choice* (New York: Columbia University Press)
Diggelmann, Oliver, and Maria Nicole Cleis, 'How the Right to Privacy Became a Human Right', *Human Rights Law Review*, 2014 <https://doi.org/10.1093/hrlr/ngu014>
Dijkstra, Arie, 'Personalization/Computer-Tailoring in Persuasive Technology: Tailoring Ingredients Target Psychological Processes.', in *PPT@ PERSUASIVE*, 2016, pp. 6–12
Ding, Xuan, Lan Zhang, Zhiguo Wan, and Ming Gu, 'A Brief Survey on De-Anonymization Attacks in Online Social Networks', in *2010 International Conference on Computational Aspects of Social Networks* (IEEE, 2010), pp. 611–15 <https://doi.org/10.1109/CASoN.2010.139>
Dive, Lisa, and Ainsley J Newson, 'Reconceptualizing Autonomy for Bioethics', *Kennedy Institute of Ethics Journal*, 28.2 (2018), 171–203
Dror, Yehezkel, 'Values and the Law', *The Antioch Review*, 17.4 (1957), 440 <https://doi.org/10.2307/4610000>
Dworkin, Gerald, 'Acting Freely', *Noûs*, 4.4 (1970), 367 <https://doi.org/10.2307/2214680>
———, 'Autonomy and Behavior Control', *The Hastings Center Report*, 6.1 (1976), 23 <https://doi.org/10.2307/3560358>
———, 'Paternalism', *The Stanford Encyclopedia of Philosophy*, 2020 <https://plato.stanford.edu/archives/fall2020/entries/paternalism/>
———, 'The Nature of Autonomy', *Nordic Journal of Studies in Educational Policy*, 2015.2 (2015), 28479 <https://doi.org/10.3402/nstep.v1.28479>
———, *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press,

1988)
Dwyer, Catherine, 'Task Technology Fit, the Social Technical Gap and Social Networking Sites', *AMCIS 2007 Proceedings*, 2007, 374
Egelman, Serge, A Felt, and David A Wagner, 'Choice Architecture and Smartphone Privacy: There's a Price for That', in *WEIS*, 2012
Elci, Meral, and Arzu Sert, 'Subliminal Advertising and Its Ethical Dimensions in the Social Media Age', in *Handbook of Research on Effective Advertising Strategies in the Social Media Age* (IGI Global, 2015), pp. 368–87
Enzweiler, Matthew James, 'Swatting Political Discource: A Domestic Terrorism Threat', *Notre Dame Law Review*, 90 (2015), 2001
Ericsson, *IoT Connection Outlook* <https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/iot-connections-outlook>
———, *Mobility Report*, 2021
Eser, Albin, 'Justification and Excuse Symposium: The New German Penal Code', *American Journal of Comparative Law*, 24.4 (1976), 621–37 <https://heinonline.org/HOL/P?h=hein.journals/amcomp24&i=629>
EU, Council of the, 'Press Release: Digital Services Act: Council and European Parliament Provisional Agreement for Making the Internet a Safer Space for European Citizens'
European Commission, 'Action Plan against Disinformation', 2018
European Court of Human Rights - Press Unit, 'Factsheet: New Technologies', 2022 <https://www.echr.coe.int/documents/fs_new_technologies_eng.pdf>
European Data Protection Board, *Guidelines 3/2022 on Dark Patterns in Social Media Platform Interfaces: How to Recognise and Avoid Them*, 2022 <https://edpb.europa.eu/system/files/2022-03/edpb_03-2022_guidelines_on_dark_patterns_in_social_media_platform_interfaces_en.pdf>
———, *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default*, 2020 <https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf>
European Parliament, *Amendments Adopted by the European Parliament on 20 January 2022 on the Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and Amending Directive 2000/31/EC (COM(2020)0825*, 2022
Fallis, D., 'Lying and Omissions', in *The Oxford Handbook of Lying*, ed. by J. Meibuaer (Oxford: Oxford University Press, 2018), pp. 183–92
Farley, B., and W. Clark, 'Simulation of Self-Organizing Systems by Digital Computer', *Transactions of the IRE Professional Group on Information Theory*, 4.4 (1954), 76–84 <https://doi.org/10.1109/TIT.1954.1057468>
Favaretto, Maddalena, Eva De Clercq, Jens Gaab, and Bernice Simone Elger, 'First Do No Harm: An Exploration of Researchers' Ethics of Conduct in Big Data Behavioral Studies', ed. by Daniel Jeremiah Hurst, *PLOS ONE*, 15.11 (2020), e0241865 <https://doi.org/10.1371/journal.pone.0241865>
Feather, N. T., and J. G. Simon, 'Causal Attributions for Success and Failure in Relation to Expectations of Success Based upon Selective or Manipulative Control', *Journal of Personality*, 39.4 (1971), 527–41 <https://doi.org/10.1111/j.1467-6494.1971.tb00060.x>
Feinberg, Joel, 'Autonomy', in *The Inner Citadel -Essays on Individual Autonomy*, ed. by John Christman (Oxford University Press)
———, 'Legal Paternalism', *Canadian Journal of Philosophy*, 1.1 (1971), 105–24
———, *The Moral Limits of the Criminal Law Volume 3: Harm to Self* (Oxford University Press, 1989) <https://doi.org/10.1093/0195059239.001.0001>
Feller, Avi, Emma Pierson, Sam Corbett-Davies, and Sharad Goel, 'A Computer Program Used

for Bail and Sentencing Decisions Was Labeled Biased against Blacks. It's Actually Not That Clear', *The Washington Post*, 17 (2016)
Fisher, Marc, John Woodrow Cox, and Peter Hermann, 'Pizzagate: From Rumor, to Hashtag, to Gunfire in DC', *Washington Post*, 6 (2016), 8410–15
Fishman, Clifford, 'Technology and the Internet: The Impending Destruction of Privacy by Betrayers, Grudgers, Snoops, Spammers, Corporations and the Media', *George Washington Law Review*, 72 (2004), 1503
Floarea, Aurel-Dorian, and Valentin Sgarciu, 'Smart Refrigerator: A next Generation Refrigerator Connected to the IoT', in *2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (IEEE, 2016), pp. 1–6 <https://doi.org/10.1109/ECAI.2016.7861170>
Floridi, L, 'Enveloping the World: The Constraining Success of Smart Technologies', *CEPE 2011: Crossing Boundaries*, 2011, 111
Floridi, Luciano, *Philosophy and Computing: An Introduction* (London/New York: Routledge, 1999)
———, 'Semantic Conceptions of Information', *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University, 2019)
———, *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality* (Oxford: Oxford University Press, 2014)
———, *The Onlife Manifesto: Being Human in a Hyperconnected Era* (Springer Nature, 2015)
Fogg, Brian Jeffrey, 'A Behavior Model for Persuasive Design', in *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09* (New York, New York, USA: ACM Press, 2009), p. 1 <https://doi.org/10.1145/1541948.1541999>
———, 'Persuasive Computers', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '98* (New York, New York, USA: ACM Press, 1998), pp. 225–32 <https://doi.org/10.1145/274644.274677>
———, *Persuasive Technology: Using Computers to Change What We Think and Do* (San Francisco: Morgan Kaufmann Publishers, 2003)
———, 'Persuasive Technology', *Ubiquity*, 2002.December (2002), 2 <https://doi.org/10.1145/764008.763957>
———, 'The Behavior Grid', in *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09* (New York, New York, USA: ACM Press, 2009), p. 1 <https://doi.org/10.1145/1541948.1542001>
Fradkov, Alexander L., 'Early History of Machine Learning', *IFAC-PapersOnLine*, 53.2 (2020), 1385–90 <https://doi.org/10.1016/j.ifacol.2020.12.1888>
Fragkos, Emmanouil, 'Question for Written Answer E-004810/2021 to the Commission (Legislation against the Manipulation of the Human Brain through Neuroscience)', 2019 <https://www.europarl.europa.eu/doceo/document/E-9-2021-004810_EN.pdf>
Frankfurt, Harry, 'Freedom of Will and Concept of a Person', *The Journal of Philosophy*, 68.1 (1971), 5–20
Franklin, Stan, and Art Graesser, 'Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents BT - Intelligent Agents III Agent Theories, Architectures, and Languages', ed. by Jörg P Müller, Michael J Wooldridge, and Nicholas R Jennings (Berlin, Heidelberg: Springer Berlin Heidelberg, 1997), pp. 21–35
Fried, Charles, 'Two Concepts of Interests: Some Reflections on the Supreme Court's Balancing Test', *Harvard Law Review*, 76.4 (1963), 755 <https://doi.org/10.2307/1338701>
Friedman, Batya, and Helen Nissenbaum, 'Bias in Computer Systems', *ACM Transactions on Information Systems*, 14.3 (1996), 330–47 <https://doi.org/10.1145/230538.230561>
Friedman, Marilyn, 'Autonomy, Gender, Politics' (Oxford University Press, 2003)

<https://doi.org/10.1093/0195138503.001.0001>
Frik, Alisa, and Luigi Mittone, 'Factors Influencing the Perception of Website Privacy Trustworthiness and Users' Purchasing Intentions: The Behavioral Economics Perspective', *Journal of Theoretical and Applied Electronic Commerce Research*, 14.3 (2019), 89–125
Fritz, Thomas, Elaine M. Huang, Gail C. Murphy, and Thomas Zimmermann, 'Persuasive Technology in the Real World', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM, 2014), pp. 487–96 <https://doi.org/10.1145/2556288.2557383>
Froehlich, Thomas J, 'A Not-So-Brief Account of Current Information Ethics: The Ethics of Ignorance, Missing Information, Misinformation, Disinformation and Other Forms of Deception or Incompetence.', *BiD*, 39, 2017
Froomkin, Michael, 'The Death of Privacy?', *Stanford Law Review*, 52.5 (2000), 1461 <https://doi.org/10.2307/1229519>
Fuchs, Helmut, *Strafrecht Allgemeiner Teil I*, 10th edn (Wien: Verlag Österreich, 2018)
Gable, Robert S., and Ralph Kirkland Gable†, 'Remaking the Electronic Tracking of Offenders into a "Persuasive Technology"', *Journal of Technology in Human Services*, 34.1 (2016), 13–31 <https://doi.org/10.1080/15228835.2016.1138839>
Gambs, Sébastien, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez, 'De-Anonymization Attack on Geolocated Data', *Journal of Computer and System Sciences*, 80.8 (2014), 1597–1614 <https://doi.org/10.1016/j.jcss.2014.04.024>
Garrett, R. Kelly, and Brian E. Weeks, 'The Promise and Peril of Real-Time Corrections to Political Misperceptions', in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW '13* (New York, New York, USA: ACM Press, 2013), p. 1047 <https://doi.org/10.1145/2441776.2441895>
Gartner, M., 'Regulatory Acknowledgment of Individual Autonomy in European Digital Legislation: From Meta-Principle to Explicit Protection in the Data Act', *European Data Protection Law Review*, 8.4 (2023), 462–73 <https://doi.org/10.21552/edpl/2022/4/6>
Gartner, Maximilian, 'Fit for the Future: A Pragmatic Account of Human Autonomy to Understand Emerging Issues in The Internet of Everything', in *Technology and Society: The Evolution of the Legal Landscape*, ed. by Marie Bourguignon, Tom Hick, Sofie Royer, and Ward Yperman (Antwerpen: Gompel&Svacina, 2021)
———, 'LBS as Vectors of Influence (Forthcoming)', in *Proceedings of the ICA*, 2021, p. tbd.
———, 'Research Proposal: Influenceable Autonomy and Predictable Freedom in the IoE', in *Proceedings of the Seventh JURIX 2019 Doctoral Consortium Co-Located with 32nd International Conference on Legal Knowledge and Information Systems (JURIX 2019)*, ed. by Monica Palmirani (CEUR Workshop Proceedings, 2018)
Gellert, Raphaël, and Serge Gutwirth, 'The Legal Construction of Privacy and Data Protection', *Computer Law & Security Review*, 29.5 (2013), 522–30 <https://doi.org/https://doi.org/10.1016/j.clsr.2013.07.005>
Ghaemi, Sara, and Guenther Brauner, 'User Behavior and Patterns of Electricity Use for Energy Saving', in *6. Internationale Energiewirtschaftstagung an Der TU Wien [Proceedings]*, 2009
Gigerenzer, Gerd, Klaus Fiedler, and Henrik Olsson, 'Rethinking Cognitive Biases as Environmental Consequences', *Ecological Rationality: Intelligence in the World*, 2012, 80–110
Gilbar, Roy, and José Miola, 'ONE SIZE FITS ALL? ON PATIENT AUTONOMY, MEDICAL DECISION-MAKING, AND THE IMPACT OF CULTURE', *Medical Law Review*, 23.3 (2015), 375–99 <https://doi.org/10.1093/medlaw/fwu032>
Glancy, Dorothy J, 'The Invention of the Right to Privacy', *Ariz. L. Rev.*, 21 (1979), 1

Goh, Yeow Chong, Xin Qing Cai, Walter Theseira, Giovanni Ko, and Khiam Aik Khor, 'Evaluating Human versus Machine Learning Performance in Classifying Research Abstracts', *Scientometrics*, 125.2 (2020), 1197–1212 <https://doi.org/10.1007/s11192-020-03614-2>

Goldfarb, Avi, and Catherine Tucker, 'Shifts in Privacy Concerns', *American Economic Review*, 102.3 (2012), 349–53 <https://doi.org/10.1257/aer.102.3.349>

Goodson, Ivor F., and J. Marshall Mangan, 'Computer Literacy as Ideology', *British Journal of Sociology of Education*, 17.1 (1996), 65–79 <https://doi.org/10.1080/0142569960170105>

Gray, Colin M., Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs, 'The Dark (Patterns) Side of UX Design', in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM, 2018), pp. 1–14 <https://doi.org/10.1145/3173574.3174108>

Gray, Colin M., Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford, 'Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective', in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM, 2021), pp. 1–18 <https://doi.org/10.1145/3411764.3445779>

Greenawalt, Kent, 'The Perplexing Borders of Justification and Excuse', *Columbia Law Review*, 84.8 (1984), 1897–1927

Greenberg, Saul, Sebastian Boring, Jo Vermeulen, and Jakub Dostal, 'Dark Patterns in Proxemic Interactions', in *Proceedings of the 2014 Conference on Designing Interactive Systems* (New York, NY, USA: ACM, 2014), pp. 523–32 <https://doi.org/10.1145/2598510.2598541>

Grifoni, Patrizia, Arianna D'Ulizia, and Fernando Ferri, 'Context-Awareness in Location Based Services in the Big Data Era', 2018, pp. 85–127 <https://doi.org/10.1007/978-3-319-67925-9_5>

Gumbis, Jaunius, Vytaute Bacianskaite, and Jurgita Randakeviciute, 'Do Human Rights Guarantee Autonomy?', *Cuadernos Constitucionales de La Cátedra Fadrique Furió Ceriol*, 62, 2008, 77–93

Guyer, Paul, 'Kant on the Theory and Practice of Autonomy', in *Kant's System of Nature and Freedom - Selected Essays*, ed. by Paul Guyer (Oxford: Oxford University Press, 2005), pp. 115–45

Hadjistavropoulos, Thomas, and David Cruise Malloy, 'Ethical Principles of the American Psychological Association: An Argument for Philosophical and Practical Ranking', *Ethics & Behavior*, 9.2 (1999), 127–40 <https://doi.org/10.1207/s15327019eb0902_4>

Hagendorff, Thilo, 'The Ethics of AI Ethics: An Evaluation of Guidelines', *Minds and Machines*, 30.1 (2020), 99–120 <https://doi.org/10.1007/s11023-020-09517-8>

Hagras, Hani, 'Toward Human-Understandable, Explainable AI', *Computer*, 51.9 (2018), 28–36

Hall, Margaret Isabel, 'Mental Capacity in the (Civil) Law: Capacity, Autonomy, and Vulnerability', *McGill Law Journal*, 58.1 (2013), 61–94 <https://doi.org/10.7202/1013386ar>

Hallowell, N, 'Balancing Autonomy and Responsibility: The Ethics of Generating and Disclosing Genetic Information', *Journal of Medical Ethics*, 29.2 (2003), 74–79 <https://doi.org/10.1136/jme.29.2.74>

Hamilton, Isobel Asher, 'Amazon Built an AI Tool to Hire People but Had to Shut It down Because It Was Discriminating against Women', *Business Insider*, 2018

Hampton, Jean, 'Contractarianism', *The Cambridge Dictionary of Philosophy* (Cambridge University Press, 1999)

Hans, Jonas, ‘Technology as a Subject for Ethics’, *Social Research*, 49.4 (1982), 891–98 <http://www.jstor.org/stable/40971222>

Hanssen, Ingrid, ‘An Intercultural Nursing Perspective on Autonomy’, *Nursing Ethics*, 11.1 (2004), 28–41 <https://doi.org/10.1191/0969733004ne664oa>

———, ‘From Human Ability to Ethical Principle: An Intercultural Perspective on Autonomy’, *Medicine, Health Care and Philosophy*, 7.3 (2005), 269–79 <https://doi.org/10.1007/s11019-004-9032-1>

Harbo, Tor-Inge, ‘The Function of the Proportionality Principle in EU Law’, *European Law Journal*, 16.2 (2010), 158–85 <https://doi.org/10.1111/j.1468-0386.2009.00502.x>

Harrington, Donna, and Ralph Dolgoff, ‘Hierarchies of Ethical Principles for Ethical Decision Making in Social Work’, *Ethics and Social Welfare*, 2.2 (2008), 183–96 <https://doi.org/10.1080/17496530802117680>

Hart, Henry M Jr., ‘The Aims of the Criminal Law Sentencing’, *Law and Contemporary Problems*, 23.3 (1958), 401–41 <https://heinonline.org/HOL/P?h=hein.journals/lcp23&i=419>

Hawkins, Richard, Robin Mansell, and W Edward Steinmueller, ‘Toward Digital Intermediation in the Information Society’, *Journal of Economic Issues*, 33.2 (1999), 383–91

HAYNE, STEPHEN C., and RONALD E. RICE, ‘Attribution Accuracy When Using Anonymity in Group Support Systems’, *International Journal of Human-Computer Studies*, 47.3 (1997), 429–52 <https://doi.org/10.1006/ijhc.1997.0134>

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, ‘Deep Residual Learning for Image Recognition’, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–78

———, ‘Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification’, in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–34

He, Yi, Jian-chao Han, and Shao-hua Zeng, ‘Classification Algorithm Based on Improved ID3 in Bank Loan Application’, in *Information Engineering and Applications* (Springer, 2012), pp. 1124–30

Herse, Sarita, Jonathan Vitale, Meg Tonkin, Daniel Ebrahimian, Suman Ojha, Benjamin Johnston, and others, ‘Do You Trust Me, Blindly? Factors Influencing Trust Towards a Robot Recommender System’, in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (IEEE, 2018), pp. 7–14 <https://doi.org/10.1109/ROMAN.2018.8525581>

Higgins, R., ‘Derogations under Human Rights Treaties’, *British Yearbook of International Law*, 48.1 (1977), 281–319 <https://doi.org/10.1093/bybil/48.1.281>

High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 2019 <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419>

Hill Jr, Thomas E, ‘Autonomy and Benevolent Lies’, *J. Value Inquiry*, 18 (1984), 251

Hoffrage, Ulrich, and Julian N. Marewski, ‘Social Scoring Als Mensch-System-Interaktion’, in *Social Credit Rating* (Wiesbaden: Springer Fachmedien Wiesbaden, 2020), pp. 305–29 <https://doi.org/10.1007/978-3-658-29653-7_17>

Hofstede, Geert, *Culture’s Consequences: International Differences in Work-Related Values* (sage, 1984), v

———, ‘Dimensionalizing Cultures: The Hofstede Model in Context’, *Online Readings in Psychology and Culture*, 2.1 (2011), 919–2307

Holvast, Jan, ‘History of Privacy’, ed. by Vashek Matyáš, Simone Fischer-Hübner, Daniel Cvrček, and Petr Švenda (Berlin, Heidelberg: Springer Berlin Heidelberg, 2009), pp. 13–

42
Holzinger, Andreas, ‘From Machine Learning to Explainable AI’, in *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, 2018, pp. 55–66
Hornsby, Adrian, and Rod Walsh, ‘From Instant Messaging to Cloud Computing, an XMPP Review’, in *IEEE International Symposium on Consumer Electronics (ISCE 2010)*, 2010, pp. 1–6
Hosanagar, Kartik, Ramayya Krishnan, and Liye Ma, ‘Recomended for You: The Impact of Profit Incentives on the Relevance of Online Recommendations’, in *ICIS 2008 Proceedings*, 2008, p. 31
Hoskins, Guy Thurston, ‘Beyond “Zero Sum”: The Case for Context in Regulating Zero Rating in the Global South’, *Internet Policy Review*, 8.1 (2019) <https://doi.org/10.14763/2019.1.1392>
Hou, Yubo, Dan Xiong, Tonglin Jiang, Lily Song, and Qi Wang, ‘Social Media Addiction: Its Impact, Mediation, and Intervention’, *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 13.1 (2019) <https://doi.org/10.5817/CP2019-1-4>
Huang, Haosheng, Georg Gartner, Jukka M. Krisp, Martin Raubal, and Nico Van de Weghe, ‘Location Based Services: Ongoing Evolution and Research Agenda’, *Journal of Location Based Services*, 12.2 (2018), 63–93 <https://doi.org/10.1080/17489725.2018.1508763>
Hung, Wei-Tien, and Guo-Bao Liou, ‘The Influence of Social Media on Outdoor Recreation Participation’, 2022
Hyland, Ken, ‘Academic Publishing and the Myth of Linguistic Injustice’, *Journal of Second Language Writing*, 31 (2016), 58–69 <https://doi.org/10.1016/j.jslw.2016.01.005>
Icoz, Orhan, An\il Kutuk, Onur Icoz, and others, ‘Social Media and Consumer Buying Decisions in Tourism: The Case of Turkey’, 2018
Ienca, Marcello, and Roberto Andorno, ‘Towards New Human Rights in the Age of Neuroscience and Neurotechnology’, *Life Sciences, Society and Policy*, 13.1 (2017), 1–27
International Telecommunication Union, *Measuring Digital Development - Facts and Figures*, 2020
———, ‘Measuring Digital Developments: Facts and Figures 2020’, 2020
Iyengar, Sheena S., and Mark R. Lepper, ‘When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?’, *Journal of Personality and Social Psychology*, 79.6 (2000), 995–1006 <https://doi.org/10.1037/0022-3514.79.6.995>
J Williams, Alvin, and BabuP George, ‘Truck Drivers--the under-Respected Link in the Supply Chain: A Quasi-Ethnographic Perspective Using Qualitative Appreciative Inquiry’, *Operations and Supply Chain Management: An International Journal*, 6.3 (2014), 85–93
Jacobs, Naomi, ‘Two Ethical Concerns about the Use of Persuasive Technology for Vulnerable People’, *Bioethics*, 34.5 (2020), 519–26 <https://doi.org/10.1111/bioe.12683>
Jennings, Bruce, ‘Reconceptualizing Autonomy: A Relational Turn in Bioethics’, *Hastings Center Report*, 46.3 (2016), 11–16
Jensen, Herbert L., ‘Using Neural Networks for Credit Scoring’, *Managerial Finance*, 18.6 (1992), 15–26 <https://doi.org/10.1108/eb013696>
Ji, Shouling, Prateek Mittal, and Raheem Beyah, ‘Graph Data Anonymization, De-Anonymization Attacks, and De-Anonymizability Quantification: A Survey’, *IEEE Communications Surveys & Tutorials*, 19.2 (2017), 1305–26 <https://doi.org/10.1109/COMST.2016.2633620>
Jiang, Tonglin, Yubo Hou, and Qi Wang, ‘Does Micro-Blogging Make Us “Shallow”? Sharing Information Online Interferes with Information Comprehension’, *Computers in Human Behavior*, 59 (2016), 210–14 <https://doi.org/10.1016/j.chb.2016.02.008>
Jobin, Anna, Marcello Ienca, and Effy Vayena, ‘The Global Landscape of AI Ethics

Guidelines', *Nature Machine Intelligence*, 1.9 (2019), 389–99
Joffe, Steven, M Manocchia, J C Weeks, and P D Cleary, 'What Do Patients Value in Their Hospital Care? An Empirical Perspective on Autonomy Centred Bioethics', *Journal of Medical Ethics*, 29.2 (2003), 103–8
Johnson, Hollyn M., and Colleen M. Seifert, 'Sources of the Continued Influence Effect: When Misinformation in Memory Affects Later Inferences.', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20.6 (1994), 1420–36 <https://doi.org/10.1037/0278-7393.20.6.1420>
Jones-Jang, S. Mo, Tara Mortensen, and Jingjing Liu, 'Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't', *American Behavioral Scientist*, 65.2 (2021), 371–88 <https://doi.org/10.1177/0002764219869406>
Joseph, Sarah, Jenny Schultz, and Melissa Castan, 'The International Covenant on Civil and Political Rights: Cases, Materials, and Commentaries', *The American Journal of International Law*, 2004 <https://doi.org/10.2307/3216728>
Juth, Niklas, and Frank Lorentzon, 'The Concept of Free Will and Forensic Psychiatry', *International Journal of Law and Psychiatry*, 33.1 (2010), 1–6
Kahrs, Hans J, 'Kausalität Und Überholende Kausalität Im Zivilrecht', in *Kausalität Und Überholende Kausalität Im Zivilrecht* (De Gruyter, 2015)
Kang, Ruogu, Stephanie Brown, and Sara Kiesler, 'Why Do People Seek Anonymity on the Internet?', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM, 2013), pp. 2657–66 <https://doi.org/10.1145/2470654.2481368>
Kant, Immanuel, *Groundwork for the Metaphysics of Morals* (New Haven: Yale University Press, 2002)
———, 'Grundlegung Zur Metaphysik Der Sitten', in *Immanuel Kant: Werke in Zwölf Bänden. Band 7*, ed. by Wilhelm Weischedel (Frankfurt am Main: Suhrkamp, 1977)
Kapantai, Eleni, Androniki Christopoulou, Christos Berberidis, and Vassilios Peristeras, 'A Systematic Literature Review on Disinformation: Toward a Unified Taxonomical Framework', *New Media & Society*, 23.5 (2021), 1301–26 <https://doi.org/10.1177/1461444820959296>
Kaptein, Maurits Clemens, 'Formalizing Customization in Persuasive Technologies', in *PERSUASIVE 2015*, ed. by T. Mactavish and S. Basapur, 2015, pp. 27–38
Kaptein, Maurits Clemens, Panos Markopoulos, Boris de Ruyter, and Emile Aarts, 'Persuasion in Ambient Intelligence', *Journal of Ambient Intelligence and Humanized Computing*, 1.1 (2010), 43–56 <https://doi.org/10.1007/s12652-009-0005-3>
Kaptein, Maurits, and Dean Eckles, 'Selecting Effective Means to Any End: Futures and Ethics of Persuasion Profiling', 2010, pp. 82–93 <https://doi.org/10.1007/978-3-642-13226-1_10>
Kapust, Daniel, 'Skinner, Pettit and Livy: The Conflict of the Orders and the Ambiguity of Republican Liberty', *History of Political Thought*, 25.3 (2004), 377–401
Kara, M. A., 'Applicability of the Principle of Respect for Autonomy: The Perspective of Turkey', *Journal of Medical Ethics*, 33.11 (2007), 627–30 <https://doi.org/10.1136/jme.2006.017400>
Karger, David R., and William Jones, 'Data Unification in Personal Information Management', *Communications of the ACM*, 49.1 (2006), 77–82 <https://doi.org/10.1145/1107458.1107496>
Karlita, Tita, Nadia Azahro Choirunisa, Rengga Asmara, and Fitri Setyorini, 'Cat Breeds Classification Using Compound Model Scaling Convolutional Neural Networks.', 2022 <https://doi.org/10.2991/assehr.k.220301.150>
Kasper, Debbie V S, 'The Evolution (Or Devolution) of Privacy', *Sociological Forum*, 20.1

(2005), 69–92 <http://www.jstor.org/stable/4540882>
Katsikari, Christina, Leonidas Hatzithomas, Thomas Fotiadis, and Dimitrios Folinas, 'Push and Pull Travel Motivation: Segmentation of the Greek Market for Social Media Marketing in Tourism', *Sustainability*, 12.11 (2020), 4770
Keding, Christoph, and Philip Meissner, 'Managerial Overreliance on AI-Augmented Decision-Making Processes: How the Use of AI-Based Advisory Systems Shapes Choice Behavior in R&D Investment Decisions', *Technological Forecasting and Social Change*, 171 (2021), 120970 <https://doi.org/10.1016/j.techfore.2021.120970>
Kelders, Saskia, 'Involvement as a Working Mechanism for Persuasive Technology', in *PERSUASIVE 2015*, ed. by T. MacTavish and S. Basapur, 2015, pp. 3–14 <https://doi.org/DOI: 10.1007/978-3-319-20306-5_1>
Kelsen, Hans, 'Causality and Imputation', *Ethics*, LXI.1 (1950)
Kelton, Kari, Kenneth R. Fleischmann, and William A. Wallace, 'Trust in Digital Information', *Journal of the American Society for Information Science and Technology*, 59.3 (2008), 363–74 <https://doi.org/10.1002/asi.20722>
Kight, Raymond, and Sandra Burri Gram-Hansen, 'Do Ethics Matter in Persuasive Technology?', 2019, pp. 143–55 <https://doi.org/10.1007/978-3-030-17287-9_12>
Klasnja, Predrag, and Wanda Pratt, 'Healthcare in the Pocket: Mapping the Space of Mobile-Phone Health Interventions', *Journal of Biomedical Informatics*, 45.1 (2012), 184–98 <https://doi.org/10.1016/j.jbi.2011.08.017>
Kleene, S. C., 'Representation of Events in Nerve Nets and Finite Automata', in *Automata Studies. (AM-34)* (Princeton University Press, 1956), pp. 3–42 <https://doi.org/10.1515/9781400882618-002>
Klumpp, Matthias, 'Automation and Artificial Intelligence in Business Logistics Systems: Human Reactions and Collaboration Requirements', *International Journal of Logistics Research and Applications*, 21.3 (2018), 224–42
Knijnenburg, Bart P., Martijn C. Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell, 'Explaining the User Experience of Recommender Systems', *User Modeling and User-Adapted Interaction*, 22.4–5 (2012), 441–504 <https://doi.org/10.1007/s11257-011-9118-4>
Koene, Ansgar, Elvira Perez, Christopher James Carter, Ramona Statache, Svenja Adolphs, Claire O'Malley, and others, 'Ethics of Personalized Information Filtering', 2015, pp. 123–32 <https://doi.org/10.1007/978-3-319-18609-2_10>
Kokolakis, Spyros, 'Privacy Attitudes and Privacy Behaviour: A Review of Current Research on the Privacy Paradox Phenomenon', *Computers & Security*, 64 (2017), 122–34 <https://doi.org/10.1016/j.cose.2015.07.002>
Kokott, J., and C. Sobotta, 'The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR', *International Data Privacy Law*, 3.4 (2013), 222–28 <https://doi.org/10.1093/idpl/ipt017>
Kolenik, Tine, and Matjaz Gams, 'Persuasive Technology for Mental Health: One Step Closer to (Mental Health Care) Equality?', *IEEE Technology and Society Magazine*, 40.1 (2021), 80–86 <https://doi.org/10.1109/MTS.2021.3056288>
Koops, Bert-Jaap, Bryce Clayton Newell, Tjerk Timan, Ivan Skorvanek, Tomislav Chokrevski, and Masa Galic, 'A Typology of Privacy ', *University of Pennsylvania Journal of International Law*, 38.2, 483–576 <https://heinonline.org/HOL/P?h=hein.journals/upjiel38&i=489>
Kopec, Matthew, 'Deceptive Omissions, Half-Truths, and the Moral Exemplar in Clinical Ethics', *The American Journal of Bioethics*, 21.5 (2021), 33–35 <https://doi.org/10.1080/15265161.2021.1906993>
Kosters, Mark, and Jeroen Van der Heijden, 'From Mechanism to Virtue: Evaluating Nudge

Theory', *Evaluation*, 21.3 (2015), 276–91 <https://doi.org/10.1177/1356389015590218>
Kramer, A. D. I., J. E. Guillory, and J. T. Hancock, 'Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks', *Proceedings of the National Academy of Sciences*, 111.24 (2014), 8788–90 <https://doi.org/10.1073/pnas.1320040111>
Kramer, Matthew, *Where Law and Morality Meet* (Oxford: Oxford University Press, 2004)
Krischkowsky, Alina, Bernhard Maurer, and Manfred Tscheligi, 'Captology and Technology Appropriation: Unintended Use as a Source for Designing Persuasive Technologies', 2016, pp. 78–83 <https://doi.org/10.1007/978-3-319-31510-2_7>
Kronman, Anthony T, 'Paternalism and the Law of Contracts', *Yale Lj*, 92 (1982), 763
Kühling, Jürgen, and Manuel Klar, 'Privatheit Und Datenschutz in Der EU Und Den USA – Kollision Zweier Welten?', *Archiv Des Öffentlichen Rechts*, 141.2 (2016), 165 <https://doi.org/10.1628/000389116X14684978889181>
Kuhn, Steven, 'Formalism', *The Cambridge Dictionary of Philosophy* (Cambridge University Press, 1999)
Kuncel, Nathan R, David M Klieger, and Deniz S Ones, 'In Hiring, Algorithms Beat Instinct', *Harvard Business Review*, 92.5 (2014), p32--32
Kupfer, Joseph, 'Privacy, Autonomy, and Self-Concept', *American Philosophical Quarterly*, 24.1 (1987), 81–89
Kwon, Joon-Myoung, Ki-Hyun Jeon, Hyue Mee Kim, Min Jeong Kim, Sung Min Lim, Kyung-Hee Kim, and others, 'Comparing the Performance of Artificial Intelligence and Conventional Diagnosis Criteria for Detecting Left Ventricular Hypertrophy Using Electrocardiography', *EP Europace*, 22.3 (2020), 412–19 <https://doi.org/10.1093/europace/euz324>
Lai, Kenneth, Xinyuan Tu, and Svetlana Yanushkevich, 'Dog Identification Using Soft Biometrics and Neural Networks', in *2019 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2019), pp. 1–8 <https://doi.org/10.1109/IJCNN.2019.8851971>
Lai, Poey Chin, 'The Literature Review of Technology Adoption Models and Theories for the Novelty Technology', *JISTEM-Journal of Information Systems and Technology Management*, 14 (2017), 21–38
Lanzing, Marjolein, '"Strongly Recommended" Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies', *Philosophy & Technology*, 32.3 (2019), 549–68 <https://doi.org/10.1007/s13347-018-0316-4>
Larsson, Stefan, 'On the Governance of Artificial Intelligence through Ethics Guidelines', *Asian Journal of Law and Society*, 7.3 (2020), 437–51
Lavrusik, Vadim, 'Washington Post Announces Personalized News Aggregation Site', *Mashable*, 2011
Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, and others, 'The Science of Fake News', *Science*, 359.6380 (2018), 1094–96 <https://doi.org/10.1126/science.aao2998>
Lee, J. D., and K. A. See, 'Trust in Automation: Designing for Appropriate Reliance', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46.1 (2004), 50–80 <https://doi.org/10.1518/hfes.46.1.50_30392>
Lee, Min Kyung, 'Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management', *Big Data & Society*, 5.1 (2018), 205395171875668 <https://doi.org/10.1177/2053951718756684>
Leese, Matthias, 'The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-Discriminatory Safeguards in the European Union', *Security Dialogue*, 45.5 (2014), 494–511 <https://doi.org/10.1177/0967010614544204>
Lehrer, Keith, 'Freedom, Preference and Autonomy', *The Journal of Ethics*, 1.1 (1997), 3–25

<http://www.jstor.org/stable/25115533>
Lenaerts, Koen, and José A Gutiérrez-Fons, 'To Say What the Law of the EU Is: Methods of Interpretation and the European Court of Justice', *Colum. J. Eur. L.*, 20 (2013), 3
Letsas, George, 'The ECHR as a Living Instrument: Its Meaning and Legitimacy', *Constituting Europe: The European Court of Human Rights in a National, European and Global Context*, 2 (2013), 106
Lewandowsky, Stephan, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook, 'Misinformation and Its Correction: Continued Influence and Successful Debiasing', *Psychological Science in the Public Interest*, 13.3 (2012), 106–31 <https://doi.org/10.1177/1529100612451018>
Li, Huaxin, Qingrong Chen, Haojin Zhu, Di Ma, Hong Wen, and Xuemin Shen, 'Privacy Leakage via De-Anonymization and Aggregation in Heterogeneous Social Networks', *IEEE Transactions on Dependable and Secure Computing*, 17.2 (2020), 350–62 <https://doi.org/10.1109/TDSC.2017.2754249>
Lockton, Dan, David H.T. Harrison, and Neville A Stanton, 'Choice Architecture and Design with Intent', 2009 <https://doi.org/10.14236/ewic/NDM2009.59>
Lockton, Dan, David Harrison, and Neville Stanton, 'Design with Intent: Persuasive Technology in a Wider Context', in *Persuasive Technology* (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 274–78 <https://doi.org/10.1007/978-3-540-68504-3_30>
Loftus, E. F., 'Planting Misinformation in the Human Mind: A 30-Year Investigation of the Malleability of Memory', *Learning & Memory*, 12.4 (2005), 361–66 <https://doi.org/10.1101/lm.94705>
Loske, Dominic, and Matthias Klumpp, 'Intelligent and Efficient? An Empirical Analysis of Human--AI Collaboration for Truck Drivers in Retail Logistics', *The International Journal of Logistics Management*, 2021
Lowry, Paul Benjamin, Jun Zhang, Chuang Wang, and Mikko Siponen, 'Why Do Adults Engage in Cyberbullying on Social Media? An Integration of Online Disinhibition and Deindividuation Effects with the Social Structure and Social Learning Model', *Information Systems Research*, 27.4 (2016), 962–86 <https://doi.org/10.1287/isre.2016.0671>
Lu, Sylvia, 'Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence ', *Vanderbilt Journal of Entertainment & Technology Law*, 23.1, 99–160 <https://heinonline.org/HOL/P?h=hein.journals/vanep23&i=111>
Di Luzio, Adriano, Alessandro Mei, and Julinda Stefa, 'Mind Your Probes: De-Anonymization of Large Crowds through Smartphone WiFi Probe Requests', in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications* (IEEE, 2016), pp. 1–9 <https://doi.org/10.1109/INFOCOM.2016.7524459>
Lyons, Joseph B., Garrett G. Sadler, Kolina Koltai, Henri Battiste, Nhut T. Ho, Lauren C. Hoffmann, and others, 'Shaping Trust Through Transparent Design: Theoretical and Experimental Guidelines', 2017, pp. 127–36 <https://doi.org/10.1007/978-3-319-41959-6_11>
Ma-Kellams, Christine, 'Cultural Variation and Similarities in Cognitive Thinking Styles Versus Judgment Biases: A Review of Environmental Factors and Evolutionary Forces', *Review of General Psychology*, 24.3 (2020), 238–53 <https://doi.org/10.1177/1089268019901270>
Maan, Saskia, Bo Merkus, Jaap Ham, and Cees Midden, 'Making It Not Too Obvious: The Effect of Ambient Light Feedback on Space Heating Energy Consumption', *Energy Efficiency*, 4.2 (2011), 175–83 <https://doi.org/10.1007/s12053-010-9102-6>
MacCallum, Gerald C, 'Negative and Positive Freedom', in *The Liberty Reader* (Routledge, 2017), pp. 100–122

MacDonald, Diana, ‘Anti-Patterns and Dark Patterns’, in *Practical UI Patterns for Design Systems* (Berkeley, CA: Apress, 2019), pp. 193–221 <https://doi.org/10.1007/978-1-4842-4938-3_5>

Mackenzie, Catriona, ‘Relational Autonomy, Normative Authority and Perfectionism’, *Journal of Social Philosophy*, 39.4 (2008), 512–33 <https://doi.org/10.1111/j.1467-9833.2008.00440.x>

Mackenzie, Catriona, and Natalie Stoljar, ‘Introduction - Autonomy Refigured’, in *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*, ed. by Catriona Mackenzie and Natalie Stoljar (Oxford: Oxford University Press, 2000), pp. 3–35

Maclean, Alisdair, ‘Autonomy, Consent and Persuasion’, *European Journal of Health Law*, 13.4 (2006), 321–38 <https://doi.org/10.1163/157180906779160274>

Maier, Maximilian, František Bartoš, T. D. Stanley, David R. Shanks, Adam J. L. Harris, and Eric-Jan Wagenmakers, ‘No Evidence for Nudging after Adjusting for Publication Bias’, *Proceedings of the National Academy of Sciences*, 119.31 (2022) <https://doi.org/10.1073/pnas.2200300119>

Manson, Marshall, ‘Facebook Zero: Considering Life after the Demise of Organic Reach’, *Social@ Ogilvy White Paper*, 2014, 1–7

Marriam-Webster (Dictionary), ‘Disinformation’, 2021

Martin, Kirsten, and Katie Shilton, ‘Putting Mobile Application Privacy in Context: An Empirical Study of User Privacy Expectations for Mobile Devices’, *The Information Society*, 32.3 (2016), 200–216 <https://doi.org/10.1080/01972243.2016.1153012>

Mason, Christina, Peter van der Putten, and Max van Duijn, ‘How Identity and Uncertainty Affect Online Social Influence’, 2020, pp. 174–90 <https://doi.org/10.1007/978-3-030-61841-4_12>

Mason, Richard O., ‘Applying Ethics to Information Technology Issues’, *Communications of the ACM*, 38.12 (1995), 55–57 <https://doi.org/10.1145/219663.219681>

Matthews, John, Khin Than Win, Harri Oinas-Kukkonen, and Mark Freeman, ‘Persuasive Technology in Mobile Applications Promoting Physical Activity: A Systematic Review’, *Journal of Medical Systems*, 40.3 (2016), 72 <https://doi.org/10.1007/s10916-015-0425-x>

Mavronicola, Natasa, ‘What Is an “Absolute Right”? Deciphering Absoluteness in the Context of Article 3 of the European Convention on Human Rights’, *Human Rights Law Review*, 12.4 (2012), 723–58

Mayer, Richard E., Emily Griffith, Ilana T. N. Jurkowitz, and Daniel Rothman, ‘Increased Interestingness of Extraneous Details in a Multimedia Science Presentation Leads to Decreased Learning.’, *Journal of Experimental Psychology: Applied*, 14.4 (2008), 329–39 <https://doi.org/10.1037/a0013835>

McCarthy, John, *Ascribing Mental Qualities to Machines.*, 1979

McClain, Linda C, ‘Inviolability and Privacy: The Castle, the Sanctuary, and the Body Symposium: The Sacred Body in Law and Literature’, *Yale Journal of Law & the Humanities*, 7.1 (1995), 195–242 <https://heinonline.org/HOL/P?h=hein.journals/yallh7&i=207>

McCoy, Matthew, ‘Autonomy, Consent, and Medical Paternalism: Legal Issues in Medical Intervention’, *The Journal of Alternative and Complementary Medicine*, 14.6 (2008), 785–92

McKenna, Michael, ‘Compatibilism’, *The Stanford Encyclopedia of Philosophy*, 2004

Mead, M, and R Métraux, ‘The Study of Culture at a Distance. Chicago: Univ’ (Chicago Press, 1953)

Mele, Alfred, ‘History and Personal Autonomy’, *Canadian Journal of Philosophy*, 23.2 (1991),

271–80
Metzger, Miriam J., and Andrew J. Flanagin, 'Credibility and Trust of Information in Online Environments: The Use of Cognitive Heuristics', *Journal of Pragmatics*, 59 (2013), 210–20 <https://doi.org/10.1016/j.pragma.2013.07.012>
Meynen, Gerben, 'Autonomy, Criminal Responsibility, and Competence', *Journal of the American Academy of Psychiatry and the Law Online*, 39.2 (2011), 231–36 <https://jaapl.org/content/39/2/231>
Midden, Cees, and Jaap Ham, 'The Illusion of Agency: The Influence of the Agency of an Artificial Agent on Its Persuasive Power', 2012, pp. 90–99 <https://doi.org/10.1007/978-3-642-31037-9_8>
Middlebrook, Callie, 'The Grey Area: Instagram, Shadowbanning, and the Erasure of Marginalized Communities', *SSRN Electronic Journal*, 2020 <https://doi.org/10.2139/ssrn.3539721>
Mik, Eliza, 'The Erosion of Autonomy in Online Consumer Transactions', *Law, Innovation and Technology*, 8.1 (2016), 1–38 <https://doi.org/10.1080/17579961.2016.1161893>
Milano, Silvia, Mariarosaria Taddeo, and Luciano Floridi, 'Recommender Systems and Their Ethical Challenges', *AI & SOCIETY*, 35.4 (2020), 957–67 <https://doi.org/10.1007/s00146-020-00950-y>
Mill, John Stuart, *On Liberty* (Kitchener, Ontario: Batoche Books Limited)
Mincholé, Ana, and Blanca Rodriguez, 'Artificial Intelligence for the Electrocardiogram', *Nature Medicine*, 25.1 (2019), 22–23 <https://doi.org/10.1038/s41591-018-0306-1>
Mingers, John, and John Brocklesby, 'Multimethodology: Towards a Framework for Mixing Methodologies', *Omega*, 25.5 (1997), 489–509 <https://doi.org/10.1016/S0305-0483(97)00018-2>
Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi, 'The Ethics of Algorithms: Mapping the Debate', *Big Data and Society*, 3.2 (2016) <https://doi.org/10.1177/2053951716679679>
MMller, Karsten, and Carlo Schwarz, 'Making America Hate Again? Twitter and Hate Crime Under Trump', *SSRN Electronic Journal*, 2018 <https://doi.org/10.2139/ssrn.3149103>
Moka-Mubelo, Willy, *Reconciling Law and Morality in Human Rights Discourse* (Cham: Springer, 2017)
Montgomery, Alan L., and Michael D. Smith, 'Prospects for Personalization on the Internet', *Journal of Interactive Marketing*, 23.2 (2009), 130–37 <https://doi.org/10.1016/j.intmar.2009.02.001>
Moore, Gordon E., 'Cramming More Components onto Integrated Circuits', *Electronics*, 38.8 (1965), 114ff.
Muraven, Mark, and Roy F. Baumeister, 'Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle?', *Psychological Bulletin*, 126.2 (2000), 247–59 <https://doi.org/10.1037/0033-2909.126.2.247>
Murmann, Uwe, 'Problems of Causation with Regard to (Potential) Actions of Multiple Protagonists', *Journal of International Criminal Justice*, 12.2 (2014), 283–94
Narayanan, Arvind, Arunesh Mathur, Marshini Chetty, and Mihir Kshirsagar, 'Dark Patterns: Past, Present, and Future', *Queue*, 18.2 (2020), 67–92 <https://doi.org/10.1145/3400899.3400901>
Nass, Clifford, and Youngme Moon, 'Machines and Mindlessness: Social Responses to Computers', *Journal of Social Issues*, 56.1 (2000), 81–103 <https://doi.org/10.1111/0022-4537.00153>
Naughton, John, 'The Evolution of the Internet: From Military Experiment to General Purpose Technology', *Journal of Cyber Policy*, 1.1 (2016), 5–28 <https://doi.org/10.1080/23738871.2016.1157619>

Nawratek, Krzysztof, ‘Top-Down Revolutions. Negative and Positive Autonomy’, in *Total Urban Mobilisation* (Singapore: Springer Singapore, 2019), pp. 37–46 <https://doi.org/10.1007/978-981-13-1093-5_5>

Nelson, Mark R, ‘We Have the Information You Want, but Getting It Will Cost You! Held Hostage by Information Overload.’, *XRDS: Crossroads, The ACM Magazine for Students*, 1.1 (1994), 11–15

Neuhouser, Frederick, ‘Freedom, Dependence, and the General Will’, *The Philosophical Review*, 102.3 (1993), 363–95 <https://doi.org/10.2307/2185902>

Neuman, Gerald L, ‘American Convention on Human Rights’, *Max Planck Encyclopedia of Public International Law*, 2010

Newman, David T, Nathanael J Fast, and Derek J Harmon, ‘When Eliminating Bias Isn’t Fair: Algorithmic Reductionism and Procedural Justice in Human Resource Decisions’, *Organizational Behavior and Human Decision Processes*, 160 (2020), 149–67

Nkonde, Mutale, Maria Y. Rodriguez, Leonard Cortana, Joan K. Mukogosi, Shakira King, Ray Serrato, and others, ‘Disinformation Creep: ADOS and the Strategic Weapon-Ization of Breaking News’, *Harvard Kennedy School Misinformation Review*, 2021 <https://doi.org/10.37016/mr-2020-52>

Noack, Brent, and others, ‘Big Data Analytics in Human Resource Management: Automated Decision-Making Processes, Predictive Hiring Algorithms, and Cutting-Edge Workplace Surveillance Technologies’, *Psychosociological Issues in Human Resource Management*, 7.2 (2019), 37–42

Norris, Pippa, *Digital Divide* (Cambridge University Press, 2001) <https://doi.org/10.1017/CBO9781139164887>

Nouwens, Midas, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal, ‘Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence’, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM, 2020), pp. 1–13 <https://doi.org/10.1145/3313831.3376321>

Nowak, Manfred, *U.N. Covenant on Civil and Political Rights - CCPR Commentary*, 2nd edn (Kehl: N.P. Engel, Publisher, 2005)

Nygren, Thomas, Jenny Wiksten Folkeryd, Caroline Liberg, and Mona Guath, ‘Students Assessing Digital News and Misinformation’, 2020, pp. 63–79 <https://doi.org/10.1007/978-3-030-61841-4_5>

O’Keefe, D., *Persuasion: Theory and Research*, 2nd edn (Beverly Hills: Sage Publications, 2002)

O’Keefe, Daniel, ‘Persuasion’, in *The Handbook of Communication Skills*, ed. by Owen Hargie, 4th edn (New York: Routledge, 2018), pp. 319–37

Odom, Marcus D, Anand Kumar, and Laura Saunders, ‘Web Assurance Seals: How and Why They Influence Consumers’ Decisions’, *Journal of Information Systems*, 16.2 (2002), 231–50

Ogorek, Regina, *Richterkönig Oder Subsumtionsautomat?: Zur Justiztheorie Im 19. Jahrhundert* (Frankfurt am Main: Vittorio Klostermann, 2008)

ÓhÉigeartaigh, Seán S., Jess Whittlestone, Yang Liu, Yi Zeng, and Zhe Liu, ‘Overcoming Barriers to Cross-Cultural Cooperation in AI Ethics and Governance’, *Philosophy & Technology*, 33.4 (2020), 571–93 <https://doi.org/10.1007/s13347-020-00402-x>

Orazi, Davide C, and Greg Nyilasy, ‘Straight to the Heart of Your Target Audience: Personalized Advertising Systems Based on Wearable Technology and Heart-Rate Variability’, *Journal of Advertising Research*, 59.2 (2019), 137–41

Oshana, Marina, *Personal Autonomy in Society* (Ashgate: Ashgate Publishing, 2005)

Ouguergouz, Fatsah, ‘African Charter on Human and Peoples’ Rights’, *Max Planck*

Encyclopedia of Public International Law, 2010
Owens, John, and Alan Cribb, '"My Fitbit Thinks I Can Do Better!" Do Health Promoting Wearable Technologies Support Personal Autonomy?', *Philosophy & Technology*, 32.1 (2019), 23–38 <https://doi.org/10.1007/s13347-017-0266-2>
Oyebode, Oladapo, Chinenye Ndulue, Mona Alhasani, and Rita Orji, 'Persuasive Mobile Apps for Health and Wellness: A Comparative Systematic Review', 2020, pp. 163–81 <https://doi.org/10.1007/978-3-030-45712-9_13>
Oyibo, Kiemute, Rita Orji, and Julita Vassileva, 'The Influence of Culture in the Effect of Age and Gender on Social Influence in Persuasive Technology', in *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (New York, NY, USA: ACM, 2017), pp. 47–52 <https://doi.org/10.1145/3099023.3099071>
Pagallo, Ugo, 'The Collective Dimensions of Privacy in the Information Era', in *ANNUARIO DI DIRITTO COMPARATO E DI STUDI LEGISLATIVI* (Edizioni Scientifiche Italiane, 2020), pp. 115–38
Palmer, Adrian, and Stephanie Boissy, 'The Effects of Airline Price Presentations on Buyers' Choice', *Journal of Vacation Marketing*, 15.1 (2009), 39–52 <https://doi.org/10.1177/1356766708098170>
Parrilli, Davide Maria, and Rodrigo Hernandez-Ramirez, 'Re-Designing Dark Patterns to Improve Privacy', in *2020 IEEE International Symposium on Technology and Society (ISTAS)* (IEEE, 2020), pp. 253–54 <https://doi.org/10.1109/ISTAS50296.2020.9462197>
Patti, Salvatore, 'Contractual Autonomy and European Private Law', in *Rules and Principles in European Contract Law* (Intersentia), pp. 123–34 <https://doi.org/10.1017/9781780685434.007>
Paul, Richard, and Linda Elder, *The Miniature Gude to Understanding the Foundations of Ethical Reasoning*, 3rd edn (Critical Thinking, 2005)
Persson, Petra, 'Attention Manipulation and Information Overload', *Behavioural Public Policy*, 2.1 (2018), 78–106
Petty, Richard E., Duane T. Wegener, and Leandre R. Fabrigar, 'Attitudes and Attitude Change', *Annual Review of Psychology*, 48.1 (1997), 609–47 <https://doi.org/10.1146/annurev.psych.48.1.609>
Pew Research Center, *Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally*, 2019
Pitkin, Hanna Fenichel, 'Are Freedom and Liberty Twins?', *Political Theory*, 16.4 (1988), 523–52
Pluviano, Sara, Sergio Della Sala, and Caroline Watt, 'The Effects of Source Expertise and Trustworthiness on Recollection: The Case of Vaccine Misinformation', *Cognitive Processing*, 21.3 (2020), 321–30 <https://doi.org/10.1007/s10339-020-00974-8>
Poletti, F., N. V. Wheeler, M. N. Petrovich, N. Baddela, E. Numkam Fokoua, J. R. Hayes, and others, 'Towards High-Capacity Fibre-Optic Communications at the Speed of Light in Vacuum', *Nature Photonics*, 7.4 (2013), 279–84 <https://doi.org/10.1038/nphoton.2013.45>
Porat, Ariel, and Lior Jacob Strahilevitz, 'Personalizing Default Rules and Disclosure with Big Data', *Mich. L. Rev.*, 112 (2013), 1417
Pratama, Ahmad, and Moneer Al-Shaikh, 'Relation and Growth of Internet Penetration Rate with Human Development Level from 2000 to 2010', *Communications of the IBIMA*, 2012, 1–8 <https://doi.org/10.5171/2012.778309>
Preibusch, Sören, Dorothea Kübler, and Alastair R Beresford, 'Price versus Privacy: An Experiment into the Competitive Advantage of Collecting Less Personal Information', *Electronic Commerce Research*, 13.4 (2013), 423–55
Prince, Anya E R, and Daniel Schwarcz, 'Proxy Discrimination in the Age of Artificial

Intelligence and Big Data', *Iowa L. Rev.*, 105 (2019), 1257
Protection, Committee on the Internal Market and Consumer, and Committee on Civil Liberties Justice and Home-Affairs, 'Amendments - Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' <https://bit.ly/3Mboami>
Pushpanathan, Kalananthni, Marsyita Hanafi, Syamsiah Mashohor, and Wan Fazilah Fazlil Ilahi, 'Machine Learning in Medicinal Plants Recognition: A Review', *Artificial Intelligence Review*, 54.1 (2021), 305–27 <https://doi.org/10.1007/s10462-020-09847-0>
Qian, Jianwei, Xiang-Yang Li, Chunhong Zhang, Linlin Chen, Taeho Jung, and Junze Han, 'Social Network De-Anonymization and Privacy Inference with Knowledge Graph Model', *IEEE Transactions on Dependable and Secure Computing*, 16.4 (2019), 679–92 <https://doi.org/10.1109/TDSC.2017.2697854>
Qu, Xiujie, Tianbo Wei, Cheng Peng, and Peng Du, 'A Fast Face Recognition System Based on Deep Learning', in *2018 11th International Symposium on Computational Intelligence and Design (ISCID)* (IEEE, 2018), pp. 289–92 <https://doi.org/10.1109/ISCID.2018.00072>
Quinn, Philip, 'Divine Command Ethics', *The Cambridge Dictionary of Philosophy* (Cambridge University Press, 1999)
Radin, Margaret Jane, 'Boilerplate', in *Boilerplate* (Princeton University Press, 2012)
———, 'The Deformation of Contract in the Information Society', *Oxford Journal of Legal Studies*, 37.3 (2017), 505–33
Rama Murthy, Sudhir, and Monto Mani, 'Discerning Rejection of Technology', *SAGE Open*, 3.2 (2013), 215824401348524 <https://doi.org/10.1177/2158244013485248>
Ramcharan, Robin, and Bertrand Ramcharan, 'The Future of the UDHR: Building on the Past and the Present', in *Asia and the Drafting of the Universal Declaration of Human Rights* (Singapore: Springer Singapore, 2019), pp. 193–220 <https://doi.org/10.1007/978-981-13-2104-7_9>
Raper, Jonathan, Georg Gartner, Hassan Karimi, and Chris Rizos, 'A Critical Evaluation of Location Based Services and Their Potential', *Journal of Location Based Services*, 1.1 (2007), 5–45 <https://doi.org/10.1080/17489720701584069>
———, 'Applications of Location–Based Services: A Selected Review', *Journal of Location Based Services*, 1.2 (2007), 89–111 <https://doi.org/10.1080/17489720701862184>
Rappaport, T.S., 'The Wireless Revolution', *IEEE Communications Magazine*, 29.11 (1991), 52–71 <https://doi.org/10.1109/35.109666>
Rashkin, Hannah, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi, 'Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking', in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2931–37
Raub, McKenzie, 'Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices Comment', *Arkansas Law Review*, 71.2, 529–70 <https://heinonline.org/HOL/P?h=hein.journals/arklr71&i=549>
Raz, Joseph, 'Freedom and Autonomy', in *The Morality of Freedom* (Oxford University PressOxford, 1988), pp. 400–430 <https://doi.org/10.1093/0198248075.003.0015>
———, *The Authority of Law: Essays on Law and Morality* (Oxford: Oxford University Press, 1979)
Reath, Andrews, 'Ethical Autonomy', *Routledge Encyclopedia of Philosophy* (Taylor and Francis, 1998) <https://doi.org/10.4324/9780415249126-L007-1>
Redlawsk, David P., Andrew J. W. Civettini, and Karen M. Emmerson, 'The Affective Tipping Point: Do Motivated Reasoners Ever "Get It"?', *Political Psychology*, 31.4 (2010), 563–93 <https://doi.org/10.1111/j.1467-9221.2010.00772.x>

Redström, Johan, 'Persuasive Design: Fringes and Foundations', in *Proceedings of the First International Conference on Persuasive Technology for Human Well-Being*, PERSUASIVE'06 (Berlin, Heidelberg: Springer, 2006), pp. 112–122

Rees, C. F., 'Better Lie!', *Analysis*, 74.1 (2014), 59–64 <https://doi.org/10.1093/analys/ant104>

Reicher, S. D., R. Spears, and T. Postmes, 'A Social Identity Model of Deindividuation Phenomena', *European Review of Social Psychology*, 6.1 (1995), 161–98 <https://doi.org/10.1080/14792779443000049>

Reis-Dennis, Samuel, 'Understanding Autonomy: An Urgent Intervention', *Journal of Law and the Biosciences*, 7.1 (2020) <https://doi.org/10.1093/jlb/lsaa037>

Rendtorff, Jacob Dahl, 'Basic Ethical Principles in European Bioethics and Biolaw: Autonomy, Dignity, Integrity and Vulnerability--towards a Foundation of Bioethics and Biolaw', *Medicine, Health Care and Philosophy*, 5.3 (2002), 235–44

Resnick, Paul, and Hal R. Varian, 'Recommender Systems', *Communications of the ACM*, 40.3 (1997), 56–58 <https://doi.org/10.1145/245108.245121>

Reviglio, Urbano, and Rogers Alunge, '"I Am Datafied Because We Are Datafied": An Ubuntu Perspective on (Relational) Privacy', *Philosophy & Technology*, 33.4 (2020), 595–612 <https://doi.org/10.1007/s13347-020-00407-6>

Richards, Neil, and Daniel Solove, 'Privacy's Other Path: Recovering the Law of Confidentiality', *Georgetown Law Journal*, 96.1 (2007), 123–82

Rita, Orji, Mandryk Regan, and Vassileva Julita, 'Gender, Age, and Responsiveness to Cialdini's Persuasion Strategies', in *PERSUASIVE 2015*, ed. by MacTavish T. and Basapur S., 2015, pp. 147–59

Robbins, Scott, 'AI and the Path to Envelopment: Knowledge as a First Step towards the Responsible Regulation and Use of AI-Powered Machines', *AI & SOCIETY*, 35.2 (2020), 391–400 <https://doi.org/10.1007/s00146-019-00891-1>

Rogers, Everett M., 'The Digital Divide', *Convergence: The International Journal of Research into New Media Technologies*, 7.4 (2001), 96–111 <https://doi.org/10.1177/135485650100700406>

Rogers, Todd, Richard Zeckhauser, Francesca Gino, Michael I. Norton, and Maurice E. Schweitzer, 'Artful Paltering: The Risks and Rewards of Using Truthful Statements to Mislead Others.', *Journal of Personality and Social Psychology*, 112.3 (2017), 456–73 <https://doi.org/10.1037/pspi0000081>

Rosenberg, Richard, *The Social Impact of Computers*. (San Diego: Academic Press Inc, 1992)

Roßnagel, Alexander, 'Kein „Verbotsprinzip" Und Kein „Verbot Mit Erlaubnisvorbehalt" Im Datenschutzrecht', *Neue Juristische Wochenschrift: NJW*, 1 (2019), 1

Roßnagel, Alexander, and Christian Geminn, '"Privatheit" Und "Privatsphäre" Aus Der Perspektive Des Rechts - Ein Überblick', *Juristenzeitung*, 2015, 703

Rotman, David, 'We're Not Prepared for the End of Moore's Law', *MIT Technology Review*, 2020

Rouvroy, Antoinette, and Yves Poullet, 'The Right to Informational Self-Determination and the Value of Self-Development: Reassessing the Importance of Privacy for Democracy', in *Reinventing Data Protection?*, ed. by Serge Gutwirth, Yves Poullet, Paul De Hert, Cécile de Terwangne, and Sjaak Nouwt, 1st edn (Dordrecht: Springer, 2009), pp. 45–76

Rughiniş, Cosima, Răzvan Rughiniş, and Ştefania Matei, 'A Touching App Voice Thinking about Ethics of Persuasive Technology through an Analysis of Mobile Smoking-Cessation Apps', *Ethics and Information Technology*, 17.4 (2015), 295–309 <https://doi.org/10.1007/s10676-016-9385-1>

Ruijten, Peter A. M., Cees J. H. Midden, and Jaap Ham, 'Unconscious Persuasion Needs Goal-Striving', in *Proceedings of the 6th International Conference on Persuasive Technology*

Persuasive Technology and Design: Enhancing Sustainability and Health - PERSUASIVE '11 (New York, New York, USA: ACM Press, 2011), pp. 1–6 <https://doi.org/10.1145/2467803.2467807>

Russel, Stuart, and Peter Norvig, *Artificial Intelligence: A Moden Approach* (Pearson, 2020)

Samsir, S, S Suparno, and M Giatman, 'Predicting the Loan Risk towards New Customer Applying Data Mining Using Nearest Neighbor Algorithm', in *IOP Conference Series: Materials Science and Engineering*, 2020, DCCCXXX, 32004

Samuel, A. L., 'Some Studies in Machine Learning Using the Game of Checkers', *IBM Journal of Research and Development*, 3.3 (1959), 210–29 <https://doi.org/10.1147/rd.33.0210>

Sandler, Ronald L., 'Introduction: Technology and Ethics', in *Ethics and Emerging Technologies* (London: Palgrave Macmillan UK, 2014), pp. 1–23 <https://doi.org/10.1057/9781137349088_1>

Satzger, Helmut, 'Kausalität Und Gremienentscheidungen', *JURA – Juristische Ausbildung*, 36.2 (2014) <https://doi.org/10.1515/jura-2014-0028>

Saul, J., 'Just Go Ahead and Lie', *Analysis*, 72.1 (2012), 3–9 <https://doi.org/10.1093/analys/anr133>

Sauter, Wolf, 'Proportionality in EU Law: A Balancing Act?', *Cambridge Yearbook of European Legal Studies*, 15 (2013), 439–466 <https://doi.org/10.5235/152888713809813611>

Schabas, William, *The European Convention on Human Rights: A Commentary*, 1st edn (Oxford University Press, 2015) <https://doi.org/10.1093/law/9780199594061.001.0001>

Schaewitz, Leonie, and Nicole C. Krämer, 'Combating Disinformation: Effects of Timing and Correction Format on Factual Knowledge and Personal Beliefs', 2020, pp. 233–45 <https://doi.org/10.1007/978-3-030-61841-4_16>

Scheibehenne, Benjamin, Rainer Greifeneder, and Peter M. Todd, 'Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload', *Journal of Consumer Research*, 37.3 (2010), 409–25 <https://doi.org/10.1086/651235>

Schermer, Bart W., 'The Limits of Privacy in Automated Profiling and Data Mining', *Computer Law & Security Review*, 27.1 (2011), 45–52 <https://doi.org/10.1016/j.clsr.2010.11.009>

Schiller, Jochen, and Agnès Voisard, *Location-Based Services* (Elsevier, 2004) <https://doi.org/10.1016/B978-1-55860-929-7.X5000-6>

Schimmack, Ulrich, Shigehiro Oishi, and Ed Diener, 'Individualism: A Valid and Important Dimension of Cultural Differences between Nations', *Personality and Social Psychology Review*, 9.1 (2005), 17–31

Schlosser, Markus, 'Agency', ed. by Edward N Zalta, *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University, 2019)

Schlosser, Markus E., 'Agency, Ownership, and the Standard Theory', in *New Waves in Philosophy of Action* (London: Palgrave Macmillan UK, 2011), pp. 13–31 <https://doi.org/10.1057/9780230304253_2>

Schmidt, Albrecht, Michael Beigl, and Hans-W Gellersen, 'There Is More to Context than Location', *Computers & Graphics*, 23.6 (1999), 893–901 <https://doi.org/10.1016/S0097-8493(99)00120-X>

Schmidt, Andreas, 'Getting Real on Rationality—Behavioral Science, Nudging, and Public Policy', *Ethics*, 129.4 (2019), 511–43

Schneble, Christophe Olivier, Bernice Simone Elger, and David Shaw, 'The Cambridge Analytica Affair and Internet-Mediated Research', *EMBO Reports*, 19.8 (2018), e46579

Schwartz, Andrew W., 'Autonomy and Oppression: Beyond the Substantive and Content-Neutral Debate', *The Journal of Value Inquiry*, 39.3–4 (2007), 443–57

<https://doi.org/10.1007/s10790-006-3124-5>
Schwartz, Paul, and Daniel Solove, 'Reconciling Personal Information in the United States and European Union', *California Law Review*, 102.4 (2014), 877–916
Sedano, Todd, 'Code Readability Testing, an Empirical Study', in *2016 IEEE 29th International Conference on Software Engineering Education and Training (CSEET)* (IEEE, 2016), pp. 111–17 <https://doi.org/10.1109/CSEET.2016.36>
Selbst, Andrew D, and Solon Barocas, 'The Intuitive Appeal of Explainable Machines', *Fordham L. Rev.*, 87 (2018), 1085
Shao, Yingxia, Jialin Liu, Shuyang Shi, Yuemei Zhang, and Bin Cui, 'Fast De-Anonymization of Social Networks with Structural Information', *Data Science and Engineering*, 4.1 (2019), 76–92 <https://doi.org/10.1007/s41019-019-0086-8>
Shapiro, Ian, 'On Non-Domination', *University of Toronto Law Journal*, 62.3 (2012), 293–336
Shaw, Malcolm N., *International Law*, 6th edn, 2008
Shin, Donghee, and Yong Jin Park, 'Role of Fairness, Accountability, and Transparency in Algorithmic Affordance', *Computers in Human Behavior*, 98 (2019), 277–84 <https://doi.org/10.1016/j.chb.2019.04.019>
Shugan, Steven M., 'The Cost of Thinking', *Journal of Consumer Research*, 7.2 (1980), 99 <https://doi.org/10.1086/208799>
Shweder, Richard A, Nancy C Much, Manamohan Mahapatra, and Lawrence Park, 'The" Big Three" of Morality (Autonomy, Community, Divinity) and the" Big Three" Explanations of Suffering.', 1997
Singer, Marcus, 'Moral Epistemology', *The Cambridge Dictionary of Philosophy* (Cambridge University Press)
Sloan, Robert H., and Richard Warner, 'When Is an Algorithm Transparent? Predictive Analytics, Privacy, and Public Policy', *IEEE Security & Privacy*, 16.3 (2018), 18–25 <https://doi.org/10.1109/MSP.2018.2701166>
Smids, Jilles, 'The Voluntariness of Persuasive Technology', 2012, pp. 123–32 <https://doi.org/10.1007/978-3-642-31037-9_11>
Smith, Cedric M., 'Origin and Uses of Primum Non Nocere -Above All, Do No Harm!', *The Journal of Clinical Pharmacology*, 45.4 (2005), 371–77 <https://doi.org/10.1177/0091270004273680>
Smith, H Jeff, Tamara Dinev, and Heng Xu, 'Information Privacy Research: An Interdisciplinary Review', *MIS Quarterly*, 35.4 (2011), 989–1015 <https://doi.org/10.2307/41409970>
Smith, Kerri, and others, 'Neuroscience vs Philosophy: Taking Aim at Free Will', *Nature*, 477.7362 (2011), 23–25
Sneddon, Andrew, 'Advertising and Deep Autonomy', *Journal of Business Ethics*, 33.1 (2001), 15–28
———, *Autonomy*, 1st edn (London/New York: Bloomsbury Academic, 2013)
Soe, Than Htut, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovik, 'Circumvention by Design - Dark Patterns in Cookie Consent for Online News Outlets', in *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (New York, NY, USA: ACM, 2020), pp. 1–12 <https://doi.org/10.1145/3419249.3420132>
Solove, Daniel, 'A Taxonomy of Privacy', *University of Pennsylvania Law Review*, 154.3, 477–564 <https://heinonline.org/HOL/P?h=hein.journals/pnlr154&i=493>
———, 'Conceptualizing Privacy', *California Law Review*, 90.4 (2002), 1087–1155
Spahn, Andreas, 'And Lead Us (Not) into Persuasion…? Persuasive Technology and the Ethics of Communication', *Science and Engineering Ethics*, 18.4 (2012), 633–50 <https://doi.org/10.1007/s11948-011-9278-y>

Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, *Seventy-Third Session Item 74 (b) of the Provisional Agenda** Promotion and Protection of Human Rights: Human Rights Questions, Including Alternative Approaches for Improving the Effective Enjoyment of Human Rights and Fundamental Freedoms*, 2018

Specker Sullivan, Laura, and Peter Reiner, 'Digital Wellness and Persuasive Technologies', *Philosophy & Technology*, 34.3 (2021), 413–24 <https://doi.org/10.1007/s13347-019-00376-5>

Srivatsa, Mudhakar, and Mike Hicks, 'Deanonymizing Mobility Traces', in *Proceedings of the 2012 ACM Conference on Computer and Communications Security - CCS '12* (New York, New York, USA: ACM Press, 2012), p. 628 <https://doi.org/10.1145/2382196.2382262>

Steinmüller, Wilhelm, Bernd Lutterbeck, C Mallmann, U Harbort, G Kolb, and J Schneider, *Grundfragen Des Datenschutz - Gutachten Im Auftrag Des Bundesministeriums Des Innern (Drucksache VI/3826)*, 1971

Stoljar, Natalie, 'Autonomy and the Feminist Intuition', in *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*, ed. by Catriona MacKenzie and Natalie Stoljar. (New York: Oxford University Press, 2000), pp. 94–111

Straub, Detmar, Mark Keil, and Walter Brenner, 'Testing the Technology Acceptance Model across Cultures: A Three Country Study', *Information \& Management*, 33.1 (1997), 1–11

Strauss, David A., 'Persuasion, Autonomy, and Freedom of Expression', *Columbia Law Review*, 91.2 (1991), 334 <https://doi.org/10.2307/1122761>

Stroud, Natalie Jomini, 'Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure', *Political Behavior*, 30.3 (2008), 341–66 <https://doi.org/10.1007/s11109-007-9050-9>

Sunstein, Cass R, 'Nudging: A Very Short Guide', *Journal of Consumer Policy*, 37.4 (2014), 583–88

———, 'The Ethics of Nudging', *Yale Journal on Regulation*, 32.2 (2015), 413–50

Sunstein, Cass R, and Richard H Thaler, 'Libertarian Paternalism Is Not an Oxymoron', *The University of Chicago Law Review*, 2003, 1159–1202

Surden, Harry, 'Structural Rights in Privac', *SMU Law Review*, 60 (2007), 1605–29

Susser, Daniel, Beate Roessler, and Helen Nissenbaum, 'Technology, Autonomy, and Manipulation', *Internet Policy Review*, 8.2 (2019)

Taddeo, Mariarosaria, and Luciano Floridi, 'How AI Can Be a Force for Good', *Science*, 361.6404 (2018), 751–52 <https://doi.org/10.1126/science.aat5991>

Taib, Mohd, 'Loophole Allows Bias in Displays on Computer Reservations Systems', *Aviation Week & Space Technology*, 132.7 (1990), 137

Talbert, Matthew, 'Moral Responsibility', *The Stanford Encyclopedia of Philosophy*, 2022 <https://plato.stanford.edu/archives/fall2022/entries/moral-responsibility/>

Tandoc, Edson C., Zheng Wei Lim, and Richard Ling, 'Defining "Fake News"', *Digital Journalism*, 6.2 (2018), 137–53 <https://doi.org/10.1080/21670811.2017.1360143>

Tashea, Jason, 'CALCULATING CRIME: Attorneys Are Challenging the Use of Algorithms to Help Determine Bail, Sentencing and Parole Decisions', *ABA Journal*, 103.3 (2017), 54–59

Tatara, E., and A. Cinar, 'Interpreting ECG Data by Integrating Statistical and Artificial Intelligence Tools', *IEEE Engineering in Medicine and Biology Magazine*, 21.1 (2002), 36–41 <https://doi.org/10.1109/51.993192>

Tavani, H., 'Informational Privacy: Concepts, Theories and Controversies', in *Handbook of Information and Computer Ethics*, ed. by K.E. Himma and H.T. Tavani (Haboken: John Wiley, 2008), pp. 131–64

Taylor, James Stacey, ‘Introduction’, in *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy* (Cambridge: Cambridge University Press, 2005), pp. 1–30
———, *Practical Autonomy and Bioethics* (Routledge, 2010)
Taylor, Linnet, ‘From Zero to Hero: How Zero-Rating Became a Debate about Human Rights’, *IEEE Internet Computing*, 20.4 (2016), 79–83 <https://doi.org/10.1109/MIC.2016.88>
Tenbrunsel, Ann E, and David M Messick, ‘Ethical Fading: The Role of Self-Deception in Unethical Behavior’, *Social Justice Research*, 17.2 (2004), 223–36
Tene, Omer, and Jules Polonetsky, ‘Big Data for All: Privacy and User Control in the Age of Analytics’, *Northwestern Journal of Technology and Intellectual Property*, 11.5 (2013), 239–72
Terwangne, Cécile de, ‘The Right to Be Forgotten and Informational Autonomy in the Digital Environment’, in *The Ethics of Memory in a Digital Age* (Springer, 2014), pp. 82–101
Thaler, Richard, Cass Sunstein, and John Balz, ‘Choice Architecture’, in *The Behavioral Foundations of Public Policy* (Princeton University Press, 2013), pp. 428–39 <https://doi.org/10.1515/9781400845347-029>
Thomas, Rachel, and David Uminsky, ‘The Problem with Metrics Is a Fundamental Problem for Ai’, *ArXiv Preprint ArXiv:2002.08512*, 2020
Thürer, Daniel, ‘Soft Law’, *Max Planck Encyclopedia of Public International Law*, 2009
Thurman, Neil, and Steve Schifferes, ‘The Future of Personalization at News Websites’, *Journalism Studies*, 13.5–6 (2012), 775–90 <https://doi.org/10.1080/1461670X.2012.664341>
Timothy O’Connor, Christopher Franklin, ‘Free Will’, *The Stanford Encyclopedia of Philosophy2*, 2022 <https://plato.stanford.edu/archives/sum2022/entries/freewill/>
Tintarev, Nava, and Judith Masthoff, ‘A Survey of Explanations in Recommender Systems’, in *2007 IEEE 23rd International Conference on Data Engineering Workshop*, 2007, pp. 801–10
Tomuschat, Christian, ‘International Covenant on Civil and Political Rights.’, *Max Planck Encyclopedia of Public International Law*, 2019
Trappey, Charles, ‘A Meta-Analysis of Consumer Choice and Subliminal Advertising’, *Psychology \& Marketing*, 13.5 (1996), 517–30
Treves, Tullio, ‘Customary International Law’, *Max Planck Encyclopedia of Public International Law*, 2006
TSAI, GEORGE, ‘Rational Persuasion as Paternalism’, *Philosophy & Public Affairs*, 42.1 (2014), 78–112 <https://doi.org/10.1111/papa.12026>
Tsai, Ming-Chun, Shu-Ping Lin, Ching-Chan Cheng, and Yen-Ping Lin, ‘The Consumer Loan Default Predicting Model – An Application of DEA–DA and Neural Network’, *Expert Systems with Applications*, 36.9 (2009), 11682–90 <https://doi.org/10.1016/j.eswa.2009.03.009>
Tsamados, Andreas, Nikita Aggarwal, Josh Cowls, Jessica Morley, Huw Roberts, Mariarosaria Taddeo, and others, ‘The Ethics of Algorithms: Key Problems and Solutions’, *AI & SOCIETY*, 2021 <https://doi.org/10.1007/s00146-021-01154-8>
Turing, A M, ‘Computing Machinery and Intelligence’, *Mind*, 59.236 (1950), 433–60
Tversky, A, and D Kahneman, ‘Judgement under Uncertainty: Heuristics and Biases’, *Science*, 185 (1974), 1124–31
———, ‘The Framing of Decisions and the Psychology of Choice.’, *Science*, 211 (1981), 453–458
Tzanou, M., ‘Data Protection as a Fundamental Right next to Privacy? “Reconstructing” a Not so New Right’, *International Data Privacy Law*, 3.2 (2013), 88–99 <https://doi.org/10.1093/idpl/ipt004>

Valenzuela, Sebastián, Daniel Halpern, James E. Katz, and Juan Pablo Miranda, 'The Paradox of Participation Versus Misinformation: Social Media, Political Engagement, and the Spread of Misinformation', *Digital Journalism*, 7.6 (2019), 802–23 <https://doi.org/10.1080/21670811.2019.1623701>

Vance, Anthony, Christophe Elie-Dit-Cosaque, and Detmar W. Straub, 'Examining Trust in Information Technology Artifacts: The Effects of System Quality and Culture', *Journal of Management Information Systems*, 24.4 (2008), 73–100 <https://doi.org/10.2753/MIS0742-1222240403>

Varghese, Sandra, and S Remya, 'Dog Breed Classification Using CNN', in *Security Issues and Privacy Concerns in Industry 4.0 Applications* (Wiley, 2021), pp. 195–205 <https://doi.org/10.1002/9781119776529.ch10>

Vedder, Anton, 'KDD: The Challenge to Individualism', *Ethics and Information Technology1*, 1 (1999), 275–81 <https://doi.org/https://doi.org/10.1023/A:1010016102284>

Vedder, Anton, and Laurens Naudts, 'Accountability for the Use of Algorithms in a Big Data Environment', *International Review of Law, Computers & Technology*, 31.2 (2017), 206–24 <https://doi.org/10.1080/13600869.2017.1298547>

Verbeek, Peter-Paul, 'Ambient Intelligence and Persuasive Technology: The Blurring Boundaries Between Human and Technology', *NanoEthics*, 3.3 (2009), 231–42 <https://doi.org/10.1007/s11569-009-0077-8>

———, *Moralizing Technology: Understanding and Designing the Morality of Things* (University of Chicago press, 2011)

Villiger, Mark E., *Commentary on the 1969 Vienna Convention on the Law of Treaties*, *Commentary on the 1969 Vienna Convention on the Law of Treaties*, 2008 <https://doi.org/10.1163/ej.9789004168046.i-1058>

Vosoughi, Soroush, Deb Roy, and Sinan Aral, 'The Spread of True and False News Online', *Science*, 359.6380 (2018), 1146–51 <https://doi.org/10.1126/science.aap9559>

Vraga, Emily K., and Leticia Bode, 'Defining Misinformation and Understanding Its Bounded Nature: Using Expertise and Evidence for Describing Misinformation', *Political Communication*, 37.1 (2020), 136–44 <https://doi.org/10.1080/10584609.2020.1716500>

de Vries, Katja, 'Identity, Profiling Algorithms and a World of Ambient Intelligence', *Ethics and Information Technology*, 12.1 (2010), 71–85 <https://doi.org/10.1007/s10676-009-9215-9>

Wagner, Markus, 'The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems', *Vand. J. Transnat'l L.*, 47 (2014), 1371

Waldman, Ari Ezra, 'Cognitive Biases, Dark Patterns, and the "Privacy Paradox"', *Current Opinion in Psychology*, 31 (2020), 105–9 <https://doi.org/10.1016/j.copsyc.2019.08.025>

Wall, Emily, John Stasko, and Alex Endert, 'Toward a Design Space for Mitigating Cognitive Bias in Vis', in *2019 IEEE Visualization Conference (VIS)* (IEEE, 2019), pp. 111–15 <https://doi.org/10.1109/VISUAL.2019.8933611>

Waltz, Susan, 'Reclaiming and Rebuilding the History of the Universal Declaration of Human Rights', *Third World Quarterly*, 23.3 (2002), 437–48 <https://doi.org/10.1080/01436590220138378>

———, 'Universalizing Human Rights: The Role of Small States in the Construction of the Universal Declaration of Human Rights ', *Human Rights Quarterly*, 23.1 (2001), 44–72 <https://heinonline.org/HOL/P?h=hein.journals/hurq23&i=57>

Wang, Pei, 'On Defining Artificial Intelligence', *Journal of Artificial General Intelligence*, 10.2 (2019), 1–37 <https://doi.org/10.2478/jagi-2019-0002>

Warner, Richard, and Robert H Sloan, 'Making Artificial Intelligence Transparent: Fairness

and the Problem of Proxy Variables', *Criminal Justice Ethics*, 40.1 (2021), 23–39
Warren, Samuel, and Louis Brandeis, 'The Right to Privacy', *Harvard Law Review*, 4.5 (1890), 193–220
Watson, Gary, 'Two Faces of Responsibility', *Philosophical Topics*, 24.2 (1996), 227–48
Weinhofen, Henry, 'The Purpose of Punishment ', *Tennessee Law Review*, 7.3, 145–76 <https://heinonline.org/HOL/P?h=hein.journals/tenn7&i=151>
Weinmann, Markus, Christoph Schneider, and Jan vom Brocke, 'Digital Nudging', *Business & Information Systems Engineering*, 58.6 (2016), 433–36 <https://doi.org/10.1007/s12599-016-0453-1>
Weiser, Mark, 'The Computer for the 21 St Century', *ACM SIGMOBILE Mobile Computing and Communications Review*, 3.3 (1999), 3–11 <https://doi.org/10.1145/329124.329126>
Wells, Lindsay, and Tomasz Bednarz, 'Explainable Ai and Reinforcement Learning—a Systematic Review of Current Approaches and Trends', *Frontiers in Artificial Intelligence*, 4 (2021), 48
West, S, 'Increasing the Attractiveness of College Cafetaria Food: A Reactance Theory Perspective', *J Appl. Psychol*, 60 (1975), 656–58
Westin, Alan F, 'Privacy And Freedom', *Washington and Lee Law Review*, 1968
Whitten-Woodring, Jenifer, Mona S. Kleinberg, Ardeth Thawnghmung, and Myat The Thitsar, 'Poison If You Don't Know How to Use It: Facebook, Democracy, and Human Rights in Myanmar', *The International Journal of Press/Politics*, 25.3 (2020), 407–25 <https://doi.org/10.1177/1940161220919666>
Wicker, Stephen B., 'The Loss of Location Privacy in the Cellular Age', *Communications of the ACM*, 55.8 (2012), 60–68 <https://doi.org/10.1145/2240236.2240255>
Williams, Bernard, 'From Freedom to Liberty: The Construction of a Political Value', *Philosophy \& Public Affairs*, 30.1 (2001), 3–26
Wolf, Susan, 'Freedom within Reason', in *Personal Autonomy* (Cambridge University Press, 2005), pp. 258–74 <https://doi.org/10.1017/CBO9780511614194.012>
Wolfram, Catherine, Orie Shelef, and Paul Gertler, 'How Will Energy Demand Develop in the Developing World?', *Journal of Economic Perspectives*, 26.1 (2012), 119–38 <https://doi.org/10.1257/jep.26.1.119>
Won, An Na, and Won Hwa Hong, 'A Survey on Ownership of Home Appliances and Electric Energy Consumption Status According to the Number of Household Member', *Applied Mechanics and Materials*, 672–674 (2014), 2165–68 <https://doi.org/10.4028/www.scientific.net/AMM.672-674.2165>
Wong, Pak-Hang, 'Cultural Differences as Excuses? Human Rights and Cultural Values in Global Ethics and Governance of AI', *Philosophy & Technology*, 33.4 (2020), 705–15 <https://doi.org/10.1007/s13347-020-00413-8>
Wooldridge, Michael, and Nicholas R Jennings, 'Intelligent Agents: Theory and Practice', *The Knowledge Engineering Review*, 10.2 (1995), 115–52
Worden, R.P., 'A Speed Limit for Evolution', *Journal of Theoretical Biology*, 176.1 (1995), 137–52 <https://doi.org/10.1006/jtbi.1995.0183>
World Economic Forum, *Global Information Technology Report*, 2016
Wu, Liang, Fred Morstatter, Kathleen M. Carley, and Huan Liu, 'Misinformation in Social Media', *ACM SIGKDD Explorations Newsletter*, 21.2 (2019), 80–90 <https://doi.org/10.1145/3373464.3373475>
Yntema, Hessel E, 'Autonomy in Choice of Law ', *American Journal of Comparative Law*, 1.4 (1952), 341–58 <https://heinonline.org/HOL/P?h=hein.journals/amcomp1&i=349>
Young, Robert, 'The Value of Autonomy', *The Philosophical Quarterly (1950-)*, 32.126 (1982), 35–44
Young, Scott W H, Angela Tate, Doralyn Rossmann, and Hansen Mary Anne, 'The Social

Media Toll Road: The Promise and Peril of Facebook Advertising', *College \& Research Libraries News*, 75.8 (2014), 427–34

Yu, Qian, Yongxin Yang, Feng Liu, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales, 'Sketch-a-Net: A Deep Neural Network That Beats Humans', *International Journal of Computer Vision*, 122.3 (2017), 411–25

Yu, Shi, Chantal Levesque-Bristol, and Yukiko Maeda, 'General Need for Autonomy and Subjective Well-Being: A Meta-Analysis of Studies in the US and East Asia', *Journal of Happiness Studies*, 19.6 (2018), 1863–82 <https://doi.org/10.1007/s10902-017-9898-2>

Yurur, Ozgur, Chi Harold Liu, Zhengguo Sheng, Victor C. M. Leung, Wilfrido Moreno, and Kin K. Leung, 'Context-Awareness for Mobile Sensing: A Survey and Future Directions', *IEEE Communications Surveys & Tutorials*, 18.1 (2016), 68–93 <https://doi.org/10.1109/COMST.2014.2381246>

Zandy, Janet, 'Universal Declaration of Human Rights', *Max Planck Encyclopedia of Public International Law*, 2008

Zhang, Xiaolu, Luyang Yang, and Richard Sinnott, 'A Mobile Application for Cat Detection and Breed Recognition Based on Deep Learning', in *2019 IEEE 1st International Workshop on Artificial Intelligence for Mobile (AI4Mobile)* (IEEE, 2019), pp. 7–12 <https://doi.org/10.1109/AI4Mobile.2019.8672684>

Ziemele, Ineta, 'International Protection of the Right to Privacy', *Max Planck Encyclopedia of Public International Law*, 2009

Zuboff, Shoshana, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization', *Journal of Information Technology*, 30.1 (2015), 75–89 <https://doi.org/10.1057/jit.2015.5>

———, *The Age of Surveillance Capitalism*, 2019 <https://doi.org/10.1017/CBO9781107415324.004>

Zywicki, Todd J, 'Do Americans Really Save Too Little and Should We Nudge Them to Save More: The Ethics of Nudging Retirement Savings Symposium: The Ethics of Nudging - Evaluating Libertarian Paternalism: Implications, Extensions, and Applications', *Georgetown Journal of Law & Public Policy*, 14.Special 2016 (2016), 877–920

Annex 4.7.2

The Universal Declaration of Human Rights

The Universal Declaration of Human Rights (UDHR) was adopted on the 10th of December 1948 by the United Nations General Assembly. It is part of the international bill of rights, together with the Universal Declaration of Human Rights and the International Covenant on Economic Social and Cultural Rights. The declaration is legally non-binding.[730]

The International Covenant on Civil and Political Rights

The International Covenant on Civil and Political Rights (ICCPR) is a human rights regime entered into force on the 23rd of March 1976. More than 170 states have since ratified the treaty. It is part of the international bill of rights, together with the International Covenant on Civil and Political Rights and the International Covenant on Economic Social and Cultural Rights. It is enforced by the Human Rights Committee (HRC). The covenant is not effective directly, but obliges states to implement its contents.[731] The ICCPR is legally binding.

The European Convention on Human Rights

The European Convention on Human Rights, also named the Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), is a human rights regime entered into force on the 3rd of September 1953, applying to all members of the Council of Europe. It draws upon and was inspired by the Universal Declaration of Human Rights.[732] It is adjudicated and enforced by the European Court of Human Rights, sharing the same acronym. This allows unusual precise delimitations of its protected values, compared to other international regimes which will be reflected in a comparatively more expansive analysis hereafter. The ECHR is legally binding.

The European Convention on Human Rights is to be interpreted in a way that its rights are rendered practical and effective, not theoretical and illusory; as a "living instrument" which must be interpreted in the light of present day conditions and under the principle of effectiveness.[733] Accordingly, the most protective interpretation of the Convention is meant to be required.[734] This is considered as an evolutionary or evolutive approach of interpretation.[735] Additionally, Article 17 of the ECHR references the interpretation of the convention.

The Charter of Fundamental Rights of the European Union

The Charter of Fundamental Rights of the European Union (CFR) is a regional human rights regime that applies to most states of the European Union as well as to the European Union itself in the rank of the primary European Union treaties via a reference in Article 6 of the Treaty on European Union (TEU). The CFR entered into force on the 1st of December 2009. It

---

[A]DJanet Zandy, 'Universal Declaration of Human Rights', *Max Planck Encyclopedia of Public International Law*, 2008.

[731] Christian Tomuschat, 'International Covenant on Civil and Political Rights.', *Max Planck Encyclopedia of Public International Law*, 2019.

[732] William Schabas, *The European Convention on Human Rights: A Commentary*, 1st edn (Oxford University Press, 2015) <https://doi.org/10.1093/law/9780199594061.001.0001> p. 47f.

[733] Cf. ECHR (1978) Tyrer v. the United Kingdom §31, (2002) Christine Goodwin v. the United Kingdom [GC] §75, (2009) Sergey Zolotukhin v. Russia [GC] §80

[734] Cf. Schabas p.49f.

[735] Passim Eirik Bjorge, *The Evolutionary Interpretation of Treaties*, 2014 <https://doi.org/10.1093/acprof:oso/9780198716143.001.0001>.

does not apply to purely domestic situations, where the fundamental rights as instated by respective domestic laws are applicable. The CFR is legally binding.

The African Charter on Human and People's Rights

The African Charter on Human and People's Rights (ACHPR), also named the Banjul Charter, is a regional human rights regime that applies to all African states with the exception of Morocco. The ACHPR entered into force on the 21st of October 1986. It serves as a regional complementary instrument to the international bill of right; as a consequence it exhibits substantial differences to the other regimes showcased here.[736] Notably, the right to privacy is absent in the ACHPR. The ACHPR established the African Commission on Human and Peoples' Rights, and via an additional protocol the African Court on Human and Peoples' Rights, tasked with its adjudication and enforcement. The ACHPR is legally binding.[737]

The American Convention on Human Rights

The American Convention on Human Rights (ACHR) is regional human rights regime that applies to a subset of states of the Organization of American States (OAS), specifically Central and South-American states. The ACHR entered into force on the 18th of July 1978. The Inter-American Commission on Human Rights and the Inter-American Court of Human Rights are tasked with its adjudication and enforcement.[738] The ACHR is legally binding.

[736] In particular, the Banjul Charter provides for an extensive description of duties of individuals, as well as a "right of solidarity", while omitting rights established by the other regimes.
[737] Fatsah Ouguergouz, 'African Charter on Human and Peoples' Rights', *Max Planck Encyclopedia of Public International Law*, 2010.
[738] Gerald L Neuman, 'American Convention on Human Rights', *Max Planck Encyclopedia of Public International Law*, 2010.

Annex 4.7.3.2

Article 8 of the European Convention on Human Rights

Within the following Section, the term "court" is meant to denominate the European Court of Human Rights, and the term "Article" is meant to denominate an Article of the European Convention on Human Rights if not specified otherwise.

Normative Content

Article 8 of the ECHR reads as follows:

> *1. Everyone has the right to respect for his private and family life, his home and his correspondence.*
>
> *2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.*[739]

Article 8 of the ECHR is remarkably similar to Article 12 of the Universal Declaration of Human Rights.

From the text of the document, we can derive four distinct areas protected by Article 8, namely (1) private life, (2) family life, (3) home and (3) correspondence. While all of these can potentially be affected by agents of the IoE, this elaboration will focus on the protection of private life for reasons that will be self-evident.

Nature of Obligation

Principally, Article 8 of the ECHR stipulates a negative obligation for states not to arbitrarily interfere.[740] Nonetheless, positive obligations, that means obligations to act and not merely abstain from harmful actions, for states can arise to ensure protection of the rights granted.[741] While the language of paragraph 2 explicitly addresses public authorities, i.e. entities presumably under some sort of the control of a state, Article 8 can also affect relationship between two non-governmental entities interacting with each other. This can also extend to ensure compliance with these rights between private individuals, as this is necessary to protect the respect for private live mandated by Article 8.[742] The choice of action undertaken to fulfil this obligation is understood to be at the discretion of the state, but must meet a certain threshold of effectiveness of deterrence; for example grave acts with fundamental values and essential aspects of private life at risk will require sufficiently efficient criminal-law provisions.[743] Conversely, protection from acts between individuals which may violate psychological integrity may only require civil-law remedies to be considered as sufficient protection, and does not mandate deterrence under a criminal law regime.[744]

---

[739] Cf. generally for the following section ECHR (2019), Guide on Article 8 of the Convention – Right to respect for private and family life, available at https://www.echr.coe.int/Documents/Guide_Art_8_ENG.pdf.
[740] Cf. ECHR (1994), Kroon and Others v. the Netherlands, § 31
[741] Cf. ECHR (2017) , Bărbulescu v. Romania [GC], §§ 108
[742] Cf. Id. §111, ECHR Evans v. the United Kingdom [GC] § 75
[743] Cf. ECHR (2008), Case of K.U. v Finland §§43f.
[744] Cf. ECHR, X and Y v. the Netherlands, §§24ff

To trigger positive obligations of a state under Article 8 meet a heightened threshold, e.g. their underlying interest must reach a certain level of importance so that they concern fundamental values or essential aspects of private live, they highlight a disparity between the law and the social reality; and might be reviewed against the potential impact of the positive obligation on the State and whether the obligation is narrow and precise enough as opposed to being broad and indeterminate.[745] Nevertheless, the protected values under Article 8 remain unchanged, with these limits to positive obligations only affecting its enforcement by way of coercing a State into adopting certain positive steps ensuring protection between individuals.

Protective Scope

As described above an individual's private life, family life, home and correspondence are protected. While all of these can potentially be affected by agents of the IoE, this elaboration will focus on the protection of private life and correspondence as case law deals with exposure to new technologies primarily in these areas. The "living instrument"-approach (see Section 0) has led to a quite extensive interpretation of the rules codified in the ECHR, as evidenced by the wide understanding of the term "private life" and its application to situations where new technologies have emerged as a potential conflict factor, as will be explained below.

Private Life: General Scope

The ECHR has found that the terminus "private life" is wide-ranging, not accessible to exhaustive definition, and therefore generally also covers the psychological integrity of a person.[746] protection may extend to a person's inner life as well, eg, philosophical, religious or moral beliefs, emotional life.[747] mental health must also be regarded as a crucial part of private life associated with the aspect of moral integrity. It includes the right for each individual to approach others in order to establish and develop relationships with them and the outside world, so including potentially also professional and commercial activities.[748] Additionally, personal development falls within the protected scope.[749] It does not include however activities which are of essentially public nature, but public context alone is not harmful *per se*.[750] Indeed, interactions between people even when conducted in a public context may be covered.[751] This also holds when it comes to the protection of personal data.[752]

Issues falling under the umbrella term private life as understood in Article 8 ECHR can generally be divided in three categories: (1) physical, psychological and moral integrity, (2) privacy, and (3) identity and autonomy, of which the last two categories are of interest here.[753]

Privacy and Personal Data

First and foremost, the right of privacy encompasses the right to be "left alone".[754] Very explicitly, the ECHR found that "the protection of personal data is of fundamental importance

[745] Cf. ECHR (2014) Hämäläinen v. Finland [GC], § 66
[746] Cf. ECHR (2019), Nicolae Virgiliu Tănase v. Romania [GC], § 128
[747] Commentary of the Charter of Fundamental Rights of the European Union 7
[748] Cf. ECHR Botta v. Italy, §32, Fernandez Martinez v. Spain [GC §110, Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland [GC], § 130)]
[749] Cf. ECHR Bensaid v. the United Kingdom, § 47.
[750] Cf. ECHR (2019) Nicolae Virgiliu Tănase v. Romania [GC], §§ 128
[751] Cf. ECHR (2012) Von Hannover v. Germany (No. 2), [GC], §95
[752] Cf. ECHR Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland [GC], § 134
[753] Cf. ECHR (2019), Guide on Article 8 of the Convention – Right to respect for private and family life, p.20ff
[754] Cf. ECHR (1984) Malone v. The United Kingdom, [Concurring Opinion of Judge Pettiti]

to a person's enjoyment of her right to respect for private and family life as guaranteed by Article 8 of the Convention".[755] Consequently, domestic law is to ensure that personal data must be efficiently protected from misuse and abuse,[756] and therefore instate appropriate safeguards to prevent use of personal data inconsistent with Article 8.[757] This has heightened priority when automatic processing is in place.[758]

Storage of data relating to private life falls under Article 8, this is especially true for systematic collection and storage of private data by agents of a state, whereby this is intensified if some of the information is false.[759] This may be the case especially if systematic or permanent records are kept, even if it might come into existence from material in the public domain; this even when not gathered by intrusive or covert method.[760] Generally, in case data is collected, compiled, processed or published beyond a degree that is normally foreseeable, these actions fall under Article 8.[761] As outlined above in Section 0, protection of personal data must also be afforded when this data is already in the public domain.[762] Article 8 provides for informational self-determination allowing individuals to rely on their right to privacy as regards data which, albeit neutral, are collected, processed and disseminated collectively and in such a form or manner that the Article 8 rights of the individuals concerned are engaged.[763]

With respect to online activities, data connected with (dynamic) IP addresses which may help connect a person to its online activity is considered to be non-public in nature, and therefore might also be encompassed in the scope, especially given the fact that anonymity during online activities can be considered to be important, even when no specific steps are taking to conceal the user's identity.[764]

Medical data is seen as highly sensitive; collection and storage of health-related data for an extended period, and their subsequent disclosure without the patient's consent und use for unrelated purposes, as well as collection by institutions tasked with quality control has been found in cases to violate Article 8.[765] (Specifically, the court has dealt with cases in which such medical information was at risk to be used when determining an individual's job prospects.[766] Note that if such considerations are done by an informational agent, this could lead to limiting the individual's autonomy as described in Section 5.4.2.1

Location Data is similarly protected; collection and storage of data by a satellite tracking and navigation system that is attached to an individual's car has previously been found to rise to the level of interference.[767]

[755] Cf. ECHR Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland [GC], § 133
[756] Cf. ECHR (2009) Gardel v. France, § 62
[757] Cf. ECHR (1997) Z v. Finland, § 95
[758] Cf. ECHR, (2009) Gardel v. France, § 62
[759] Cf. ECHR (2000) Rotaru v. Romania [GC], § 44)
[760] Cf. ECHR (2001) P.G. and J.H. v. the United Kingdom, § 57
[761] Cf. ECHR Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland [GC], § 136
[762] Cf. FN 752
[763] Cf. ECHR Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland [GC], § 137
[764] Cf. ECHR Benedik v. Slovenia, §107ff
[765] Cf. ECHR (2017) Surikov v. Ukraine §70
[766] Passim ECHR (2017) Surikov v. Ukraine
[767] Cf. ECHR (2011) Shimovolos v. Russia §65

Identity and Autonomy

Each person is entitled to a sphere of relations between other individuals, that might necessitate measures taken by the state to ensure its integrity and protection.[768] Within this sphere, an individual is meant to be able to freely pursue the development and fulfilment of her personality, indeed Article 8 aims to protect rights important to the individual's identity, self-determination, physical and moral integrity and an individual's relationships with others and their maintenance.[769] However, right to personal development and autonomy does not protect all potential public activities an individual might want to engage in with other individuals, neither does it protect interpersonal relations of such broad and indeterminate scope that there is no conceivable direct link between the action or inaction of a state and a person's private life. [770]

Correspondence

Censorship, interception, monitoring, seizure or other impediments of correspondence (i.e. communication) is considered to be interference under Article 8.[771] This applies to electronic communications; such as telephone calls and emails(even from business premises), Internet usage[772], but also "archived" correspondence.[773] Explicitly, the court has found screening[774], copying[775], interception and monitoring (by a third party)[776], metering (logging of connections e.g. telephone numbers called)[777] and storage[778] to constitute an interference. Mass surveillance undertaken by the state naturally falls within the scope of Article 8 and requires meeting specific benchmarks.[779] It is noteworthy that there is only limited case law with respect to positive obligations (see above Section 0) to regulate the relationships of individuals with respect to the right of protection of correspondence, most notably concerning monitoring of employee-correspondence[780].

[768] Cf. ECHR (2017) A.-M.V. v. Finland §71
[769] Cf. ECHR (2017) A.-M.V. v. Finland §76, (2004) Pretty v. the United Kingdom § 82, (2002) Christine Goodwin v. the United Kingdom [GC] §90
[770] Cf. ECHR (2009), Friend and Others v. the United Kingdom §40
[771] Cf. ECHR (1975) Golder v. the United Kingdom §43
[772] Cf. ECHR (2007) Copland v. the United Kingdom §41
[773] Although searches of locally electronically stored data will usually already interfere with the right for respect for an individual's private life, cf. ECHR (2007) Wieser and Bicos Beteiligungen GmbH v. Austria §45
[774] Cf. ECHR (2007) Copland v. the United Kingdom §44
[775] Cf. ECHR (2000) Foxley v. the United Kingdom §30
[776] Cf. ECHR (1998) Lambert v. France §21
[777] Cf. ECHR (1984) Malone v. The United Kingdom §83f
[778] Cf. FN 774
[779] Passim ECHR (2015) Roman Zakharov v. Russia [GC]
[780] Cf. ECHR (2017) Bărbulescu v. Romania [GC], §§115ff

Annex 4.7.3.3

Article 9 of the European Convention on Human Rights

Again, within the following section, the term "court" is meant to denominate the European Court of Human Rights, and the term "Article" is meant to denominate an Article of the European Convention on Human Rights if not specified otherwise.

Normative Content

Article 9 of the ECHR reads as follows:

> *1. Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief and freedom, either alone or in community with others and in public or private, to manifest his religion or belief, in worship, teaching, practice and observance.*
>
> *2. Freedom to manifest one's religion or beliefs shall be subject only to such limitations as are prescribed by law and are necessary in a democratic society in the interests of public safety, for the protection of public order, health or morals, or for the protection of the rights and freedoms of others*[781]

Nature of Obligation

Similarly to Article 8, Article 9 of the ECHR stipulates a negative obligation for states not to arbitrarily interfere or place limitations on the protected rights.[782] Article 9 precludes criminal or administrative penalties, but also psychological pressure (exerted by State representatives to an individual to abandon her beliefs), [783] or imposing physical obstacles to the exercising of rights.[784]

However, just as with other provision of the ECHR positive obligations exist in which the state must actively set out to ensure protection of the rights granted. This might entail creating effective and accessible means of protecting the rights guaranteed by Article 9,[785] but does not necessarily guarantee the right to benefit from preventive measures.[786]

Protective Scope

Religious and Non-Religious Convictions

The most commonly adjudicated part of Article 9 is its protection of religion, which is not relevant for purposes of this inquiry, however it also deals more generally with thoughts, beliefs, views and convictions. Separating those concepts from religion is difficult, as the very idea of defining religion might lead to undue withdrawal of religious protection from certain individuals. In any case, most case law does deal with issues of religion and its manifestation first. Due to the vague separation between religion and other parts of thoughts, beliefs, views

[781] Cf. generally for the following section ECHR (2020), Guide on Article 9 of the Convention – Freedom of thought, conscience and religion, available at https://echr.coe.int/Documents/Guide_Art_9_ENG.pdf.
[782] See also section 0 for the distinction between positive and negative obligations.
[783] Cf. ECHR (2018), Mockuté v. Lithuania, § 123f.
[784] Cf. ECHR (2011), Boychev and Others v. Bulgaria
[785] Cf. ECHR (2017), Osmanoğlu et Kocabaş v. Switzerland, § 86
[786] Cf. ECHR (1996), Hernandez Sanchez v. Spain (Commission Decision)

or convictions some room will still be given to decisions in which the case law can conceivably be transferred to non-religious instances. An individual's conviction (or thought or view or belief) must attain a certain level of cogency, seriousness, cohesion and importance in order to fall within the protective scope.[787] Among such convictions are not merely religious views but also philosophical or deeply personal ones such as pacifism,[788] veganism,[789] medical philosophy with respect to alternative medicine, [790] or the actively reflected absence of religion, such as secularism[791], and atheism,[792] as well as convictions with respect to conscientious objections.[793] Notably, the court has named the rights reflected in Article 9, i.e. freedom of thought, conscience and religion, a "precious asset".[794]

Internal and Manifested Beliefs

Article 9 § 1 provides protection for both holding a religion or belief and manifesting it. This phrase omits the terms thought and conscience, however, the belief in question must not be religious. Protection of internally maintaining a conviction is absolute and unconditional, that is no coercive steps may be taken to change the conviction by a state. Protection of the right to manifesting one's conviction is less pronounced; as it (and only it) is subject to additional limitations as set out in Article 9 § 2.[795] Manifestation of a belief are acts that are intimately linked to the belief in question; a sufficiently close and direct nexus must be determined on the facts of each case.[796] Important in practice, the sincerity of a belief cannot be questioned by authorities of the state.[797] The extension of protection of Article 9 § 1, again with reference to cogency, seriousness, cohesion and importance, is limited. Protection is not granted with respect to just any thought; a lack of "inconvenience" ("*un désagrément suffisant"*) might preclude application.[798]

Positive and Negative Aspects of Article 9

Article 9 entails positive and negative aspects of protection. Negative aspects include issues of conscientious objections or coercion to reveal one's belief. Conscientious objection is not necessarily limited to the military domain, but broadly connects to actions which are contrary to one's conscience and convictions. However protection is strongly restricted by public interest in the civilian field.[799]

Positive aspects are connected with the manifestation of one's belief. Not every act motivated by belief is protected, indeed it must be intimately linked,[800] and Article 9 does not necessarily secure the right to act inspired or dictated by ones' beliefs.[801] Article 9 also deals with the

[787] Cf. ECHR (2013) Eweida and Others v. The United Kingdom §81 with further references
[788] Cf. ECHR (1978) Arrowsmith v. the United Kingdom (Commission Report), §69
[789] Cf. ECHR (1993) W. v. the United Kingdom (Commission decision)
[790] Cf. ECHR (1998) Nyyssönen v. Finland (Commission decision)
[791] Cf. ECHR (2011) Lautsi and Other v. Italy [GC] § 58
[792] Cf. ECHR (1986) Angeleni v. Sweden (Commission decision); (1994) Union des Athées v. France (Commission report) § 79
[793] Cf. just ECHR (2020), Factsheet – Conscientious objection, available at https://www.echr.coe.int/Documents/FS_Conscientious_objection_ENG.pdf
[794] Cf. ECHR (1993) Kokkinaki v. Greece § 31
[795] Cf. ECHR (2007) Ivanova v. Bulgaria § 79
[796] Cf. ECHR (2014) S.A.S. v. France [GC] § 55
[797] Cf. ECHR (2014) S.A.S. v. France [GC] § 56. Slightly different case law exists about the sincerity of an individuals religion in some cases.
[798] Cf. ECHR (1999) Viel v. France (Decision)
[799] Cf. ECHR (2013) Eweida and Others v. The United Kingdom §105
[800] Cf. FN 796
[801] Cf. ECHR (1997) (Kalaç v. Turkey)27

concept of paternalism, especially in the medical domain. Indeed the court explicitly notes the connection with freedom and personal autonomy.[802] Article 9 also aims to protect the mental integrity of individuals in certain cases; as individuals are protected against religious indoctrination by a state.[803]

[802] ECHR (2010), Jehovah's Witnesses of Moscow and Others v. Russia, 36
[803] Cf. ECHR (1986), Angeleni v. Sweden (Commission Decision), (1996) C.J. J.J. and E.J. v. Polan (Commission Decision)

Annex 10.2 (a)

Misinformation

On an abstract level, information (of any veracity) can be understood as well-formed, meaningful data [804]; whereas data may be understood as „a putative fact regarding some difference or lack of uniformity within some context“ [805]. This data (i.e. information) can be further be classified by its truth-value or factuality. If such data is true we tend to describe it merely as information (while the term true information would be more accurate in light of the above terminology), while if such data is untrue we may describe it as misinformation. A distinction is made between the concepts of misinformation and disinformation, although the term seems to be used interchangeably by some [806]. Misinformation may be generally defined as information that is false or misleading [807]. Disinformation shares this scope but is thought to add a qualifier of intent; the flawed information ought to be spread with intent to deceive, influence, obscure or mislead [808]. Disinformation encompasses at least partly the now infamous concept of “fake news”, which in part may be seen as encompassing elements such as news satire or parody, fabrication or manipulation [809]. One may hence consider disinformation to be a “subset” of misinformation [810]. For the purpose of simplicity, in this contribution the term misinformation will be used to encompass both misinformation in its described meaning above as well as disinformation.

Adding to challenges brought by misinformation, corrective measures are often difficult to apply. Misinformation and its effects can linger and create resistance to the introduction of more veracious information [811]. Content and timing of corrective measures may play a role in reverting the effect of misinformation [812], as seems to be the emotional attachment of the

[804] Luciano Floridi, *Philosophy and Computing: An Introduction*; Mingers and Brocklesby.
[805] Luciano Floridi, ‘Semantic Conceptions of Information’.
[806] Liang Wu and others, ‘Misinformation in Social Media’, *ACM SIGKDD Explorations Newsletter*, 21.2 (2019), 80–90 <https://doi.org/10.1145/3373464.3373475>.
[807] Vraga and Bode.
[808] Eleni Kapantai and others, ‘A Systematic Literature Review on Disinformation: Toward a Unified Taxonomical Framework’, *New Media & Society*, 23.5 (2021), 1301–26 <https://doi.org/10.1177/1461444820959296>; Daniel Chandler and Rod Munday, ‘Disinformation’, in *A Dictionary of Media and Communication* (Oxford: Oxford University Press, 2011); Marriam-Webster (Dictionary), ‘Disinformation’, 2021; Cambridge Dictionary, ‘Disinformation’, 2021; Michelle A. Amazeen and Erik P. Bucy, ‘Conferring Resistance to Digital Disinformation: The Inoculating Influence of Procedural News Knowledge’, *Journal of Broadcasting & Electronic Media*, 63.3 (2019), 415–32 <https://doi.org/10.1080/08838151.2019.1653101>.
[809] Edson C. Tandoc, Zheng Wei Lim, and Richard Ling, ‘Defining “Fake News”’, *Digital Journalism*, 6.2 (2018), 137–53 <https://doi.org/10.1080/21670811.2017.1360143>.
[810] André Calero Valdez, ‘Human and Algorithmic Contributions to Misinformation Online - Identifying the Culprit’, in *Disinformation in Open Online Media*, ed. by Christian Grimme and others (Cham: Springer International Publishing, 2020), pp. 3–15.
[811] Stephan Lewandowsky and others, ‘Misinformation and Its Correction: Continued Influence and Successful Debiasing’, *Psychological Science in the Public Interest*, 13.3 (2012), 106–31 <https://doi.org/10.1177/1529100612451018>; E. F. Loftus, ‘Planting Misinformation in the Human Mind: A 30-Year Investigation of the Malleability of Memory’, *Learning & Memory*, 12.4 (2005), 361–66 <https://doi.org/10.1101/lm.94705>.
[812] Hollyn M. Johnson and Colleen M. Seifert, ‘Sources of the Continued Influence Effect: When Misinformation in Memory Affects Later Inferences.’, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20.6 (1994), 1420–36 <https://doi.org/10.1037/0278-7393.20.6.1420>; R. Kelly Garrett and Brian E. Weeks, ‘The Promise and Peril of Real-Time Corrections to Political Misperceptions’, in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW ’13* (New York, New York, USA: ACM Press, 2013), p. 1047 <https://doi.org/10.1145/2441776.2441895>.

individual relying on the information in question [813]. Faced with contradicting information, individuals may actually be driven towards their initial (misinformation-fuelled) stance, until the factually unsupported position becomes completely unsustainable [814].

Conceivably, misinformation may relate to any matter of life. Harking back to the initial definitions, any well-formed meaningful data that is of factual (as opposed to e.g. instructional) nature can conceivably be true or untrue. Conceptually, every matter that is accessible to factual claims can also be affected to misinformation. In practice, public and academic attention is focused on areas of high risk, such as misinformation with respect to elections, national security, or public health. In these contexts, misinformation is considered to be potentially erosive of current societal structures e.g. via loss of trust or emergence of polarization and extremism, etc. by institutional observers [815] and academic researchers alike [816]. To this end, misinformation has been linked to an increase in hate crimes [817]. At the pinnacle of risk potential is misinformation that incurs direct and immediate danger to human lives, such as in the case of "swatting" [818] in which misinformation is provided to police forces to provoke the use of deadly force against a third party.

The spread of misinformation is (among other things) dependent on the domain in which it is transmitted. Critically, misinformation has been observed to spread faster than true information within open online media. [819]

[813] Leonie Schaewitz and Nicole C. Krämer, 'Combating Disinformation: Effects of Timing and Correction Format on Factual Knowledge and Personal Beliefs', 2020, pp. 233–45 <https://doi.org/10.1007/978-3-030-61841-4_16>.
[814] David P. Redlawsk, Andrew J. W. Civettini, and Karen M. Emmerson, 'The Affective Tipping Point: Do Motivated Reasoners Ever "Get It"?', *Political Psychology*, 31.4 (2010), 563–93 <https://doi.org/10.1111/j.1467-9221.2010.00772.x>.
[815] European Commission, 'Action Plan against Disinformation', 2018; Carme Colomina, Héctor SÁNCHEZ Margalef, and Richard Young, *The Impact of Disinformation on Democratic Processes and Human Rights in the World (Study Requested by the EU-DROI Subcommittee)*, 2021.
[816] Nygren and others; David M. J. Lazer and others, 'The Science of Fake News', *Science*, 359.6380 (2018), 1094–96 <https://doi.org/10.1126/science.aao2998>; Lewandowsky and others; Christina Mason, Peter van der Putten, and Max van Duijn, 'How Identity and Uncertainty Affect Online Social Influence', 2020, pp. 174–90 <https://doi.org/10.1007/978-3-030-61841-4_12>; Natalie Jomini Stroud, 'Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure', *Political Behaviour*, 30.3 (2008), 341–66 <https://doi.org/10.1007/s11109-007-9050-9>.
[817] Karsten MMller and Carlo Schwarz, 'Making America Hate Again? Twitter and Hate Crime Under Trump', *SSRN Electronic Journal*, 2018 <https://doi.org/10.2139/ssrn.3149103>; Leonardo Bursztyn and others, *Social Media and Xenophobia: Evidence from Russia* (Cambridge, MA, December 2019) <https://doi.org/10.3386/w26567>.
[818] Matthew James Enzweiler, 'Swatting Political Discource: A Domestic Terrorism Threat', *Notre Dame Law Review*, 90 (2015), 2001.
[819] Soroush Vosoughi, Deb Roy, and Sinan Aral, 'The Spread of True and False News Online', *Science*, 359.6380 (2018), 1146–51 <https://doi.org/10.1126/science.aap9559>.

Annex 10.2 (b)

Location-Based-Services

The term Location Based Services (LBS) broadly describes the combination or integration of location information with other types of services, although there seems to be no full consensus on an exact definition.[820] While the concept of LBS used to primarily denote applications for the explicit purpose of navigation or their directly adjacent uses such as mobile guides or navigation systems,[821] the field of relevant LBS applications has become more complex in the recent years. LBS are now interwoven with applications in many different domains both serious, such as disaster and emergency context, or recreational, such as gaming or fitness.[822] This mirrors are larger trend of technological processes becoming both more capable and widespread. This essay uses the term LBS device mostly with reference to devices that have capabilities to determine their geospatial location themselves, e.g. via use of GNSS. However, it may be advantageous to understand the term LBS as to denote informational agents that deal with location data in a wider sense, or further down into the information pipeline, e.g. a service that processes location data but does not collect or affirm it itself. Such wide terminology necessarily softens the scope of what ought to be called a LBS device as a consequence. While data processing is increasingly done not only on-device but server-side, this allows the use of LBS on an increasing number of devices, so long as they have interconnectivity capabilities. In this sense it may be useful to differentiate between core LBS devices that can determine their own geospatial location and other LBS devices that cannot, but are still able to provide LBS.

The increasing prevalence of LBS is aided by the adoption of devices capable of creating and processing relevant location data and of an improvement of existing infrastructure. Mobile phones, a class of devices that tend to have LBS-capabilities, have become a major access point for many people as their use continues to rise .[823] At the same time, there is an increasing amount of devices that add LBS-capabilities to their existing set of primary features such as fitness heartrate monitors or "smart-watches" that may receive GNSS information or have navigational capabilities, with these devices being considered as being part of the "Internet of Things". Interconnectivity allows interaction on the basis of an individual's location data, thereby multiplying the respective use cases of said data beyond mere confirmation of location. To this end, the increasing adoption of LBS is also supported by a consistent improvement of Internet infrastructure as displayed by continuously increasing bandwidth,[824] as characterized in the previously mentioned observation called "Edholm's Law".[825] This holds true especially for wireless interconnectivity which is improving at a faster rate than wired connections,[826] adoption of which further favours the mobile nature of LBS-devices. An LBS device may passively receive information about its location, process this information offline, i.e. on the

[820] Jochen Schiller and Agnès Voisard, *Location-Based Services* (Elsevier, 2004) <https://doi.org/10.1016/B978-1-55860-929-7.X5000-6>; Jonathan Raper and others, 'A Critical Evaluation of Location Based Services and Their Potential', *Journal of Location Based Services*, 1.1 (2007), 5–45 <https://doi.org/10.1080/17489720701584069>.
[821] Jonathan Raper and others, 'Applications of Location–Based Services: A Selected Review', *Journal of Location Based Services*, 1.2 (2007), 89–111 <https://doi.org/10.1080/17489720701862184>.
[822] Haosheng Huang and others, 'Location Based Services: Ongoing Evolution and Research Agenda', *Journal of Location Based Services*, 12.2 (2018), 63–93 <https://doi.org/10.1080/17489725.2018.1508763>.
[823] Pew Research Center.
[824] International Telecommunication Union, *Measuring Digital Development - Facts and Figures*.
[825] Cherry.
[826] Rappaport; Ericsson, *Mobility Report*.

device, and then proceed by exercising its programmed function. For example, a satellite navigation device used for car-based navigation may conduct its service, i.e. geospatial positioning and consequent route-finding, based on map data that is stored on the device. The device may collect and store, i.e. log, part of this information even after the individual service is provided, e.g. to display a list of "last-used destinations" for the convenience of the user, but that information is likely to remain on the device. This makes the data collected of little use for the provider of an LBS. However, as pointed out above, many LBS devices do now possess interconnective capabilities and location data is used in many more ways than route finding. As LBS are now often embedded into other services that also rely on interconnectivity, e.g. social media, communication, etc., data collected on the device can now be shared (back) with the LBS provider, allowing the analysis of said data irrespective of the device's capabilities or the user's active participation. The increase in capability of LBS devices, use and use cases of LBS, and capability of interconnective infrastructure together with improvements in data analysis suggests that individuals leave behind "data trails" that are both more comprehensive in a quantitative sense as well as more informative generally. This likely creates an economic incentive to maximize access to such information.[827] At the same time, increased data acquisition may also be motivated by improving the LBS via collection and analysis of non-location based information as context.[828] This positions LBS as a dualistic phenomenon with respect to data: LBS and their immediate location data collection and analysis processes may be incorporated into other information processing routines for context; at the same time LBS also benefits from additional non-location data for context all the same.

[827] Zuboff, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization'.
[828] Albrecht Schmidt, Michael Beigl, and Hans-W Gellersen, 'There Is More to Context than Location', *Computers & Graphics*, 23.6 (1999), 893–901 <https://doi.org/10.1016/S0097-8493(99)00120-X>; Ozgur Yurur and others, 'Context-Awareness for Mobile Sensing: A Survey and Future Directions', *IEEE Communications Surveys & Tutorials*, 18.1 (2016), 68–93 <https://doi.org/10.1109/COMST.2014.2381246>; Patrizia Grifoni, Arianna D'Ulizia, and Fernando Ferri, 'Context-Awareness in Location Based Services in the Big Data Era', 2018, pp. 85–127 <https://doi.org/10.1007/978-3-319-67925-9_5>.