UNIVERSITÀ DEGLI STUDI DI MACERATA

Department of Political Sciences, Communication, and International Relations

Ph.D. Course in

Global Studies. Justice, Rights, Politics

Cycle XXXIII

Title

MORAL FREEDOM

FREE TO CHOOSE IN THE ALGORITHMIC ERA

Academic Supervisors

Professor Benedetta Giovanola

Professor Benedetta Barbisan

Ph.D. Candidate

Simona Tiribelli

Scientific Coordinator

Professor Benedetta Barbisan

Academic Year

2021

To my parents
Daniela and Graziano,
with love and gratitude

INDEX

Introduction

What do we mean by *moral freedom*? Which 'necessary conditions' does it require? Do the exponential rise and use of algorithm-based information and communication technologies (digital ICTs) promote or undermine it? Are we dealing with a novel ethical challenge? If this is the case, how should we respond to it? These are a few of the main questions my dissertation aims at addressing.

These queries are increasingly ineludible in our contemporary informational societies, i.e., in societies where our pervasive use of algorithms-based ICTs, and specifically our almost totalizing presence in social networking services (SNS), has determined the rapid blurring of the distinction between online and off-line, by arising a renovated habitat from the mutual interaction, mixing, and hybridization between cyber-space and physical space. In this renewed "onlife"[1] space, we are *ubiquitous* digitally interconnected, onlife informational beings, who relentlessly produce, are fueled by, and process information. These questions are extremely meaningful, thus, in mature informational societies, where people increasingly depend on digital ICTs and, therefore, are more and more exposed to their invisible but unavoidable algorithmic design.

In fact, it is a matter of fact that in our advanced informational societies we live an environment that is not just digitalized but also algorithmically-mediated. Algorithms-based ICTs – from Internet search engines (Google, Bing, and so forth) and SNS (Facebook, Twitter, YouTube…) to the countless applications of Internet of Things (IoTs) – have become a 'necessary condition' to perform many of our most common activities, daily tasks, and large-scale projects, from the sectors of communication, administration, transports, healthcare, justice, and education, to those of industrial production, defense, finance, and energy supply.[2] In this regard,

[1] For a wide analysis on how the digital revolution (also called the "fourth revolution") by blurring the distinction between online and offline has reshaped our reality, our relations, and even ourselves as "onlife informational beings", the reference is L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford 2014.

[2] For a systematic overview of the numerous applications of algorithms-based ICTs in Europe, see F. Chiusi, S. Fischer, N. Kayser-Bril, & M. Spielkamp (eds.), "Automating Society Report 2020", *AW AlgorithmWatch*, Berlin (Germany) 2020.

COVID-19 pandemic has shown us – probably more than anything else – how much we and our societies rely to psychologically, socially, and economically survive on algorithmic ICTs.[3] From the standpoint of people's psychological health, SNS have played a prominent role, by rescuing people from a total isolation, permitting online communication, and the sharing of social experiences between individuals 'forced' to stay home. Online platforms (e.g., Zoom, Google Meet, and so on) have allowed social infrastructures as schools and universities to perform their key function, as well as many companies to do not economically fall apart, therefore, to continuing to ensure to a certain extent jobs and societal basic services.

But, algorithmic ICTs, as media philosopher Marshall McLuhan has noticed a long time ago (1964), far beyond being mere and neutral tools, not just mediate every domain of our life. They are transformative vectors: social and environmental forces[4] with both an epistemological and ontological impact. Indeed, given their role as 'informational gatekeepers'[5] – i.e., having an epistemic role and function in overseeing and, above all, governing information – algorithms-based ICTs shape the way in which we perceive and understand our reality, and thus, they impact on how we develop knowledge, for instance, just by filtering how we get information. Moreover, to the extent they forge new virtual dimensions – where we spend a huge amount of time – which in their turn hybridize our world deeply, they are also ontological forces that concretely restructure our reality, by deeply blending into

[3] For a discussion on the role of digital ICTs to mitigate the impact of COVID-19 on individuals' mental health, consider J. Torous *et alia*, "Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow", *JMIR Mental Health*, 7(3), 2020. For a critical overview of how digital ICTs have been harnessed to support societal healthcare and enhance public health response to COVID-19 worldwide, including health population surveillance, cases identification, contact tracing, and evaluation of interventions on the basis of data , see J. Budd, B.S. Miller, E.M. Manning, *et alia*, "Digital technologies in the public-health response to COVID-19", *Nature Medicine* 26, 2020, pp. 1183-1192. To analyze how digital information technology has been used to enhance resilience and continuity of business to recover from adversity resulting from the COVID-19 pandemic, see F.FH. Nah & K. Siau, "COVID-19 Pandemic – Role of Technology in Transforming Business to the New Normal", *HCI International 2020. Late-Breaking Papers: Interaction, Knowledge and Social Media. HCII 2020. Lecture Notes in Computer Science* (edited by C. Stephanidis *et alia*), 12427 (2020).

[4] R. Silverstone, *Media and Morality: On the Rise of Mediapolis*, Polity Press, Cambridge (MA) 2007.

[5] C. J. Calhoun, *Dictionary of the social sciences*, Oxford University Press, New York 2002.

and exceptionally reshaping our shared common practices, from how we make interactions and experience social relations[6] to how – through them – we develop our identity.[7] As it has been pointed out by a growing corpus of literature in the field of ethics of algorithms[8], this ascent of a renewed environment unceasingly shaped by pervasive, powerful, and continuously evolving algorithmic forces, is raising either new wonderful opportunities (e.g., decreasing time and costs in performing many activities, boosting performances, and advancing research and science) or – as they have disruptive potential – new risks and potential perils, from the way in which we develop social relations, beliefs, values, and identity, up to our democracy and freedom[9], therefore by asking imperatively for the afterthought of some of the most crucial issues above mentioned as inevitably influenced by algorithmic ICTs.

The present inquiry answers to this ethical call and aims at investigating the reshaping impact of these algorithmic forces on our freedom, and specifically, on our moral freedom, that is, our freedom to choose and act as genuine moral agents. Let me define the specific scope of this inquiry and the reasons behind the choice of this specific topic.

The overall purpose of this dissertation is precisely to investigate one of the most crucial issues in our contemporary societies in the light of the recent advancements in the algorithmic technology: the issue of our freedom, specifically, our *moral freedom*.

By moral freedom I mean our freedom to effectively become moral agents, that is, our freedom to choose and act as moral agents, and specifically, as genuine moral agents. For this reason, moral freedom also means our freedom from moral

[6] M. Parsell, "Pernicious Virtual Communities: Identity, Polarisation and the Web 2.0", in *Ethics and Information Technology*, 10 (1) 2008.

[7] M. Bakardjieva & G. Gaden, "Web 2.0 Technologies of the Self", in *Philosophy and Technology*, 25(3) 2012.

[8] B. Mittelstadt, P. Allo *et Alia*, "The Ethics of Algorithms: Mapping the Debate" in *Big Data & Society* 2016.

[9] On the controversial impact of algorithmic ICTs on the way in which we develop social relations and identities, we form opinions and beliefs, and we build our public sphere and democracy, see C. Sunstein, *Republic.com*, Princeton University Press, Princeton, NJ (USA) 2001; see also C. Sunstein, *Republic.com 2.0*, Princeton University Press, Princeton NJ (USA) 2007, and *#Republic: Divided Democracy in the Age of Social Media*, Princeton University Press, Princeton NJ (USA) 2017.

and immoral interferences to our capacity to become genuine moral agents and to develop over time a genuine moral identity.

Therefore, reflecting on moral freedom entails questioning when, namely, under what specific conditions, the human agency can be defined as morally free.

For this reason, I will inquire into the moral dimension of human choosing and agency, especially for what concerns the conditions of possibility which allow its free exercise. I will define these conditions as the *conditiones sine qua non* of our moral freedom. These are respectively:

a) the *availability of morally heterogeneous options*, namely, and agent is morally free, i.e., free to become a genuine moral agent, if she can choose and act otherwise from what *de facto* she does, that is, if and only if she can choose among different options which reflect diverse courses of actions. More specifically, an agent is free to choose and act as a moral agent, and therefore to develop genuine moral identity, if and only if she is free to form her own idea of good and moral ground projects, and so to develop and choose values, reasons, and motives through a morally heterogeneous exposure. In other words, an agent is morally free if and only if there is an availability of morally heterogeneous options that embed a plurality of values and moral orientations, on which the agent can morally reason and find the moral motives for her choices and actions;

b) *moral autonomy*, that is, the condition in which the agent is free to become a genuine moral agent if she can be the author of her choices and actions, and therefore the author of her moral identity, namely, if and only if she can reflectively endorse (or internally approve) those options, amongst those available, that respond to her own idea of good, to her values or moral ground-projects, as moral reasons or motives for her choices and actions. In other words, an agent is free to become a genuine moral agent if and only if she can choose and act in accordance to her moral values, by endorsing them as motives underlying her choices and actions, and thus as reasons on which building over time her moral identity.

However, reflecting on our moral freedom *today*, i.e., in our increasingly algorithms-based informational societies, entails also a further rethinking of moral freedom in the light of new actual and potential forms of impediment raiseable by algorithms-based ICTs – intended as "unprecedented growing resources of smart, interactive, [semi-]autonomous, and self-learning agency"[10] – on the *conditiones sine qua non* underlying the exercise of our moral freedom.

Therefore, the specific issue I will examine is *whether* – and, if this is the case, *how*, and with *which foreseeable consequences* – algorithms-based ICTs can undermine our moral freedom, that is, our freedom to choose and act as genuine moral agents in the algorithmically-governed environment of our contemporary advanced informational societies, by affecting and hampering *the conditiones sine qua non* underlying its exercise.

The specific thesis I maintain in the current work is that algorithms-based ICTs do not just can impact and influence the *conditiones sine qua non* of our moral freedom. Beyond their influencing and reshaping action on them, algorithmic ICTs can have a detrimental impact on these conditions and endanger our moral freedom, by raising a novel threat to our freedom to choose and act as moral agents, what I define as algorithmic predeterminism: an unprecedented form of predeterminism based on the predictive potential of algorithms, whose controversial application can generate intentional or accidental, moral and unmoral, interferences on people's capacity to form their own ideas of good and their own ground projects to the point of undermining their freedom to choose, act, and so become genuine moral agents.

As it will become clearer later on, information plays a crucial role when we reason about what makes us socially, politically, and – above all – morally free. In fact, our decision-making processes and behaviors, from our deliberations, options, and choices, to the formation processes of beliefs, reasons, preferences, judgments,

[10] L. Floridi & J. Cowls, "A Unified Framework of Five Principles for AI in Society", *Harvard Data Science Review*, 1(1) 2019.

and motives which inform and then steer our action, are all influenced and shaped by the perception and the intake of information. As a consequence, if the information that algorithms-based ICTs make available and present us is false, partial, inaccurate, manipulated, biased, or intentionally malicious, our moral alignment to our choices and so the morally genuine development of our identity may be compromised. More specifically, our moral freedom – our freedom to choose and act as genuine moral agents, namely, according to reasons and motives we deeply endorse –, as I will argue later on, risks to be profoundly eroded.

These ethical concerns about the influencing power of algorithm-based ICTs and their potential negative consequences on human choice behavior have tremendously grown over the last years.[11] The watershed has been undoubtedly marked by the Cambridge Analytica's scandal: the public disclosure of the 'misuse' of profiling algorithms, that regulate the functioning of our most common ICTs, to capture people's personal information, by inferencing them from billions of users' data with the goal of subtly influencing people via targeted advertising techniques and reshaping their behavior (specifically, their socio-political choice-orientation and vote) according to third-party interests.[12]

Indeed, initially, public concern about the potential of algorithms ICTs' to infer users' features (for example: preferences, interests, habits, friends, education, employment, health status, and financial standing) was only framed in the context of advertising targeted techniques for commercial goals.[13] As a consequence, the main research and solutions were just thought and focused on modernizing privacy and consumer protection regulations.[14]

[11] For a review of the current debate on the ethical aspects of algorithms see B. Mittelstadt, P. Allo *et Alia*, "The Ethics of Algorithms: Mapping the Debate" in *Big Data & Society* 2016. For an analysis of the epistemic and normative concerns related to the exponential application of machine-learning algorithms, consider A. Tsamados, N. Aggarwal *et alia*, *The Ethics of Algorithms: Key Problems and Solutions* 2020.

[12] See M. Rosenberg, "Bolton Was Early Beneficiary of Cambridge Analytica's Facebook's Data", The *New York Times*, 23 March 2018.

[13] N. M. Richards, The Dangers of Surveillance, *Harvard Law Review*, *126*(7), 2013, pp. 1934-1965.

[14] M. R. Calo, Digital Market Manipulation, *The George Washington Law Review*, *82*(4) 2014. See also J. Turow, *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*, Yale University Press, New Haven, London 2011.

Since the Cambridge Analytica's case, the scope of global awareness and outrage, as well as the range of ethical concerns, have extended considerably beyond the limits of the commercial sphere. For example, the disclosed intentionally malicious exploitation of algorithmic ICTs to influence individuals' political choice and broadly the elections have led to globally reckon with the fact there is something deeper at stake: the freedom of choice of individuals.

As a consequence, in a short time, algorithms-based ICTs, especially SNS, have revealed their controversial nature: from being positive places for users to connect with each other's day-to-day lives, share entertainment, and discuss popular culture, to becoming tools exploitable for the manipulation of public opinion and the reshaping of collective social-political choice-behavior[15] – a phenomenon observed in more than 48 countries, democratic and authoritarian alike.[16] After that infamous episode, what was only one of the fears raised by a niche of technophobic people has become a growing collective anxiety, so much so that to be globally recognized and framed by NGOs, tech policy experts, and institutional decision-makers, along with leading researchers, academics, and professionals in the field.

At the institutional level, a precise example of this recognition is provided by the adoption, on 13th of February 2019, by the European Union Committee of Ministers, of a formal document, the *Declaration on the Manipulative Capabilities of Algorithmic Processes*, which states how the targeted use of expanding volumes of citizens' aggregated data cannot just influence human behaviour in general, but may dangerously affect our exercise of freedom. The EU Committee emphasizes the role of algorithms-based ICTs to prompt users to disclose their relevant data, including personal data, for comparatively small awards of personal convenience, and stresses the peril derived from the use of these disclosed personal data to train

[15] Consider, for example, S. Vaidhyanathan, *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. Oxford University Press, Oxford 2018. See also Z Borgesius, F. J., Trilling *et alia*, Should We Worry About Filter Bubbles? *Internet Policy Review*, *5*(1) 2016.

[16] S. Bradshaw & P. Howard, *The Global Disinformation Order: 2019 Global Inventory of Organized Social Media Manipulation*, Project on Computational Propaganda (working paper), Oxford 2019.

machine-learning algorithms to prioritize search results, alter information flows, and subject people to behavioural experimentation.

As the EU Committee has remarked, the increasing possibility created by algorithms to infer users' intimate and detailed information in order to predict personal preferences, vulnerabilities, beliefs, and values from their readily available data, and then exploiting this predictive knowledge in order to sort and micro-target people through the reconfiguration of their social informational environment can lead to the corrosion of the «very foundation of the Council of Europe: Its central pillars of human rights, democracy, and rule of law are grounded on the fundamental belief in the equality and dignity of all humans as independent moral agents»[17]. Increasingly, indeed, this detailed information algorithmically inferable by users' data can be used by algorithms to sort individuals into categories and treat them differently from an informational standpoint, thereby reinforcing different forms of social, legal, cultural, and economic asymmetries, as well as exclusion and discrimination phenomena, so deeply infringing people's right to be treated and respected as equals.[18] Indeed, the algorithmic profiling and micro-targeting of individuals that often rule algorithmic decision making (ADM) have been widely criticized to generate controversial effects in terms of justice and fairness[19] and to embed a high number of flaws in terms of biases and discriminating profiling, from the ProPublica investigative report on racial biases in COMPAS risk-assessment ADM for predicting recidivism in the U.S. justice system[20] to Amazon's gender-

[17] EU Decl [13/02/2019], p. 1.

[18] The philosophical reference is J. Rawls. *A Theory of Justice*, Cambridge, MA: Harvard University Press 1971; and also, R. Dworkin, *Sovereign Virtue: The theory and Practice of Equality*. Cambridge, MA: Harvard University Press 2000.

[19] To expand, see V. Eubanks, *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY (USA): St Martin's Publishing 2018. S. Noble, *Algorithms of Oppression*, NYU Press, New York 2019; C. O'Neil, *Weapons of Math Destruction*, Penguin London 2016; R. Benjamin, *Race after Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Polity 2019.

[20] J. Angwin, *Machine Bias*. 2016, May 23.

biased recruitment algorithm[21] to the "Gender Shades Study" on gender and racial bias in ADM facial recognition software.[22]

Given that data-driven algorithms are designed to continuously achieve optimum solutions within the given parameters specified by their developers, when they operate at scale, such optimization processes inevitably prioritize certain values over others and thereby shaping the content in which users and non-users alike process information and make decisions. This reconfiguration of environment based on profiling and the categorization of people may then privilege some individuals, categories, and groups while detrimental to others and so it raises serious problems to a fair treatment of individuals in terms of both distributive and socio-relational justice. The EU document also stresses how much this algorithmic capacity and the «fine-grained, sub-conscious and personalized levels of persuasion [they can produce] may have significant effects on the cognitive autonomy of individuals and their right to form opinions and make independent decisions»[23] by adopting techniques of manipulation targeting users' thoughts and emotions, raising the possibility to alter an anticipated course of action, sometimes subliminally. The EU Committee concludes the *Declaration* by outlining how «these effects are today still underexplored, but cannot be underestimated»[24], by ultimately encouraging the EU member states in acknowledging the momentous need to deepen and address the risks raised by the manipulative capabilities of algorithms-based ICTs for our democracy and our freedom, by inviting the academia to produce interdisciplinary research regarding the capacity of algorithmic devices to reshape human behavior.

At the academic level, information technology philosophers Luciano Floridi and Mariarosaria Taddeo open the road for a deeper inquiry into the controversial influence of algorithmic ICTs on human choice behavior, inasmuch as – they claim

[21] J. Dastin, J, *Amazon scraps secret AI recruiting tool that showed bias against women.* Reuters. 2018, October 11.

[22] J. Buolamwini & T. Gebru, Gender Shades: Intersectional Accuracy Disparities. *Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR*, *81*, 2018, pp.77-91.

[23] EU Decl [13/02/2019], p. 3.

[24] *Ibidem*.

– they have the power to subtly «shape our choices and actions easily and quietly».[25] They sustain that «algorithms may exert their influencing power beyond our wishes or our understanding, undermining our control on our choices, projects, identities, and lives»[26]. The philosophers warn on the improper design and use of invisible algorithms in threatening our fragile, and yet constitutive, ability to determine our own lives and identities and keep our choices open.[27] Although they do not directly tackle the issue of freedom, nor the specific issue of moral freedom, they recognize that there is something deeper at stake in the impact of algorithms on our lives: the "openness of our choices" [28], that is another way to refer to our freedom of choice, and frame this ethical challenge as «one of the most relevant of our era, which must be addressed urgently»[29].

Nevertheless, not just the openness of our choices, as above mentioned, that is, our freedom of choice, is at stake in all the concerns evoked so far. Indeed, the thesis I argue is that at the core of all the concerns underlined above there is something morally deeper: what I define as our moral freedom, hence, a specific moral dimension of our choosing and agency, i.e., our freedom to choose and act as moral agents and so to develop our moral identity in a genuine way.

If there is a threat to our capacity to choose and act freely, and our environment is more and more permeated by invisible algorithms with the power to predict and even shape the elements that drive our choices, action, and behavior, then our moral freedom sounds to be deeply called into question. In this situation, an ethical inquiry is needed in order to assess the potential or actual impact of these algorithmic forces on our moral freedom.

I highlighted so far why the topic of moral freedom matters, and specifically, that there is a global call both from the institutional debate and academic field to deepen the issue of freedom, and especially as freedom of choice (even if it has been framed with a more generic terminology), in the light of the huge progress and wide penetration of algorithmic technologies into the social fabric of our lives.

[25] L. Floridi & M. Taddeo, "How AI Can Be a Force for Good", *Science*, 361 (6404) 2018*,* p. 752.

[26] *Ibidem*.

[27] *Ibidem*.

[28] *Ibidem*.

[29] *Ibidem*.

This is just one of the reasons motivating the present dissertation. There are indeed three further reasons behind this work.

Let me clarify them.

The first reason that motivates this dissertation is that, although the general debate around technology has very often expressed concern about freedom, so far, the analysis of the topic has been mostly addressed in the legal field among jurists and legal scholars, and specifically, the concept of freedom has been mostly understood as freedom of association and speech.[30] Instead, the philosophical issue of freedom, and specifically, the issue of our moral freedom, results almost unexplored in the fields of media ethics, ethics of information technology, and ethics of artificial intelligence (AI), and no analysis on freedom, especially on moral freedom, has been carried out through an approach deeply informed by moral philosophy *stricto sensu*.

Indeed, whilst freedom is widely acknowledged as a key moral value in the ethics and AI debate, an ethical inquiry into freedom, especially as moral freedom, drawing insights from accounts of freedom developed in moral philosophy, is largely missing in the current debate. Freedom, and specifically moral freedom, is an ethical issue *per antonomasia*, as it concerns the practical dimension of human choosing and agency, whereby any analysis that involves it is not possible without the consideration of the contribution of moral philosophy.

The first reason behind my work is hence to fulfill this specific gap in the current ethics and AI literature and so claim firstly that an ethical inquiry allows us to highlight the ethical-normative value of our moral freedom and secondly that an ethical inquiry into moral freedom in our increasingly algorithmically-governed societies is needed to adequately understand whether algorithms-based ICTs are

[30] See, for example, Council of Europe, *Guidelines on the Protection of Individuals with regard to the Processing of Personal Data in a World of Big Data*, 17 January 2017; U.N. *Human Rights Council Resolution on the Right to Privacy in the Digital Age*, U.N. Doc. A/HRC/34/7, 23 Mar. 2017. Few examples from the academic debate: E.T. Tavani, "Informational Privacy, Data Mining, and the Internet, *Ethics and Information Technology*, 1 1999; O. Tene & J. Polonetsky, "Big Data for all: Privacy and user control in the age of analytics", *Nw. J. Technology & Intellectual Property*. 2013; N. Richards, *Intellectual privacy. Rethinking Civil Liberties in the Digital Age*, Oxford University Press Oxford 2015; L. Taylor, L. Floridi, & B. van der Sloot, "Group Privacy: New Challenges of Data Technologies", *Springer*, New York, NY (USA) 2018.

positively or negatively affecting individuals' freedom to choose, act, and therefore become genuine moral agents.

The second reason behind this work lies in the prominence of the topic of moral freedom as the freedom to choose and act as moral agents. Indeed, the relevance of freedom of choice and action as a topic seems to be unquestionable. A huge number of thinkers have devoted many efforts to understand freedom, ranging from those who have tried to understand our freedom of choice by questioning its (im)possibility as a metaphysical issue to those that have tried to frame and measure it given a particular socio-political context, by developing two different debates that consider freedom according two different perspectives, the debate on free will on the metaphysical issue of freedom and the socio-political debate on the socio-political issue of freedom.

As I argue in the first chapter, however, the concept of moral freedom per se, and therefore not in reference to the free will debate or the debate on socio-political freedom, is less explored, above all in our contemporary societies. For this reason, the first section of my dissertation will commit to shedding light on the concept and the ethical-normative value of moral freedom, and specifically, to elaborate an account of moral freedom that allows us to adequately assess what connotes human agency in a moral sense. But, and most notably, this topic assumes even a greater importance today in the light of the rise of renovated environment pervaded if not already ruled by sophisticated and fine-grained human behavior conditioning techniques driven by algorithms-based ICTs. The huge predictive capacity of algorithms is today even reinforced by the growing use of ADM based on machine learning (ML) to which we are delegating more and more of our daily tasks, basic activities, and high-stake decisions. ML algorithms are indeed capable to scale massive amounts of data and infer associations on, profile, and categorize users with consequences that expand so much as the domains in which ADM are used, from social media communication and information management[31],

[31] On the topic, see E. Bozdag, "Bias in algorithmic filtering and personalization". *Ethics and Information Technology*, 15(3), 2013, pp. 209-227. S. Shapiro, Algorithmic Television in the Age of Large-scale Customization. *Television & New Media*, *21*(6) 2020; L.M. Hinman, Searching Ethics: The Role of Search Engines in the Construction and Distribution of Knowledge. In: Spink A., Zimmer M. (eds.). *Web Search. Information Science and Knowledge Management*,

advertising and marketing[32] to recruiting and employment[33], university admissions[34], housing[35], credit lending[36], criminal justice[37], policing[38], and healthcare[39]. Their profiling action combined to their pervasive use has been uncovered to be freedom-undermining in limiting users' options as access to real chances and social opportunities. The ethical inquiry into our moral freedom will not just allow us to understand the impact of algorithmic techniques on our capacity to choose and act as moral agents, namely, on the way in which we develop and choose our values and act in a way aligned to them, but it also allows to grasp the deep core of a problem that has huge social ramifications and implications, not just in terms of freedom, but also for justice and democracy.

The third reason behind the present inquiry lies in the ethical normative value of moral freedom: how I argue more broadly in the first chapter, especially in

Springer, *14*, 2018; E.B. Laidlaw, Private Power, Public Interest: An Examination of Search Engine Accountability. *International Journal of Law and Information Technology*, 17*(1)*, 2008, pp. 113–145.

[32] To expand the topic, see M. Hildebrandt, "Defining profiling: A new type of knowledge?". *Profiling the European Citizen* (edited by Hildebrandt M and Gutwirth S). The Netherlands: Springer, 2008, pp. 17-45; S. Coll, Consumption as biopower: Governing bodies with loyalty cards. *Journal of Consumer Culture*, *13*(3), 2013, pp. 201–220; Z. Tufekci, Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Journal on Telecommunications and High Technology Law*, *13*(203), 2015.

[33] See P. T. Kim, Data-Driven Discrimination at Work. *58 Wm. & Mary L. Rev*, *857* (3), 2017.

[34] See T. Simonite. *Meet the Secret Algorithm That's Keeping Students Out of College*. Wired. (2020, October 7

[35] See S. Barocas, & A.D, Selbst, "Big data's disparate impact". *SSRN Scholarly Paper*, Rochester, NY (USA): Social Science Research Network 2015.

[36] Consider J. Deville, *Leaky Data: How Wonga Makes Lending Decisions*. Charisma: Consumer Market Studies, May 20, 2013; K. Lobosco, *Facebook friends could change your credit score*. CNN Business.2013, August 27; M. Seng Ah Lee, & L. Floridi. Algorithmic Fairness in Mortgage Lending: From Absolute Conditions to Relational Trade-Offs. *Minds & Machines* 2020.

[37] See R. Berk, et al., Fairness in Criminal Justice Risk Assessments: The State of the Art. *Sociological Methods & Research*, 2018.

[38] On algorithms and criminal justice, see the work of A. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law*. New York, NY (USA): NYU Press 2017.

[39] On ADM and healthcare, see D. Danks, & A.J., London. Algorithmic Bias in Autonomous Systems. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization*, 2017, pp. 4691-4697. S. Robbins, A Misdirected Principle with a Catch: Explicability for AI. *Minds and Machines*, *29* (4), 2019, 495–514.

the third section, moral freedom is a *fundamental value* that has to be protected as an end for itself, as in it we find the exceptional and distinctive trait of our humanity: our capacity to develop and give expression to our moral standing by choosing and acting as moral agents, and specifically as genuine moral agents. In this value, we find the fundamental expression of the deeper meaning of our choices and actions, i.e., the emergence of our moral character.

To put it differently, moral freedom is what semanticizes, gives meaning to our choices and actions, and broadly, to our existence in the world, by bringing the moral dimension out from our being 'persons'.

Furthermore, moral freedom is an *intrinsic good*, as intrinsically valuable, i.e., something whose presence as capacity (irrespective of how it is used) adds to the value of our existence in the world. At the end of the first chapter, I will show why indeed a world with moral freedom is always better compared to a world without it, even if moral freedom is wrongly exercised by leading to morally undesirable outcomes. In a world with moral freedom indeed moral perfectionism, that does not coincide with moral freedom, it is a limit rather than a goal. This is because moral freedom can be further understood as an *axiological catalysator*, namely: a fundamental value whose presence confers or adds both more value or disvalue to the respect or non-respect of other values. As I will expand in the first chapter, moral freedom is what allows us to develop our moral identity or moral posture, that is, the dimension of our *ought to*, that allows ourselves to make the values we choose for ourselves moral rules for our choices, actions, and behaviors. Therefore, it needs to be safeguarded from forms of impediment to its exercise, inasmuch as it introduces an axiological difference in the dimension of our mere agency: the moral dimension, that is the dimension of our intentions that finds its moral expression in the reflective endorsement we can give or not to options, reasons, or values and allows us to align in our behavior with those we approve through the development of the obligation or ought to. In this sense, moral freedom makes choices and actions appraisable from a moral standpoint, by making choices and actions, good and bad, not on the basis of their consequences, but of the agent's intentions. In this sense, moral freedom makes agent's good and bad choices, respectively, better and worse, insofar as chosen in a context of freedom, of moral presence, and possibility of authorship of the agent. So, without it, it would be difficult to consider our good and bad actions as truly good or bad, as well as to attribute moral responsibility – and in certain cases also legal imputability. Indeed, without it, it would be difficult to evaluate whether the subject is the real agent of a choice, or whether something in the social environment she lives in did not allow her to choose alternatively, and whether she was the author of her choices and actions, or something has determined them in her place. In short, when we can guarantee moral freedom, we can assess, for better or worse, whether the subject who acts and chooses is morally present in her decisions and behavior. As a result, the protection of moral freedom as the freedom to choose and act assumes great ethical importance.

The protection of moral freedom entails today that the inquiry on the forms of impediment to its exercise expands in the light of this new algorithmically permeated environment, and therefore specifically assess if there are novel forms that may threaten it, beyond the existing ones.

In order to carry out my ethical inquiry into moral freedom and specifically on potential or actual technological forms of impediment to it, the methodology employed to conduct this philosophical inquiry is inspired by what John Rawls has called *reflective equilibrium*[40], which consists in working back and forth among our considered judgments or "intuitions" (though Rawls avoided to use the term "intuitions" in this context) about social cases or particular critical issues, the principles or rules that we believe ought to govern them, and the theoretical

[40] J. Rawls, *A Theory of Justice*, Cambridge, MA: Harvard University Press 1971. The term 'reflective equilibrium' was coined by Rawls and popularized in his *A Theory of Justice* as a method for arriving at the content of the principles of justice. Here I employ this method for arriving at the content of the concept of moral freedom. Specifically, Rawls argues that human beings have a "sense of justice" that is a source of both moral judgment and moral motivation. According to the Rawlsian definition, we begin with "considered judgments" that arise from the sense of justice. These may be judgments about general moral principles (of any level of generality) or specific moral cases. If our judgments conflict in some way, we proceed by adjusting our various beliefs until they are in "equilibrium", which is to say that they are stable, not in conflict, and provide consistent practical guidance. According to Rawls a set of moral beliefs in an ideal reflective equilibrium describes or characterizes the underlying principles of the human sense of justice. After Rawls, the method has been broadly defined as a process consisting in engaging with an issue of great importance, locating it within existing debates (when there are), considering it from the most relevant standpoints, and evaluate which angle of way to approach it is capable of shedding the most valuable light on it.

considerations that we believe bear on accepting these considered judgments, principles, or rules, revising any of these elements wherever necessary in order to achieve an acceptable coherence among them. This is specifically the philosophical methodology informing the moral reasoning underlying the present work.

The method of reflective equilibrium succeeds, and therefore we achieve a reflective equilibrium, when we arrive at an acceptable coherence among these beliefs. An acceptable coherence requires that our beliefs not only be consistent with one other (a weak requirement), but that some of these beliefs provide support or provide the best explanation for others. Moreover, in the process, we may not only modify prior beliefs but add new beliefs as well.[41] Indeed, as Schroeter pointed out[42], there need be no assurance the reflective equilibrium is stable – we may modify it as new elements arise in our thinking. For this reason, the inquiry on moral freedom needs to be constantly renovated with the continuous change of our world and the rise of new forms that can enhance or endanger it.

In practical contexts, as ours, insofar as the ethical inquiry on moral freedom I pursue is in the particular algorithmic connoted environment we reside today, this deliberation may help us to conclude about what we ought to do when we had not at all been sure earlier. According to Scanlon[43], we arrive at an optimal equilibrium when the component judgments, principles, and theories are ones we are un-inclined to revise any further because together they have the highest degree of acceptability or credibility for us.

Viewed most generally, a "reflective equilibrium" is the end-point of a deliberative process in which we reflect on and revise our beliefs about an area of inquiry. The inquiry might be as specific as the moral question, "What is the right thing to do in this case?" or the logical question, "Is this the correct inference to make?". Alternatively, the inquiry might be much more general, asking which theory or account – for example, in this case – of moral freedom we should accept, or which principles of inductive reasoning we should use, as the most capable of shedding the most valuable light on the issue considered. The process and method

[41] *Ibidem.*

[42] F. Schroeter, "Reflective Equilibrium and Anti-theory", *Noûs*, 38(1): 110-134 2004.

[43] T.M. Scanlon, "Rawls on Justification", in The Cambridge *Companion to Rawls*, S. Freeman (ed.), Cambridge: Cambridge University Press, 2002 pp. 139–167.

itself, also beyond the specific Rawlsian connotation[44], is generally called as the "method of reflective equilibrium".

The method of reflective equilibrium drives the overall conceptual operation that underlies my dissertation. Specifically, it clarifies how to address and frame an ethical inquiry based on a common-sense intuition, like those we have observing contemporary social phenomena, and like the one that inspired this specific work, i.e., a potential threat to our moral freedom posed by algorithmic technology. these intuitions need to be indeed philosophically argued, therefore, in a way that shows a level of consistency between our ideas and the observable social phenomenon.

The method hence highpoints how to philosophically frame and develop intuition-based questions about a given potential problem and how to tackle it by drawing insights from theories developed in philosophy, in this case, in moral philosophy, by evaluating them in the light of what can give the most consistent and fruitful understanding of the phenomenon considered.

If the methodology I adopt is now clear, let me clarify the thesis I claim and the goals that the present work aims at achieving.

The specific thesis I maintain is that the "algorithmic governance", that is structuring by design in our mature informational societies, is creating a new form of impediment, what I call *algorithmic predeterminism*, to our moral freedom, by silently undermining those necessary conditions (or *conditiones sine qua non*) which secure at a minimum threshold and so enable our freedom to choose and act as moral agents. I claim also that the elaboration of a new conceptual lens to understand our privacy, what I define *moral privacy*, by detecting a zone for the protection of our moral freedom, can provide a precious tool both to algorithms-based ICTs designers and tech policymakers to assess and recognize – case by case

[44] The method of reflective equilibrium is widely debated. It is important to consider that there are also alternative account of reflective equilibrium which retains the importance of revisability and emphasizes the positive role of examining our moral intuitions, but rejects the appeal to coherentism in favor a treating our intuitive moral judgments as the right sort to count as foundational, even if they are still defeasible. For example, see J. McMahan, "Moral Intuition", in *Blackwell Guide to Ethical Theory*, H. LaFollette (ed.), Oxford: Blackwell 2000 and P. Nichols, "Wide reflective equilibrium as a method of justification in bioethics", *Theoretical medicine and Bioethics*, 33(5) 2012, pp. 325-341.

– when our moral freedom risks to be compromised and therefore to act consequently to mitigate or avoid this risk.

The dissertation aims at two distinct, but profoundly intertwined, goals.

The first goal is to elaborate a normative conception of moral freedom that can allow us to make an adequate evaluative judgment about the current or potential impact of algorithm-based ICTs on our moral freedom as our freedom of choice and action as moral agents, and which also enables us to shed light on the practical implications raised by this impact on our moral freedom in the specific context of our contemporary informational societies.

The ethical-normative account proposed may offer something more than an abstract ethical principle that stands on its own, as it provides a broader conceptual and prescriptive guide to algorithms-based ICTs architects to rethink the design and evaluate the implementation of algorithms-based technologies. Indeed, one of the most severe critics in the field of ethics applied to technology is that in the numerous frameworks proposed so far the ethical principles picked out are too general and risk to be empty labels and so to be instrumentalized for unethical goals. [45] The

[45] One of the main problems in the field of ethics of ICTs and algorithms is indeed the use of ethical principles as empty labels, emptied from the philosophical richness which is proper to them. This unfruitful use has generated two main phenomena, what have been defined as "Ethics Washing" and "Ethics Bashing". Ethics indeed is increasingly used by companies as an acceptable façade that justifies deregulation, self-regulation or market driven governance, and is increasingly identified with technology companies' self-interested adoption of appearances of ethical behavior. This growing instrumentalization of ethical language by tech companies is called specifically "ethics washing". Beyond AI ethics councils, ethics washing includes other attempts at simplifying the value of ethical work, which often form part of a corporate communications strategy: the hiring of in-house moral philosophers who have little power to shape internal company policies; the focus on humane design – e.g. nudging users to reduce time spent on apps – instead of tackling the risks inherent in the existence of the products themselves. The technology community's criticism and scrutiny of instances of ethics washing often borders into the opposite fallacy, which we call "ethics bashing". This is a tendency, common amongst social scientists and non-philosophers, to trivialize "ethics" and "moral philosophy" by reducing more capacious forms of moral inquiry to the narrow conventional heuristics or misused corporate language they seek to criticize. Equating serious engagement in moral argument with the social and political dynamics within ethics boards, or understanding ethics as a political stance which is antithetic to – instead of complementary to – serious engagement in democratic decision-making, is a frequent and dangerous fallacy. The misunderstandings underlying ethics bashing are at least three-fold: (a) philosophy and "ethics" are seen as a communications strategy and as a form of cover-up or façade for unethical behavior, (b) the role and importance of moral philosophy is downplayed and portrayed as mere "ivory tower" intellectualization of complex problems that need to be dealt with in practice; and (c) philosophy is understood in opposition and as alternative to political representation and social organizing. See E.

conception of moral freedom that I elaborate does not want to add another principle to the long list already developed in the existing frameworks, but rather aims at informing with a normative account those ethical principles that, as described, should deal with the protection of our freedom of choice and action. This goal, i.e., the elaboration of this kind of normative conception, is one side of the coin. The other side coincides with the elaboration of a conceptual proposal capable of introducing in the tech field the protection of moral freedom.

The second goal is indeed that of elaborating a conception of privacy that can cover a zone that is still undetermined in the theories on privacy in the tech field, what I call *moral privacy*, namely, the protection of those criteria underlying our moral freedom. This further declination of the concept of privacy can provide a specific tool both to private regulative bodies and institutional policymakers to evaluate and regulate the deployment and the functioning of certain technologies in the light of their impact on our crucial freedom to form genuine moral identity.

Both these concepts, moral freedom and moral privacy as two sides of the same coin, can constitute a precious compass to navigate the complexity of a reality reshaped and increasingly governed by powerful and interconnected algorithmic choice-architectures.

To achieve these goals, the dissertation is articulated in three main chapters.

The first chapter focuses on the elaboration of an account of moral freedom that allows to recognize properly what it means to choose and act as moral agents, and therefore to adequately evaluate the impact of algorithms-based ICTs on our freedom to choose and act as moral agents in our informational societies.

The first section of the chapter aims to show the complexity of the concept of human freedom and the shortcomings in the huge and heterogeneous debate on it when it deals to distinguishing moral freedom from free will and from socio-political freedom. In doing so, in this section, I shed light on two concepts of moral freedom, a positive concept of moral freedom as freedom to become genuine moral agents and to form our moral identity, and a negative concept of moral freedom, as

Bietti. *From Ethics Washing to Ethics Bashing. A View on Tech Ethics from Within Moral Philosophy*. FAT Conference, Barcelona (Spain), January 2020, p. 2. The work proposed instead aims at showing how moral philosophy can fruitfully inform and enrich the debate on ethics and AI.

freedom from potential and actual, moral and immoral interferences to our capacity to choose and act as moral agents and so to become genuine moral agents.

The second section takes the cues from the analysis of some of the main theories in the free will debate and in the socio-political debate, insofar they concern the exercise of free choice and action in order to bring out what conditions, not sufficient but at least necessary, need to be guaranteed to make the exercise of our freedom of choice and action as moral agents possible. The process of detecting these necessary conditions underlying moral freedom presents both some overlaps and differences with some perspectives characterizing both the philosophical debate on free will and that one on socio-political freedom. Therefore, in this second section, I take the moves from these underlined conditions, i.e., "the availability of alternative options" and "human autonomy", and I revise them in the light of the account of positive and negative moral freedom proposed: I respectively define them as "the availability of morally heterogeneous options" and "moral autonomy".

Finally, in the third section, I clarify what is the ethical-normative value of our moral freedom, i.e., why it is a criterion that ought to be safeguarded to ensure the protection of the moral dimension of our agency and of our living together.

The second chapter of the dissertation, by opening the dialogue between moral philosophy and technology, questions whether the necessary conditions for the exercise of our moral freedom – detailed in chapter one – are safeguarded by ICTs. More specifically, I focus on the role played by the algorithmic functioning of our ICTs (e.g., Google Search, Facebook Newsfeed, Amazon's, Netflix's, and Spotify's recommendation systems, Siri's voices, GPS, etc.), inasmuch as algorithms-based ICTs seem to constitute what Cass Sunstein and Richard Thaler define as the "choice-architectures"[46] of the emerging hybrid environment where we reside, i.e., that reshapes the contexts in which we make our choices, perform our actions, and ultimately develop our identities. Thus, the first section of the chapter is devoted to the exploration of the "algorithmic governance" in our informational societies in order to clarify what I specifically define as *algorithmic*

[46] R. Thaler, & C. Sunstein, *Nudge: Improving decisions about health, wealth and happiness*. London (UK): Penguin 2009.

choice-architectures, by analyzing their peculiar main features and their pluralistic intertwined functions.

The second and third sections develop the core of my thesis and argue the double-level impact of algorithmic architectures on our moral freedom. More specifically, the second section analyzes the *epistemological* reshaping impact produced by the inter-play of two diverse types of algorithms (i.e., profiling and filtering algorithms) on the first condition detected as necessary for the exercise of moral freedom, i.e., the availability of morally heterogeneous options, which in turn implies a morally heterogeneous exposure of the subject in social relations and moral values so as to be able to form her own idea of good, along with moral values and moral ground projects towards which steering her actions, behavior, and finally, the development of her moral identity. The third section, in turn, analyzes the deep *ontological* reshaping impact of algorithmic systems specifically produced by algorithmic recommendation systems (RS) on the second necessary condition for the exercise of our moral freedom, that is our moral autonomy, intended as the relational-determination of our choices and actions that we express in the dimension of intrinsic endorsement, i.e., in the way in which we endorse and actualize morally heterogeneous values – that we develop through the morally heterogenous exposure in relations and values as requested by the first conditions – into our action and behavior and by doing so we develop as moral agents the authorship of our moral identity.

The third chapter tries to detect the risk of a new form of impediment to our moral freedom and aims at providing conceptual tools to prevent or mitigate it.

The first section of the chapter re-elaborates the argumentation provided in the second chapter and explores the formation of what I define as *algorithmic predeterminism*, that is the algorithmic capacity to predict the moral elements driving our choices and so reshaping them according to pre-determined goals in a way that can suspend our reflective endorsement and therefore undermine our moral freedom. I analyze the extent of this algorithmic impact and its actual and potential consequences in diverse societal domains where algorithms are broadly applied, ranging from the sector of healthcare to that of justice.

The second section shows how the philosophical and legal debate in the technological field which deals with the protection of our freedom through the theory of informational privacy, although providing meaningful lens to start to understand the tool of privacy in a way that can protect our freedom broadly conceived, still lacks to consider our moral freedom, and specifically, to address and provide tools to tackle the phenomenon of algorithmic predeterminism. Specifically, this section sheds light on how to fulfill this gap, namely, the absence of a specific privacy lens to frame, understand, and then mitigate the risk posed by the predetermining potential of algorithmic ICTs on our moral freedom, through the definition of a novel lens to develop a conception of privacy for the protection of freedom which considers and includes our moral freedom. I will define this lens as *moral privacy*, to indicate, by following the criteria underlined in the first chapter, the specific zone that must be protected in the algorithmic societies to secure our free choosing and agency as moral agents at a least at a minimum threshold.

The third and last section is focused on recognizing which specific social agents (institutional and technological) need to be called into action when it comes to applying the conceptual tools developed to prevent jeopardies to the free exercise of our choices and actions, the ones I will define as the *champions of moral freedom*.

Finally, some caveats.

Firstly, before I get underway, I need to clarify the terminology I use in this work. Indeed, the term 'moral freedom' has been used over time with diverse shades of meaning. On one side, and mostly, as a synonym of 'free will', hence to indicate precisely the 'free nature of the will'[47]; on the other side, as a substitute for 'freedom to fall'[48], and hence to denote our 'freedom to do good or evil'. In this work, I use the term 'moral freedom' as previously defined to indicate our freedom to choose

[47] For a clear and wide use of the term 'moral freedom' as a synonym of 'free will', consider the work of N. Hartmann, *Moral Freedom*, The MacMillan Company, New York 1932.

[48] For instance, in the debate on freedom and moral enhancement. For a recent example: J. Danaher, "Moral Enhancement and Moral Freedom: A Critique of the Little Alex Problem", in *Royal Institute of Philosophy Supplement*, 83 (10) 2018, pp. 233-250. For a wider understanding of the context that underlie this specific use of the term, see J. Harris, "Moral Enhancement and Freedom", in *Bioethics* 25 2011, pp. 102-111; and also I. Persson and J. Savulescu, "Moral Bioenhancement, Freedom and Reason", in *Bioethics*, 9 2016, pp. 263-268.

and act as moral agents, and specifically, as genuine moral agents, namely, freedom to form our own idea of goods, values, and moral ground projects and with them our moral identity (or moral posture).

This use means that I will not carry on an inquiry on the free or unfree nature of our will, as I focus my analysis on the conditions of possibility for the exercise of our freedom of choice and action for what concerns its moral dimension. Nevertheless, there are many overlaps in the criteria to assess the 'free nature' of the will and those to evaluate our freedom of choice and action as moral agents. I do not reconstruct the philosophical debate on the free nature of the will, but I cannot avoid mentioning some standpoints of its main exponents, just insofar they are pertinent to the understanding of the conditions enabling the exercise of the free choice. A similar warning is valid also for the academic debate on socio-political freedom. Its deeper analysis is beyond the scope of the present work, but some references to its champions will be unavoidable in our inquiry into moral freedom.

Secondly, although the issue of moral freedom is strictly linked to the issues of moral responsibility and social justice, which deserve as well to be rethought in the light of the algorithmic governance, this further conceptual inquiry – for reasons of space and internal coherence – does not fall within the scope of my analysis.

Thirdly, the issue of moral freedom is a topic of global interest, inasmuch as it concerns a common human dimension, that one of agency, which goes beyond geographic boundaries. Nevertheless, I limit my inquiry on moral freedom and its threats to advanced informational societies, those where the algorithmic governance is highly perceivable. Although aware of both the intranational and global digital divide which nonetheless characterize also our informational societies worldwide, as well as aware of the diverse geopolitical orientations which connote them, I will not explicitly weigh the role of these social, political, and geographical variants in the inquiry I carry on.

First Chapter

On Moral Freedom: A Philosophical Inquiry

Jeroen van den Hoven, one of the pioneers in the field of moral philosophy applied to digital ethics, underlined years ago that there is no other way for moral thinking in the field of ICTs than pursuing a conceptual inquiry into the specific ethical key-concepts that play a crucial role in the description, analysis, and evaluation of ICTs in order to inform their design, their shaping, and their development in an ethically meaningful way.[49] This chapter aims at laying the conceptual foundations of the present ethical inquiry by drawing insights from theories developed in moral philosophy to bring out an account of freedom capable to shed light and then adequately evaluate the deepest impact of algorithmic ICTs on our lives: the impact of algorithmic systems on what I define our moral freedom.

But what do I mean by moral freedom?

In the first section, I try to navigate the complexity of the multidimensional concept of freedom and the countless theories elaborated to define the topic in the philosophical field by shedding light on a novel concept of freedom that has not been sufficiently explored so far, neither specifically framed and addressed per se: what I conceptualize as moral freedom. Moral freedom concerns the dimension of practical agency and more specifically of practical choice but it does not coincide with either free will or socio-political freedom, though it partially overlaps with both of them. Therefore, by highlighting the shortcomings and overlaps in the current debate when it comes to defining the concept of moral freedom, I elaborate a specific account of it, and show why this account of moral freedom can allow us to understand our freedom to choose in our contemporary societies, in order words, to understand how we – as moral agents – act in the world, and therefore, to adequately evaluate later the impact of algorithmic ICTs on our agency and lives.

[49] J. Van den Hoven, "The use of normative theories in computer ethics" in *The Cambridge Handbook of Information and Computer Ethics* (ed. L. Floridi), Cambridge University Press, Cambridge 2010, p. 61.

In the second section of this chapter, I further define the account of moral freedom proposed by clarifying two key elements that are necessary to meet our conception of moral freedom, what I define as the *conditiones sine qua non for the exercise* of our moral freedom.[50] To do so, I take the moves from the critical analysis of the philosophical accounts that in the debates on free will and socio-political freedom mostly discuss the main conditions required to exercise our freedom of choice and action, only insofar as they are valuable to understand in which specific conditions our agency can be defined as morally free; the conditions detected are respectively "the availability of alternative options" and "human autonomy". In pursuing this ethical inquiry, I revise these necessary conditions and bring out the conditiones sine qua non underlying the exercise at a minimum threshold of our moral freedom, what I re-conceptualize as: the "availability of morally heterogeneous alternative options" and our "moral autonomy".

Finally, after having unpacked the concept of moral freedom and clarified the *conditiones sine qua non* for its exercise, in the third section, I shed light on the ethical normative dimension[51] of our moral freedom, and show why the latter is a normative value that ought to be imperatively secured and promoted for the protection of the moral dimension of our agency and of our living together.

I.1 Navigate the Complexity: From Freedom to Moral Freedom

Reasoning about the philosophical concept of freedom is a highly complex task, as many are the forms and specific declinations that freedom can assume. It is fundamental to distinguish the peculiar valence of the concept of freedom when we discuss the issue of free will from that it assumes when we refer to social and

[50] Conditions that are necessary, even though not sufficient on their own, to guarantee at a minimum threshold the exercise of our moral freedom.

[51] The introduction of this dimension will provide a normative guidance in the last chapter to clarify how we should think about the design of algorithmic ICTs in a way that safeguard the moral dimension of our choice, agency, and living together.

political freedom: in these two cases two different kinds of philosophical problems are at stake.[52]

Whilst both conceptions of freedom relate to the dimension of choice and agency, in the former, the peculiar valence the concept of freedom assumes is metaphysical, and freedom of choice is grounded on the free or unfree nature of the will[53]; instead, in the latter, the peculiar valence of the concept is socio-political, namely, the focus is on what it takes to live freely within a particular socio-political order and culture, and freedom of choice is declined also in reference to many forms of coercion that are politically or socially exercisable. In this case, our freedom of choice and action is considered beyond the free or unfree nature of the will, and regards the possibility to make or not certain religious, political, or social choices, such as choosing what to believe or with whom to associate, without being socially persecuted, discriminated or politically punished for that choice.[54]

The notion of moral freedom emerges in both debates on freedom, but it is less explored, and never conceptualized per se. The purpose of this first section is to fill this gap, untie the use of the concept of moral freedom as a mere synonym of free will, as well as shed light on its valence also in the liberal tradition of the social and political debate.

When it comes to analyzing the debate on freedom of choice as freedom of the will, moral philosophers very often failed to distinguish the concept of moral

[52] Even if there are examples of unitary treatises of these two concepts of freedom, as that offered by Thomas Hobbes in the *Leviathan* (1651).

[53] There is a long-standing debate that has engaged many disciplines from moral philosophy to moral psychology, neuroscience, and biology on the 'real possibility' of free will and the related question of whether human choices and actions are causally determined and whether this prevents individuals from having free will. For an historical reconstruction of the debate on the metaphysical nature of "free will" and the different causal theories that have been detected over time as threats to this freedom, from divine foreknowledge to biological, social, and physical determinism, see M. Mori, *Libertà, necessità, determinismo*, Il Mulino, Bologna 2001. For a critical analysis of the philosophical debate on free will and the main four different philosophical families (i.e., *libertarianism*, *compatibilism*, *agnosticism*, and *illusionism*), consider the analysis provided by M. De Caro, *Il libero arbitrio. Una Introduzione*, Laterza Editori, Roma-Bari 2004.

[54] To give more examples: can you be free if the government imposes sanctions on you for following your conscience on religious matters? Can you be free if your workplace, college or school has a speech code that prevents you from saying and doing certain things? These kinds of questions are central to the socio-political tradition of liberalism (broadly conceived).

freedom from that of free will[55], by leading the succeeding discussion to treat the two concepts as not more than mere synonyms and so to turn off the focus on moral freedom per se. However, within the debate on free will, three "exceptions" can be identified as attempts to distinguish free will from moral freedom, even if they have not been developed in full extent.

The first attempt to distinguish moral freedom from free will can be found in Gabbert[56], who opens his *Moral Freedom* by declaring his intention to unpack this key concept. He claims that our moral freedom is a unique thing and if we want to understand it we need to clarify what we mean by moral and by free. He claims that what can define our actions as moral does not lie in the "free-characteristic" chance of human will, i.e., its being or not free by nature, but in the fact that our actions stand for a system of behavior (and specifically of values) that we can or cannot approve (endorsement). Only from this latter standpoint, we can morally appraise our actions, according to Gabbert, and not from the chance of their being

[55] Even before, the majority of philosophers in the debate on free will fail to distinguish freedom of choice and action from freedom of the will. This is an important distinction as, beyond the difficulty to understand whether the nature of our will is free or unfree, we can reckon that we exercise a certain kind or degree of practical freedom in our empirical context, as it has been pointed out by T. Hobbes, Of Liberty and Necessity in *Hobbes and Bramhall on Liberty and Necessity* (ed. By V. Chappell), Cambridge University Press, Cambridge 1654 [1999], pp. 15-42. R. Albritton, Freedom of Will and Freedom of Action in *Proceedings and Addresses of the American Philosophical Association*, Vol. 59, No. 2 (Nov., 1985), pp. 239-251. See also the distinction provided by I. Kant in the *First Critique* (1787) between "transcendental freedom" (i.e., free will *stricto sensu*) and freedom in a *practical respect* (that could be known by experience, as freedom to act – direct your conduct – for the sake of more remote sensible goods). On this point, I. Kant, *Critique of Pure Reason* (trad. and ed. by P. Guyer and A. W. Wood), Cambridge University Press, Cambridge 1998. This distinction between freedom of choice and action and free will is today time-needed, as the use itself of the term of free will is becoming widely contested. The term indeed is today considered misleading for at least three reasons. Firstly, the term "free will" in its original sense as capacity of choice according to the will has been recognized as fallen in disuse. Secondly, it is widely noticed the theoretical obsolescence also of the concept itself of "will", whose specific adequacy has been highly contested both in the field of philosophy of mind and in that one of the cognitive sciences. Lastly, the problem of freedom, in its more abstract sense, concerns as much the freedom of choice as the freedom to act, and many philosophers in the debate considers absurd the idea itself of a "free will" literally intended, as they retain that freedom can be predicable just for human actions. With this gradual fall in disuse of the term, philosophers will be called to reason whether it is necessary to distinguish free will from freedom of choice or to start to use the latter as a synonym of the former. To expand this issue, see M. De Caro, *Il libero arbitrio. Una Introduzione*, Laterza Editori, Roma-Bari 2004.

[56] M.R. Gabbert, "Moral Freedom", *The Journal of Philosophy*, 24 (17), 1927, pp. 464-472.

free or unfree by the nature of the will. However, Gabbert does not go further in exploring this distinction, and instead privileges the analysis of the issue of freedom of choice and unpredictability. To sum up, the system of values that we can approve or not via our choices and actions is what according to Gabbert morally connotes our agency, i.e., what makes it morally significant and evaluable, and therefore is the dimension where we should look for the expression of our moral freedom.

In a similar direction, we can find a second attempt, carried out by Myrton Frye[57]. In his *Moral Freedom and Power*, Myrton Frye distinguishes moral freedom from intellectual freedom, with the former referring to the conduct, while the latter to thought. Specifically, the author further defines moral freedom in a way that is not fully traceable directly back to the free or unfree nature of the will in the definition of moral freedom as "the subject's power to approve or disapprove the actions that she chooses to perform"[58]. He says even more about moral freedom. This freedom is defined by a specific moral feature, what he claims to be the moral criticism: "what a person does, she may be called upon to justify, and this means to show that the action she performed was the act she *ought to* have done"[59]. The moral criticism that connotes morally our freedom seems to bring out, beyond the dimension of subject's approval or disapproval of choices and actions, also previously mentioned, another dimension, the dimension of *ought to*, that is the Kantian-inspired deontological one. This dimension concerns what is normative for us, or better, what (e.g., value, reason, principle…) we choose and endorse as normative for our conduct. Myrton Frye does not proceed further in this analysis, as he ends to focus on the metaphysical limitations to this power. Nevertheless, from his contribution an important moral feature of our agency emerges, that is the one of obligation, i.e., to be able to oblige ourselves and specifically our conduct on the basis of the values (or principles, reasons, and so forth) we approve and endorse, and therefore to look at it for the understanding of our moral freedom in a not-coincident way to free will, as the freedom to develop our *ought to*, that is a specific moral feature of our agency.

[57] A. Myrton Frye. "Moral Freedom and Power". The Journal of Philosophy, 28 (10), 1931, pp. 253-260.

[58] *Ivi*. p. 254

[59] *Ibidem*.

The third and last attempt is that of Hartmann, who too, in the same direction of the previous two, frames the problem of moral freedom as the problem of the freedom of the will (or the metaphysic of morals). Nevertheless, in his *Moral Freedom* [60], he highlights how the concept of moral freedom cannot just concern the freedom of the "will" as literally understood, as the concept of freedom of the will is too narrow and does not emphasize the key-dimension of moral values that steer our choices and actions. Although he devoted his work to the argument against the impossibility of free will, Hartmann tries to define further moral freedom. Specifically, he sheds light on moral freedom as a capacity or "individual's decisional power to transform values from potential to actual"[61]. The strong metaphysical valence of Hartmann's definition of moral freedom is perceptible, anyway, it helps us again to shed light on a key moral dimension of agency, that one of values, that is intrinsically morally-connoted, and above all, again, on the power of approving or endorsing values, as a specific capacity of the individual to actualize values into her choices and actions.

The moral dimension of approval (or endorsement) of the subjects of their choices and actions, that one of *ought to*, and that one of values are strictly related, as we will see in a while, and can help us to conceptualize moral freedom per se.

I will not go further in the debate on free will for now, because, as I anticipated in the introductory caveats, my analysis on moral freedom is not aimed at investigating the free or unfree nature of the will[62] through a metaphysical inquiry; however the free will debate is relevant for my analysis insofar as it allows to bring out the key characteristics that allow to define human agency as moral and then elaborate an account of moral freedom to adequately understand the moral dimension of our choice and agency in the current world.

[60] N. Hartmann, *Moral Freedom*, The MacMillan Company, New York 1932.

[61] *Ibidem*.

[62] Another reason behind the choice to not adopt the concept of free will to understand our moral freedom in our contemporary societies is that it concerns metaphysics and the nature of our will, while I am interested in the kind of moral freedom we can experience in our contemporary societies. The concept and the related debate are anyway extremely useful to understand freedom of choice and action as a result of what it takes to the will to be defined as free.

Basing on the connoting aspects of moral freedom emerging from the debate on free will, I can so provide my very first definition of moral freedom.

Moral freedom distinguishes from free will as what is in power and what is in act. Since free will concerns the (free or unfree) state-of-the-nature of the will, whether it is free or unfree this is a characteristic that is already in act; this means that whether we think correctly or incorrectly that we have it or not, however, we have no power to change or interfere with it, as it pertains to the metaphysical order of our reality. Moral freedom instead concerns our freedom to express our moral agency, i.e., our power to become moral agents.

Let me better define it.

Moral freedom is our freedom to become moral agents, specifically, genuinely moral agents, that means to be able to develop moral reasons, values, and moral ground projects in a genuine way – i.e., via the exposition to heterogeneous relations, attachments, practices, and so that are not developed under the force of external impositions or determination (for example, to comply with the authority and the law) – and to actualize them via our choices and actions.

Moral freedom is our freedom and power to be(come) moral,
and specifically, genuine moral agents.

Here someone might say that we are always moral beings and therefore this power cannot be undermined. However, we are not referring here to our essence as human beings, but we refer to how we act as agents in the world. Even if morality is intrinsic aspect to our being humans (that is an ontological question that I do not go into, as it is outside the scope of my analysis), the thesis I support here is that we are not always moral agents, i.e., subjects that act as moral agents.

Here at least two other objections could be raised: the first could ask whether for example children, having not yet developed an idea of good and evil, can be defined as moral agents. My answer is linked to the previous point, we do not take into consideration here the moral standing of the individuals, but when their actions can be specifically characterized as moral. The second point can be raised by those who advocate a consequentialist ethical paradigm. They may argue that the morality of our actions depends on the consequences in terms of good or evil they produce.

So, given that my actions can always be good or bad on the basis of the effects they produce in the world, our actions can always be considered as moral – thus, without considering the level of intentions, to recall instead another ethical paradigm, that is, the Kantian matrix deontological one. Here, what is crucial to underline is that the account of moral freedom I propose is in line with the ethical theories of moral constructivism, which see the morality of the subject in the internal dimension of intentions (that are defined with a wide array of terms from reasons to mental states). To avoid to enter into a metaethical analysis, I just underline that I rarely use the term 'intentions', and privilege in order to refer to the internal adhesion (or not adhesion) of the subject to values, reasons or principles, what many exponents of constructivism define as "reflective endorsement"[63] of the values that inform our choices. In this approval or reflective endorsement of certain values we develop the obligation, our ought to, that is, how we actualize those values into our choices and actions by obliging ourselves to them and building over time our moral identity.

To sum up and unpack our concept of moral freedom:

I. Moral freedom *is our freedom and power to be(come) moral, and specifically, genuine moral agents.*

ii. *Freedom to become genuine moral agents* means *freedom to develop genuine moral identities*.

iii. *Freedom to develop genuine moral identities* means *freedom to develop genuine moral reasons, values, and ground-projects* that inform my choices and actions.

iv**.** *Freedom to develop genuine moral reasons, values, and ground-projects* means *freedom to choose and act in accordance*

[63] See, for example, C.M. Korsgaard, C.M. *The Sources of Normativity*. Cambridge (UK): Cambridge University Press 1996.

with reasons, values, and moral ground-projects I endorse (and by doing so developing *my ought to*)

By paraphrasing Berlin[64], this is a *positive concept* of moral freedom: moral freedom as *freedom and power to* construct, build my own moral projects, and with them, build and affirm my moral identity, according to reasons and values that I can genuinely develop and endorse (through the approval or endorsement I develop my ought to), and where their moral genuineness rely on my approval and the level of moral heterogeneity of values, relations, attachments to which I can be exposed.

This is just a part of the account of moral freedom here at play. There is also indeed a "negative concept" of moral freedom.

Indeed, as it has been anticipated before, the moral dimension of freedom is also called into question within another debate on freedom of choice, though never analyzed per se, the debate that analyzes freedom of choice at a socio-political level, namely, that one that asks broadly speaking what it takes to be free in a particular social and political order.

To shed light on this negative sense of moral freedom, a helpful clue is provided by List and Valentini[65], insofar they stress specifically a moral dimension of freedom neglected in the socio-political debate on freedom thus far. They clearly admit to restrict their discussion just to socio-political freedom, setting aside what we previously discussed as the metaphysical debate on free will; as well as, to focus on what Berlin calls a "negative sense" of freedom, and so setting aside from freedom understood in a positive sense.

Their analysis is very relevant for our inquiry at least for three reasons: first, they discuss freedom as a concept, and not its measurement, that instead is very common around theories focusing on freedom of choice in the social and political debate; second, they acknowledge that facts on freedom matter morally and that also moral constraints on freedom require justification[66]; third, they shed light on

[64] I. Berlin (1969). *Two Concepts of Freedom*. Oxford (UK): Oxford University Press.

[65] C. List & L. Valentini. "Freedom as Independence." *Ethics* 126, no 4 (2016), pp. 1043-1074.

[66] In simple words, the right to ask on what grounds are you restricting my freedom.

an unexplored lens that can go beyond diatribes between the liberals and the republicans[67], that are out of the interest of the present work.

List and Valentini move from a basic liberal definition of negative freedom according to which "an agent's freedom to do something is the *actual* absence of *relevant* constraints on the agent's doing something"[68] to further develop it through the sift of the moralization question[69], i.e., whether the constraint-absence condition can be qualified by some moralized exemption clause according to which morally permissible constraints (for example, "non-arbitrary" or "just" ones) do not count as freedom-restricting. They argue that the concept of social freedom must involve a "robustness requirement", according to which the mere possibility of the imposition of constraints, even if not still actualized, may restrict an agent's

[67] I will give an overview of the debate on socio-political freedom in the next chapter, as it will be relevant to bring out the necessary conditions underlying our moral freedom, for now, it can be enough to underline how the much recent philosophical work on socio-political freedom revolves around debates between liberals and republicans. Liberals, following Isaiah Berlin, define freedom as the absence of constraints on action, where the constraints that matter can be spelt out in various ways. Republicans, especially in Philip Pettit's influential interpretation, instead argue that freedom requires non-domination: the guaranteed or robust absence of arbitrary constraints.

[68] The specific reference is the liberal account in the tradition of Isaiah Berlin, according to which freedom is the *actual* absence of relevant constraints. I. Berlin (1969). *Two Concepts of Freedom.* Oxford (UK): Oxford University Press. Nevertheless, critics have argued that this focus on actual constraints, or possibilities of action present in the actual world, has problematic implications. Consider the often-cited case of a slave with a non-interfering master. In this hypothetical scenario, the master could in principle interfere with the slave's actions (e.g., under the legal institution of slavery as it existed in the United States prior to 1865), but it so happens that the master refrains from interfering, and many actions are actually open to the slave. On the liberal conception, the slave would count as free to perform these actions – a conclusion that many find unsatisfactory in light of the paradigmatically unfree status associated with slavery. Furthermore, whatever our ordinary-language intuitions about the case may be, the slave is certainly subject to modal constraints on action that—to understate things dramatically – stand in need of justification. The master's power to interfere is enough to raise a justificatory burden. To address this problem with the liberal conception, republicans suggest, we need to move to a conception of freedom that demands robustness, so that freedom can already be undermined by the mere possibility of constraints, suitably interpreted, i.e., by the possibility of actions-being-rendered-impossible. In our example, the master can make use of such a possibility at any time, even if he does not currently do so. On a robust conception of freedom, an agent is free to do X only if she enjoys the absence of constraints, both in the actual world and in other nearby possible worlds. This can clearly account for the situation of the slave, since his status makes him susceptible to being constrained by his master, independently of whether the master actually exercises this power. To expand this discussion see C. List & L. Valentini. "Freedom as Independence." *Ethics* 126, no 4 (2016), p. 10.

[69] C. List & L. Valentini. "Freedom as Independence." *Ethics* 126, no 4 (2016), p. 4.

freedom[70]. In doing so, the concept of negative freedom is clarified as "an agent's freedom to do something is the robust (*actual and potential*) absence of constraints *simpliciter* on the agent's doing something", where by constraint they widely argue and specify the robust absence of both *moral* and *immoral* interference. This means that according to List and Valentini, freedom of choice in a socio-political context is from either immoral interference, whereby immoral interference they refer those that can be classified as morally unjust, wrongful, and illegitimate; but freedom of choice is also from moral interferences, namely, from those interferences that can be defined as moral as they can be "morally justified" with the criterion of tracking and respecting agents' interests. List and Valentini argue specifically that even those interferences that can be defined as moral, as justified by a moral explanation showing that they are morally permissible interferences (for example, inasmuch as developed by taking into account the interests or dispositions of the agents), are freedom-restricting and therefore no clause within the definition of freedom can exempt them. The difficulty of faithfully considering and reading the interests and dispositions of the subjects, which is not always made explicit, leads to request not only a moral justification also for the interferences that can be defined as moral, but to always to consider them (also when morally justified) as freedom-conditioning.

We can clarify this last point with an example: the institutional order to stay home during the first phase of the pandemic has generated a huge debate on the constitutional or unconstitutional nature of that act. From a moral standpoint, for many, the ban imposed is not morally unjust, i.e., unmoral, as it has been morally justified by national security due to the containment of the spread of the virus. Even if is a moral interference, this cannot be considered as a non-limitation to freedom, and so blindly accepted. That remains from a conceptual standpoint an interference

[70] To this point, they agree with Pettit's republican condition. Nevertheless, republican conception of freedom departs from the liberal one on a second dimension too: freedom is defined as *non-domination*, the robust absence of *arbitrary* constraints. Arbitrariness is interpreted as a moralized notion: something is arbitrary if it is unjust, illegitimate, capricious, or not governed by the right principles. Presumably, a slave is unfree because his master can constrain him arbitrarily, i.e., unjustly, illegitimately, and without taking seriously the slave's interests. The republican conception contrary to ordinary language use, re-classifies just, legitimate, and non-arbitrary restrictions of freedom as no restrictions of freedom at all.

to our freedom. This clarification provides to decision-makers, that are in charge to translate the concept into policies, a lens to distinguish when something X does not constitute a limitation to freedom, when X is an immoral interference or wrongful limitation to our freedom (and so requires to be eliminated or fixed), and when X is a moral interference (that can be accepted to a certain extent) but is anyway a limitation and therefore it needs to be morally justified and regulated case by case.

I cannot reconstruct the steps of the wide argumentation provided by List and Valentini, because for their extent and complexity it would easily lead us astray. Anyway, this argument's worth is to sheds light on a negative moral dimension of freedom in the socio-political debate and so to provide us a lens through which we can highlight a *negative concept* of moral freedom:

ii. *Moral Freedom* as freedom from *potential* and *actual* and *moral* and *immoral interferences* on our power to become genuinely moral agents.

To further specify this negative concept of moral freedom:

ii. *Freedom from unmoral or illegitimate interference* (that therefore must be fixed or removed by default, as morally illegitimate or wrong and thus unacceptable);

iii. *Freedom from moral interference* (interference always requiring a moral justification, i.e., that the reasons behind the limitation will be disclosed, and that when is possible should be subjected a regulation or mitigation, inasmuch as this limitation always interfere with the possibility for the agents to choose and act according to their own reasons and values).[71]

[71] An example can make this clearer: even whether you can interfere with my political choice by admitting just TV channels with a certain political orientation, that express also morally preferable ideas, this interference with my exposition to political ideas needs to be motivated and regulated, as it affects the way in which I develop my political orientation in a substantial way. This does not mean that this kind of interferences can be avoided, but they need to be acknowledged and publicly

This negative concept of moral freedom is important because it clarifies how the freedom to develop our moral identity is also expressed in immoral and moral non-interference. Obviously, the latter is more controversial, above all, when there are cases in which justifiable moral interferences can be essential to maintain or pursue public order or social good.

In this regard, indeed, it is important to clarify that these two concepts of moral freedom do not want to present a black-and-white scenario, where if there is interference our moral freedom is canceled, and vice versa, to fully enjoy it we need a state of absolute social non-interference. These two concepts of moral freedom show us that moral freedom is a multidimensional concept and thinking of it in the contemporary collective context involves trade-offs, i.e., situations in which we enjoy more positive freedom, others in which we enjoy more freedom in a negative sense, wherein the latter, some immoral interferences can be fixed, and the same for moral interferences, some of them can be mitigated, while others just regulated. However, these two concepts are dynamic and have a regulatory function, namely, they provide us with normative ideals and lenses through which to understand and then evaluate our moral freedom in our advanced informational societies.

I believe that in our globalized and multicultural information societies, this account of moral freedom is the soundest to adequately understand the meaning and the value of our choosing and agency as moral agents, both for the development of our moral identity, and the flourishing of the moral dimension of our living together.

I will devote more space to the value of our moral freedom in the concluding section of this first chapter. To complete our account on our moral freedom, indeed, it is necessary to identify what are the necessary conditions underlying its exercise.

I.2 The 'Conditions of Possibility' to Exercise Moral Freedom

In the foregoing chapter, I highlighted a positive and negative concept of moral freedom, and claimed that both help us to elaborate a sound account of moral

made clear, as to provide a specific meta-information that can be used by the subject in her moral reasoning to decide what she will endorse as reasons in her decision-making process for their action.

freedom. In this section, I try to shed light on which are the necessary conditions to guarantee at a minimum threshold our moral freedom, in both this positive and negative sense[72]. To do so, I take the moves from the free will debate and the socio-political debate on freedom of choice in order to see whether they examine the necessary conditions underlying the exercise of freedom of choice and action, whether there is a degree of convergence between these two debates on the issue, and whether this analysis can help us to bring out and develop the necessary conditions underlying our freedom to choose and act as moral agents[73].

In the long-standing debate on free will, the necessary conditions underlying freedom of choice and action have been widely discussed, by welcoming different formulations. Nevertheless, two conditions are largely acknowledged[74]: what the medieval philosophy had already defined as a) *libertas indifferentiae*, i.e., an agent is free to the extent that she can do and even not perform a certain action, namely, if and only if she can act otherwise from what the facto she does, and b) *libertas spontaneitatis*, i.e., an agent is free to the extent she can do what she wants to do, that is, if and only if in her agency she self-determine herself.

These two medieval terms have survived in some modern authors like Descartes, Hobbes, and Hume, and today they are widely used in debate on freedom of choice as freedom of the will to denote the *conditiones sine qua non* (therefore, necessary but not sufficient)[75] underlying the possibility of our freedom of choice and action.

[72] The highlighting of these necessary conditions will provide us an evaluation field to analyze and assess the impact of algorithmic technology on our moral freedom – inquiry at the core of the second chapter of the dissertation.

[73] For the sake of our specific purpose, I will not reconstruct these two far-reaching debates, as this operation would lead us far away from our account of moral freedom. This means that, for example, for the free will debate, the investigation on the diverse accounts inside the philosophical tradition that aim to prove or confute from experience, a priori reflection, and various scientific findings whether humans have "free will" (or whether is reasonable to believe that they have it) is set aside from this work. I will not engage my analysis in evaluating or taking a philosophical position in the wide array of compatibilist, incompatibilist or libertarian accounts proposed to ground the nature of free will. The analysis proposed here does not want to enter, debate or add something more to the already huge and analytically highly sub-structured discussion about the controversial existence and the kind of nature of freedom of the will. As I outlined before, this is a metaphysical side of the coin.

[74] M. De Caro, *Il libero arbitrio. Una Introduzione*, Laterza Editori, Roma-Bari 2004.

[75] Another and more controversial question is whether these conditions together are also sufficient conditions for realizing freedom of choice and action.

Indeed, it is clear that if an agent is free to perform a certain action, she will be also free to do not perform it. At the same time, it is also intuitively clear that to guarantee this freedom, the choice between potential courses of action cannot be casual, namely, it cannot result from factors out of control of the agent (as when, for instance, an agent decides whether to make a certain action on the basis of the toss of a coin, or when she is coerced to make a certain choice).

These terms have been substituted over time. At the current state of the art of the free will debate, the two *conditiones sine qua non* underlying our freedom of choice and action widely, although not universally, acknowledged are defined as follows:

1) the *availability of alternative options* in the decision-making process (i.e., the possibility for the agent to do otherwise from what *de facto* she does)[76];

2) and *human autonomy* over the decision-making process (i.e., the condition of the agent to be in control of her own choices, or at least the condition in which she can participate in a relevant way to the process that leads to the actualization of a certain course of action instead of one another)[77].

[76] There is a wide and technical debate about the availability of alternative options (also called as PAP: principle of alternative possibilities) in the free will debate, and specifically, in relation to moral responsibility. The technical analysis of it is however beyond of the scope of the present analysis. To expand the condition of availability of alternative options (or freedom to do otherwise) in free will debate, see J. Edwards. *Freedom of Will*, ed. Paul Ramsey, New Haven: Yale University Press 1754 [1957]. R. Holton. *Willing, Wanting, Waiting*, New York: Oxford University Press 2009; R. Reid, *Essays on the Active Powers of the Human Mind*, ed. Baruch Brody, Cambridge, MA: MIT Press 1788 [1969], and H. G. Frankfurt, "Alternate Possibilities and Moral Responsibility", *The Journal of Philosophy*, 66(23) 1969, pp. 829–839. Reprinted in Fischer 1986, pp. 143–52; in Frankfurt 1988, pp. 1–10; and in Widerker and McKenna 2003, pp. 17–25, and also H. G. Frankfurt, "What We Are Morally Responsible For", in L.S. Cauman et al. (eds.), *How Many Questions? Essays in Honor of Sidney Morgenbesser*, Indianapolis, IN: Hackett Publishing Company, 1983, pp. 321–335. Reprinted in Frankfurt 1988, pp. 95–103 and in Fischer and Ravizza 1993, pp. 286–295.

[77] The condition of autonomy in the free will debate is even more debated than the PAP in relation to the free or unfree nature of the will. Although discussing the diverse formulations of autonomy as self-determination is beyond the scope of the present analysis, precious references to expand the issue are P. Thomas. *Self-determination: The Ethics of Action*, volume 1, Oxford: Oxford University Press 2017. B. Michael, "Planning Agency, Autonomous Agency," in *Personal Autonomy*, ed. James Stacey Taylor, New York: Cambridge University Press 2005. L. Ekstrom. "A Coherence Theory of Autonomy," *Philosophy and Phenomenological Research*, 53 1993, pp. 599-616. H. Frankfurt. "Autonomy, Necessity, and Love," in *Vernunftbegriffe in der Moderne: Stuttgarter*

The first condition, i.e., the availability of alternative options, is usually denoted in the free will debate as the subject's power to do otherwise, while the second, i.e., human autonomy, is usually described in terms of power of self-determination, that is, the subject's power to choose according to reasons and motives on which she is in control (the idea of self-determination or self-governing individual).

If we now turn to the debate on socio-political freedom of choice, it is quite interesting to note that we can find the two conditions mentioned above as well, even if formulated differently. Specifically, their connotation varies depending on the specific tradition or branch within the socio-political debate we look into.

Skinner helps us to dig into this debate, as he maps all the different socio-political conceptions of freedom that have been defended since the birth of modern liberal political philosophy in the 17th Century, by showing how in this genealogy of freedom of choice in a social-political sense it is possible to identify three main branches to it[78]:

1) Those who outline that "to be free" means "to be free from interferences", where with interferences they refer to the use of physical force or coercive threats to influence or constrain individuals' choice (this first conception corresponds to Berlin's idea of 'negative' liberty) [79];

2) Those who sustain that "to be free" means to be able to act in a way that is consistent with your authentic self, and it entails some consistency between individuals' choices and personal values (Berlin's idea of 'positive' liberty);

Hegel-Kongress 1993, eds. H.F. Fulda and R.P. Horstmann, Stuttgart: Klett-Cotta 1994. C.M. Korsgaard, "The Normative Constitution of Agency," in Manuel Vargas and Gideon Yaffe (eds.), *Rational and Social Agency: The Philosophy of Michael Bratman*, New York: Oxford University Press, 2014, pp. 190–214.

[78] See Q. Skinner, *Liberty before Liberalism*, Cambridge University Press, Cambridge 2012; and Q. Skinner, *Hobbes and Republican Liberty*, Cambridge University Press, Cambridge 2008.

[79] For a focus on the socio-political freedom understood by the liberal tradition associated with Isaiah Berlin as *non-interference* see I. Berlin, *Two Concepts of Freedom*, Oxford University Press, Oxford 1969.

3) Those who endorse the idea of "to be free" as "free from domination", where domination occurs when individuals' actions are subjected to the will of another person or entity (this third conception instead clearly corresponds to Berlin's idea of 'negative' liberty) [80].

We can add also another (and fourth) conception of freedom in the social-political debate: the conception of freedom as independence[81], we analyzed before, which has been defended by taking the virtues of its liberal and republican counterparts (i.e., 1. freedom as non-interference and 3. freedom as non-domination): those who sustain this conception consider the possibility of freedom in the potential and actual absence of moral and immoral interferences (or interferences *simpliciter*).

This distinction is helpful as clarifies how each of these branches can imply a different formulation of the conditions *sine qua non* underlying the freedom of choice and action as conceptualized by the traditions characterizing the socio-political debate, insofar as different are the conceptions of freedom implicated by each of those branches. Therefore, this distinction is helpful, inasmuch as it shows the complexity of the attempt to clarify these conditions when it comes to deal with a massive and heterogeneous corpus of literature, as that one produced by socio-political freedom debate – whose analysis anyway does not fall under the scope of our inquiry.

Nevertheless, it can be noticed that each of these branches and definitions of freedom seems to acknowledge the idea that social and political freedom requires the condition of autonomy of the agent[82] – although there are differences[83] when it

[80] On this third conception, a major example in the republican tradition is the political theory offered by Philip Pettit in P. Pettit, *Republicanism: A Theory of Freedom and Government*, OUP, Oxford 2001; and in P. Pettit, "The Instability of Freedom as Non-Interference: The Case of Isaiah berlin" in *Ethics* 121, no 4 (2011), pp. 693-716. See also P. Pettit, *Just Freedom: A Moral Compass for a Complex World*, WW Norton, New York 2014.

[81] C. List & L. Valentini. "Freedom as Independence." *Ethics* 126, no 4 (2016).

[82] The problem here is that in the socio-political debate very often autonomy is used as an equivalent of freedom (this is particularly the case of those accounts of positive freedom where the meaning of freedom is very closer to what is generally is meant with autonomy).

[83] Killmister, J. (2017). *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance*. London (UK): Routledge.

comes to critically reckon with the sub-conditions of autonomy and its particular declinations (e.g., relational or as self-rule) that each of these branch's highlights.

To give but one example of this complexity, consider the liberal theory of autonomy that was first proposed by Raz back in the 1980s[84]. This theory focuses on three conditions that need to be satisfied if a particular choice is to count as autonomous:

> If a person is to be maker or author of his own life then he must have the mental abilities to form intentions of a sufficiently complex kind, and plan their execution. These include minimum rationality, the ability to comprehend the means required to realize his goals, the mental faculties necessary to plan actions, etc. For a person to enjoy an autonomous life he must actually use these faculties to choose what life to have. There must in other words be adequate options available for him to choose from. Finally, his choice must be free from coercion and manipulation by others, he must be independent.[85]

The three conditions of autonomy embedded in this quoted passage are:

(a) the agents must have a *minimum rationality* to plan actions that will allow them to achieve their goals (this condition identifies the extent to which one can make decisions that are based on identifying, weighing, and assessing options for their fit with one's preferences and plans);

(b) the agents must have *adequate options available to choose* from (i.e., the condition of availability of alternative options);

(c) the agents must be *independent* (wherewith it Raz intends autonomous as free from coercion and manipulation when making and implementing their choices).

In Raz's liberal account of autonomy, autonomy includes also the condition of availability of alternative options to choose, previously pointed out in the debate on free will, and the concept of autonomy is specifically connoted as independence, i.e., freedom from internal and external constraints, and specifically, from immoral interferences, as for example, those represented by coercion and manipulation (liberal tradition, 1 branch). To sum up, autonomy is described as the independence of one's deliberation and choice from manipulation by others. The concept of

[84] J. Raz, *The Morality of Freedom*, OUP, Oxford 1986.

[85] *Ivi*, p. 373.

autonomy as independence sees the autonomous subject as the sole 'author' of her own life and choices. A similar conception of autonomy emerges from the theory of freedom as independence elaborated by List and Valentini (branch 4), where however, as previously discussed, the interferences considered are either actual or potential, as in the republican tradition, but also, beyond the immoral interference, the moral interferences too are considered as freedom-undermining.

By looking at the second branch on the definition of socio-political freedom (branch number 2), the specific idea of autonomy that emerges concerns instead the authenticity (genuineness) of moral values and the ability of the agent to choose and act with sufficient resources and power to make one's values effective. Here the condition of autonomy is seen more as the power to rule or govern oneself, but in order to govern oneself, the agent must be in a position to act competently based on desires (values, conditions, etc.) that are in some sense one's own. This picks out the two families of conditions often proffered in this conception of autonomy: competency conditions and authenticity conditions. Competency includes various capacities, for example, rational thought or freedom from debilitating pathologies, self-control, and so on. Authenticity conditions often include the capacity to reflect upon and endorse one's desires, values and so forth (autonomy as endorsement).

It is important to notice how in the same tradition, the liberal one, we can at least distinguish a sense of autonomy that emerges from a negative concept of freedom, that is, autonomy as independence, and another sense of autonomy that instead emerges from a positive concept of freedom, that is, autonomy as self-rule, where the authenticity (or genuineness) factor mainly relies on the endorsement or internal approval of the agent's values, desires, reasons (and so forth) driving her choices.

Finally, by looking at the third branch (that represents republicanism, broadly speaking), autonomy is mainly used in a way equivalent to freedom, and specifically as absence of domination, where domination occurs when individuals' choices and actions are subjected to the will of another person or entity.

Nevertheless, the idea of autonomy as non-domination deeply differs from the liberal idea of autonomy. Republican tradition indeed has widely criticized the individualism connoting the liberal tradition and the concept itself of autonomy as independence and self-rule (endorsement) elaborated.

Specifically, the liberal concept of autonomy is criticized as a proceduralist conception, because liberalism would focus on the procedure through which a person can come to endorse options and values, as crucial in the determination of her autonomy. According these critics, the liberal idea of autonomy avoid to question whether the autonomy as self-rule or self-government of the agent may be understood independently of the perhaps socially defined values in terms of which people develop themselves. In particular, exponents from republicanism and communitarianism claim that the liberal procedural view runs counter to the manner in which most of us define ourselves, and hence diverges problematically from the aspects of identity that motivate action, ground moral commitments, and by which people formulate life plans.[86] Autonomy, it is argued, implies the ability to reflect wholly on oneself, to accept or reject one's values, connections, and self-defining features.[87] But we are all not only deeply enmeshed in social relations and cultural patterns, we are also defined by such relations.[88]

These critical considerations have sparked some to develop an alternative conception of autonomy meant to replace individualistic notions. This replacement has been called "relational autonomy"[89]. Relational conceptions of autonomy stress the ineliminable role that relatedness plays in both persons' self-conceptions,

[86] See for example MJ. Sandel (1982). *Liberalism and the Limits of Justice*, Cambridge: Cambridge University Press, 2nd ed., 1999.

[87] There have been many responses to these charges on behalf of a liberal outlook. The most powerful response is that autonomy need not require that people be in a position to step away from all of their connections and values and to critically appraise them. Mere piecemeal reflection is all that is required. As Kymlicka puts it: "No particular task is set for us by society, and no particular cultural practice has authority that is beyond individual judgement and possible rejection" W. Kymlicka. *Liberalism, Community and Culture*, Oxford: Clarendon, 1989, p. 50.

[88] For example, we use language to engage in reflection but language is itself a social product and deeply tied to various cultural forms. In any number of ways we are constituted by factors that lie beyond our reflective control but which nonetheless structure our values, thoughts, and motivations. See C. Taylor. *The Ethics of Genuineity*, Cambridge, MA: Harvard University Press 1991.

[89] See C. Mackenzie, & N. Stoljar (eds.), 2000a. *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, New York: Oxford University Press.

relative to which autonomy must be defined, and the dynamics of deliberation and reasoning. These views offer an alternative to traditional models of the autonomous individual, but it must be made clear what position is being taken on the issue: on the one hand, accounts on relational autonomy can be considered as resting on a non-individualist conception of the person and then claim that insofar as autonomy is self-government and the self is constituted by relations with others, then autonomy is relational; these accounts may be understood as claiming that whatever selves turn out to be, autonomy fundamentally involves social relations rather than individual traits.[90] Some such views also waiver between claiming that social and personal relations play a crucial causal role in the development and enjoyment of autonomy and claim that such relations constitute autonomy.[91]

Another relational element to autonomy that has been developed connects social support and recognition of the person's status to her capacities for self-trust, self-esteem, and self-respect. The argument in these approaches is that autonomy requires the ability to act effectively on one's own values (either as an individual or member of a social group), but that oppressive social conditions of various kinds threaten those abilities by removing one's sense of self-confidence required for effective agency. Social recognition and/or support for this self-trusting status is required for the full enjoyment of these abilities.[92]

Whether the complexity in bringing out a certain convergence in the broad and heterogeneous treatments offered by free will debate and socio-political debate is now clear, insofar many are the different formulations of the necessary conditions underlying our freedom to choose and act, at the same time, there are factors that – broadly speaking – seem to be enough acknowledged by both the debates. Indeed, by looking at these two far-reaching debates, the availability of alternative options and autonomy, either as independence, self-rule (authenticity in the endorsement),

[90] M. Oshana, "Personal Autonomy and Society," *Journal of Social Philosophy*, 29(1): 1998, pp. 81–102.

[91] For the discussion, see C. Mackenzie, "Three Dimensions of Autonomy: A Relational Analysis," in Veltman and Piper (eds.), 2014 pp. 15–41.

[92] See for example R. Arneson, "Autonomy and Preference Formation," in Jules Coleman and Allen Buchanan (eds.), *In Harm's Way: Essays in Honor of Joel Feinberg*, Cambridge: Cambridge University Press, 1991, pp. 42–73. P. Benson, "Feminist Intuitions and the Normative Substance of Autonomy," in J.S. Taylor (ed.), 2005 pp. 124–42. A. Westlund, "Autonomy and Self-Care," in Veltman and Piper (eds.), 2015, pp. 181–98.

or relational, seems to be conditions widely recognized as underlying the exercise of freedom of choice, both in a positive and negative sense.

In the light of the account of moral freedom previously elaborated, I take the moves from these two conditions mainly considered in the free will debate, i.e., the availability of alternative options and human autonomy, as they particularly clarify what it takes to consider a choice as free at a minimum threshold, and I revise them in a way that is informed by the socio-political debate on freedom of choice, insofar as the socio-political debate highlights either overlaps with free will debate (e.g., the conditions of availability of options and the conception of autonomy as self-rule or self-determination via reflective endorsement) and differences (e.g., in the recognition of the socio-relational dimension of autonomy) both extremely useful to sub-define our two-concept account of moral freedom that focuses on how we choose and act as moral agents effectively, i.e., in our contemporary societies.

Indeed, by analyzing these debates, it is possible to bring out the *conditions sine qua non* of our moral freedom, and therefore to underline an account of moral freedom that recognizes both the dimension of the individual and the socio-relational one, inasmuch as moral freedom is related to how we choose as individuals, especially as individual moral agents, but placed in socio-relational contexts, such as those of contemporary advanced informational societies, that are increasingly intercultural and globalized.

These conditions do not coincide with those detected by free will debate (broadly speaking), nor can be fully traced back to those formulated by the liberal tradition or the republican one.

The *conditiones sine qua non* for the exercise of moral freedom concern the individual dimension of approval (so the condition of autonomy intended as the agent's capacity of reflective endorsement) but in a way that is deeply informed by the socio-relational dimension, which specifically relates to another condition, that is, the condition of the availability of alternative options. These conditions in fact respond at a minimum threshold to what it takes to choose and act as moral agents, therefore to an account of moral freedom, as previously discussed, that see the authorship of the agents' choices and actions in the dimension of the endorsement but in a way that is not detached or isolated by the socio-relation dimension of the agent's living.

Therefore, I define moral freedom's necessary conditions as follows:

a) The *availability of morally heterogeneous options*: moral freedom requires the possibility to choose and act otherwise as moral agents. The condition of the possibility to choose and act otherwise as moral agents entails the possibility to choose and act in a qualitative-different context of options, and specifically in context of morally-heterogeneous options that reflects morally heterogeneous reasons and values. This is a very important aspect that requires a sufficient moral exposure of the subjects to social contexts, relations, attachments, values (and so forth) morally diversified. Indeed, a vice-versa situation, namely, with a no-qualitative diversified moral exposure, would prevent the agents to develop morally reasons and values that are alternative to those to which they have been exclusively exposed, and so they would not be able to act differently, to form their own idea of good, their moral projects and reasons, as well as to critically test their adhesion to the reasons and values endorsable, that is what is required to develop a genuine moral identity.

b) *Moral autonomy*: freedom to choose and act as moral agents and form genuine moral identity requires that the agent is considered the author of their choices and actions. However, this autonomy as authorship is not defined as full independence or complete self-governance of the subject, inasmuch the condition of autonomy is always informed by the socio-relational dimension via the availability of morally heterogeneous options. In this sense, the condition of autonomy is always seen as a *self-relational determination*. In this sense, moral autonomy consists in the possibility for the subjects to be the authors of their choices and actions by exercising the reflective endorsement on the morally heterogeneous options available, that in turn are deeply informed by the socio-relational dimension in the morally heterogeneous availability of options (values),

> hence, by endorsing them as moral motives for their choices and actions. The socio-relational dimension informs deeply the individual dimension in which the subjects choose and act as moral agents. The reflective endorsement, as the condition in which the subject expresses her self-determination is deeply informed by the social-relational dimension, so by moral autonomy we refer hereinafter to the capacity of self-relational determination of the agents.

As I underlined above, these are necessary but not sufficient conditions for our moral freedom. Indeed, as the two concepts of moral freedom elaborated in the first section may show, moral freedom is a complex and multi-dimensional concept. We can experience in some cases more moral freedom in a positive sense, while in other more moral freedom in a negative sense. This means that full-moral freedom or no-moral freedom scenario is highly unlikely, as previously explained.

The two conditions highlighted, namely, "the availability of morally heterogeneous options" and "moral autonomy", allow us to evaluate when and specifically which contexts are more moral-freedom restricting, and those instead that are more moral-freedom enhancing, by so sketching an evaluative field – at a minimum threshold – through which assessing new potential forms of impediment to our moral freedom.

I.3 The Ethical-Normative Value of Moral Freedom

In the previous sections, I elaborated a novel account of moral freedom, by drawing insights from theories developed in the free will debate and socio-political debate on freedom of choice. I defined moral freedom as our freedom to be(come) moral agents and specifically genuine moral agents. I have conceptualized moral freedom both in a positive sense, as freedom to develop our moral identity through choices and actions in a consistent way with our moral values, reasons, and ground projects, and in a negative sense, as freedom from actual and potential, moral and immoral interferences.

Then, specifically in the second section, always drawing insights from the free will debate and the socio-political debate on freedom of choice's underlying conditions, I brought out the necessary conditions underlying the possibility of our moral freedom, that are specifically, the availability of morally heterogeneous and moral autonomy as reflective endorsement, and I claimed that they can secure just a minimum threshold our moral freedom.

In this third and last section, I focus on the ethical-normative value of moral freedom, which means that I consider moral freedom to have a specific normative value, and I claim this through five arguments.

Firstly, it sounds necessary to clarify what we mean by normativity in moral philosophy. Korsgaard, one of the main exponents of moral constructivism and normative theories, provides a very clear definition of normativity[93].

> Normativity pervades our lives. We not merely have beliefs: we claim that we and others *ought to* hold certain beliefs. We not merely have desires: we claim that we and others *ought to act on* some of them, but not on others. [italics is mine]

When we seek the normative dimension of key ethical concept like ours of moral freedom, we are not looking merely for an explanation of a moral practice as the exercise of freedom of choice. We are asking what can justify the claims that moral concept makes on us. In other words, we are asking what can justify the protection of moral freedom as a fundamental and inviolable ethical value: this is what we call 'the normative question'. Thus, when you consider as normative an ethical concept, you are considering that you ought to respect it. Korsgaard outlines that this is the force of normative claims, "the right of these concepts to give laws on us"[94].

> The normative question is a first-person question that arises for the moral agent who must actually do what morality says. When you want to know what a philosopher's theory of normativity is, you must place yourself in the position of an agent on whom

[93] C.M. Korsgaard, *The Sources of Normativity*, Cambridge University Press, Cambridge 1996, p.1.

[94] Korsgaard, *The Sources of Normativity*, p.9.

> morality is making a difficult claim. You then ask the philosopher: *must I really do this*? *Why must I do it*? And his answer is his answer to the normative question.[95]

In order to understand the reason why our moral freedom should be respected as a fundamental ethical value, we have to raise these sorts of ethical questions. The answer to them is the normative value of our moral freedom.

There are at least five reasons that can allow us to argue for the ethical-normative value of our moral freedom, i.e., as something that adds a crucial value to our life and ought to be safeguarded as inviolable.

The first reason to argue for the ethical-normative value of moral freedom lies in the fact that in moral freedom we find the exceptional and distinctive trait of our humanity: in our moral freedom we find the dimension in which we develop and express the unique and deeper meaning of our choices and actions, that is, our moral identity (or moral posture), namely, what we develop over time as a result of the moral dispositions, values, horizons of meanings, and moral ground projects developed and endorse. Indeed, moral freedom concerns our capacity to develop, and give expression to, our moral standing, by choosing and acting as moral agents, and specifically as genuine moral agents. Put it in other words, moral freedom is what can allow us to semanticize, give meaning to our choices and actions, and broadly, to our existence in the world, by bringing the moral character out of our being persons, by bringing out our moral identity as moral agents from our being humans. Indeed, by moral freedom we can develop our own horizons of meaning, moral values, ideas of what is good, as well as our own ground projects and life goals, namely, everything that can morally motivate our choices and actions, and therefore our identity, towards a moral direction, rather than another.

Thus, because of the great value moral freedom adds to our lives and identity, its protection is a *fundamental value* that must be protected as an end for itself.

[95] Korsgaard, *The Sources of Normativity*, p. 16.

The second reason that allows us to reckon with the ethical-normative value of moral freedom is that it is an *intrinsic good*, i.e., it is intrinsically valuable, namely, something whose presence (irrespective of how it is used) adds value to our existence in the world. Indeed, moral freedom as the power to become moral, especially, genuine moral agents sheds light on how in this potential that we all can become – potentially – genuine moral agents lies its richness and value: in this opportunity to express ourselves as moral agents lies the value of being able to give meaning to our lives in a moral sense and to ourselves as persons.

By intrinsic good I mean that a world with moral freedom is always better compared to a world without it – even when our moral freedom can produce immoral outcomes. Indeed, our freedom to become moral agents also allows its contrary, i.e., freedom to become immoral (in the field of moral philosophy, this freedom is defined "freedom to fall"[96], i.e., the freedom to do wrong). Freedom to fall is not a limitation to our moral freedom, but it is its wrong exercise. Moral freedom in fact does not coincide with moral perfectionism, that is more a limit rather than a goal, inasmuch as moral perfectionism would imply no freedom to choose alternatively: even when that alternative course of action is not morally desirable, its possibility is always morally preferable to its impossibility. In a world with merit and blame, freedom to fall or freedom to become immoral is contemplated, and lead to a specific kind of moral identity: the immoral one (that is worthy to remind is not a-moral). However, to become immoral is not a loss of opportunity, rather a specific way to morally define your own identity.

To sum up, moral freedom is intrinsically valuable, as – even when it leads to immoral outcomes – without it we would not be able to develop our moral identity in a genuine way, to express the exceptionality of our moral character in our choices and actions, which in turn would lose their moral meaning and their moral extent.

[96] For example, see J. Harris, "Moral Enhancement and Freedom", in *Bioethics* 25 2011, pp. 102-111; and also I. Persson and J. Savulescu, "Moral Bioenhancement, Freedom and Reason", in *Bioethics*, 9 2016, pp. 263-268.

Connected to the previous reason, the third reason that allows us to maintain the ethical-normative value of moral freedom lies in the fact that moral freedom is also an *axiological catalysator*, i.e., a value whose presence can confer either more value or more disvalue to the respect or non-respect of other values.

As I argued previously, moral freedom is what enables us to form and choose our values, endorse them into our choices and actions, and therefore develop our moral identity (or moral posture). This means that it introduces an axiological difference in the dimension of our mere agency: the moral dimension, that is the dimension of our intentions which find their moral expression in the reflective endorsement we can give or not to options, reasons, and values towards which we choose to align our agency and behavior. In this sense, moral freedom makes choices and actions appraisable from a moral standpoint, by making choices and actions, respectively, good and bad, not on the basis of their consequences, but of the basis of agents' intentions. In this sense, moral freedom makes an agent's good and bad choices, respectively, also better and worse, insofar as chosen in a context of freedom, of moral presence, and possibility of authorship of the agent. This also entails that it would be difficult without moral freedom to consider our good and bad actions truly good or bad (as well as moral responsibility – and in certain cases – also legal imputability for certain choices and action). Indeed, without it, it would be difficult to evaluate whether the subject is the real agent of a choice, or whether something in the social environment she lives did not allow her to choose alternatively, whether she was the author of her choices and actions, or instead something has determined them in her place.

In short, whether we can guarantee moral freedom, we can assess, for better or worse, whether the subject who acts and chooses is morally present and free in her decisions and behavior. It follows that the protection of moral freedom as the freedom of choice and action as genuine moral agents assumes great ethical importance.

The fourth reason to define moral freedom as an ethical-normative value is that moral freedom is not just a fundamental value for the individual but also for the moral flourishing of our living together. Indeed, moral freedom as freedom from moral and immoral interference on the development of our moral identities, hence, on how we form and choose reasons, values, and ground projects through a morally heterogeneous exposure, grounds the possibility of a social dimension morally more open to the dialogue, mutual understanding, recognition, and even respect for differences in culturally heterogeneous values and practices, insofar as moral freedom allows the sharing and contamination of moral values.

Indeed, moral freedom as freedom to become genuine moral agents means freedom to choose which are the reasons we want to endorse as motives for our agency and behavior, and this means the possibility to develop a *more convinced* culture of giving-reasons (and hence moral justifications), that is fundamental in boosting the moral dimension of our living together, as well as a more cohesive social sphere, by favoring openness to mutual understanding, to sharing moral commitments, and therefore to develop joint cooperation towards social goals and common goods.

The fifth and last reason underlying our conceptualization of moral freedom as an ethical-normative value is that moral freedom morally grounds the respect of any other ethical value, as it concerns the development of morality as *ought to*, i.e., the development of our moral obligation towards any other value. Moral freedom indeed allows us to choose what moral values endorse into our choices and then make them moral law or moral rules for our conduct.

To sum up our inquiry so far: in this first chapter of the dissertation, I focused on the concept of moral freedom and tried to unpack it, drawing insights from the theories of freedom of choice. Specifically, I have highlighted two concepts of moral freedom, a positive concept of moral freedom, as freedom to become genuine moral agents, and a negative concept of moral freedom as freedom from actual and potential, moral and immoral interference on our development of genuine moral identities. Then, I have clarified what are the conditiones *sine qua non* underlying the exercise of our moral freedom, by drawing insights from theories developed in the free will debate and in the social and political one, and I defined these conditions as a) “the availability of morally heterogeneous availability” and b) “moral autonomy” as self-relational determination. I ended by arguing the ethical normative value of moral freedom, hence, why it is a fundamental value that ought to be secured from actual and potential new forms of impediments and I underlined how the ethical inquiry on moral freedom today must expand in the light of new potential moral freedom-constraining forces, as those triggered by algorithms-based technologies.

However, the complexity of moral freedom is just one side of the coin.

In the next chapter, I deal with another complexity: that one characterizing the phenomenon of algorithmic governance that is establishing in our contemporary informational societies, and then question whether this algorithmic governance can undermine our moral freedom.

SECOND CHAPTER

The Rise of Algorithmic Governance

That today our informational societies are more and more permeated by algorithms seems to be a matter of fact. Examples abound.

Individuals continuously interact with algorithm-based ICT starting from recommendation systems (RS) which make unceasingly daily suggestions about what a user may like and choose (from a song, a movie, a product, or even a friend).[97] Online service providers daily mediate how and particularly what piece of information is accessed through personalization and filtering algorithms.[98] There are algorithms determining who is the most likely to be guilty of tax evasion[99] and algorithms which trade stocks in Wall Street[100] or which help people in dating and mating[101]. As a consequence, we are daily nestled into a pervasive 'network' of algorithms, which in turn are more and more assuming a key role and social power in our societies and personal lives.

Indeed, especially in recent year with the huge availability and tremendous advances in algorithms' development, ever more daily tasks, personal choices, and high-stakes decisions – previously just left to humans – started to be increasingly delegated to algorithms-based ICTs: social services and both public and private

[97] For a general overview on algorithmic recommendation systems, see S. Milano, M. Taddeo, and L. Floridi, "Recommender Systems and Their Ethical Challenges" in *AI & SOCIETY* 2020. Consider also the works of D. Paraschakis, *Algorithmic and Ethical Aspects of Recommender Systems in E-Commerce*, Malmö universitet, Malmö 2018; and the analysis of N. Perra and L. Rocha, "Modelling Opinion Dynamics in the Age of Algorithmic Personalization" in *Scientific Reports* 9 (1) 7261 2019.

[98] S. Newell and M. Marabelli, "Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of datification" in T*he Journal of Strategic Information Systems* 24(1) 2015, pp. 3-14.

99 T. Zarsky, "Transparent predictions". University of Illinois Law Review, 2013 (4).

[100] S. Patterson, *Dark Pools: The Rise of AI Trading Machines and the Looming Threat to Wall Street.* Random House 2013.

[101] E. Siegel. *Predictive Analytics: The Power to Predict who will Click, Buy, Lie or Die*. John Wiley and Sons 2013.

infrastructures such as schools and hospitals[102], financial institutions[103], courts[104], local governmental bodies[105], and national governments[106] are ever more relying on algorithms-based ICTs to make significant and life-changing decisions, whereby how they are designed has become more and more a pivotal issue.

This increasing social power and authority given to algorithms-based ICTs in our contemporary societies is acknowledged by a growing number of scholars as the "rise of algorithmic governance" or the rise of – in its terminological variants – "algorithmic regulation"[107] and "algorithmic governmentality"[108]. Whenever you are denied a job opportunity by an algorithmic RS or you are negated a loan by a credit-scoring algorithm, whenever you are told which way to drive by a GPS routing-algorithm, or you have been diagnosed a certain disease or are prompted to exercise in a certain way, to take some drugs, or eat a certain food by health-oriented apps, you can rightly claim that you live within a society governed by algorithms.

[102] See, for example, Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations" in *Science* 366 (6464) 2019, pp. 447-453; see also N. Zhou *et alia*, "Concordance Study Between IBM Watson for Oncology and Clinical Practice for Patients with Cancer in China", in *The Oncologist* 24 (6) 2019, pp. 812-19. Consider also J. Morley, C. Machado, C. Burr *et alia*, "The Debate on the Ethics of AI in Health Care: A Reconstruction and Critical Review" in *SSRN Electronic Journal* 2019.

[103] See M.S.A. Lee and L. Floridi, "Algorithmic Fairness in Mortgage Lending: From Absolute Conditions to Relational Trade-Offs" in *SSRN Electronic Journal* 2020. See also N. Aggarwal, "The Norms of Algorithmic Credit Scoring" in *SSRN Electronic Journal* 2020.

[104] Look at B. Green and C. Yiling, "Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments' in *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, pp. 90-99. ACM Press Atlanta, GA, USA 2019. See also M. Yu and G. Du, "Why Are Chinese Courts Turning to AI?", *The Diplomat*, January 2019.

[105] On this specific issue see the masterpiece of V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, New York, NY 2017. See also D. Lewis, "Social Credit Case Study: City Citizen Scores in Xiamen and Fuzhou", *Medium: Berkman Klein Center Collection*, 8 October 2019.

[106] To expand this specific point, see R. Labati, A.G. Donida, E. Muñoz, V. Piuri *et alia*, "Biometric Recognition in Automated Border Control: A Survey" in *ACM Computing Surveys*, 49 (2), pp. 1-39 2016. See also T. Hauer, "Society Caught in a Labyrinth of Algorithms: Disputes, Promises, and Limitations of the New Order of Things" in *Society* 56 (3) 2019, pp. 222-230; and H. Roberts, J. Cowls *et alia*, "The Chinese Approach to Artificial Intelligence: An Analysis of Policy and Regulation" in *SSRN Electronic Journal* 2019.

[107] K. Yeung, "Algorithmic Regulation: A Critical Interrogation" in *Regulation and Governance* 12, no 3 2018, pp. 505-523.

[108] A. Rouvroy, "Algorithmic Governmentality and the End (s) of Critique" in *Society of the Query*, 2 2013; and "Algorithmic governmentality: a passion for the real and the exhaustion of the virtual", *All watched over by algorithms*, Berlin January 2015.

Legal scholars and philosophers in the field of ethics of AI and algorithms have started to address the phenomenon of algorithmic governance and its potential consequences, by generating three main debates according to specific sets of ethical problems detected in relation to the rise of the algorithmic governance:

- The first debate is about *privacy* and *surveillance*[109] and it analyzes the algorithmic governance in the light of the widespread 'datification' of our societies. Specifically, those involved in this debate focus on how the massive production and large availability of individuals' data along with the growing algorithmic power and ability to sort, pars and mine information from data raise many concerns about individuals' privacy and surveillance.

- The second debate is on *bias* and *inequality*[110] and is specifically focused on the way in which the algorithmic governance phenomenon can replicate and reinforce discrimination and social injustice in contemporary societies due to the spread of historical discriminating biases embedded in algorithmic training datasets or as inferable by proxy.[111]

- The third debate focuses on the problems of *transparency* and *procedure* related to algorithmic governance and specifically to the increasing opacity and not-explainability of algorithms (so defined as "black boxes"[112]), as

[109] For a specific focus on this first debate, see J. Polonetsky and O. Tene, "Privacy and Big Data: Making ends meet" in *Stanford Law Review* 66 2013, pp. 25-33.

[110] For a deep focus on this second debate, consider the works of K. Crawford, "The hidden biases of Big Data" in *Harvard Business Revie*, 1 April 2013; T. Zarsky, "Automated predictions: perception, law and policy" in *Communications of the ACM*, 15, no 9 2012, pp. 33-35; T. Zarsky, "Transparent prediction" in *University of Illinois Law Review*, 4 (1504) 2013; R. Binns, "Fairness in Machine Learning: Lessons from Political Philosophy" in *Journal of Machine Learning Research* 81 2018, pp. 1-11; V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, New York, NY 2017; S. Noble, *Algorithms of Oppression*, NYU Press, New York 2019; C. O'Neil, *Weapons of Math Destruction*, Penguin London 2016.

[111] In other words, very often, machine learning algorithms are trained to learn how to predict future behaviors by spotting patterns in large databanks of past behaviors. These training databanks, as well as the patterns that are extrapolated from them, can be biased, inasmuch as they can contain historical human biases and hence replicate them to a large extent.

[112] F. Pasquale, *The Black Box Society*, Harvard University Press, Cambridge (MA) 2015. As Frank Pasquale as widely argued in his *The Black Box Society*, they are 'black boxes': they shape

> they are very often too complex in the way they technically operate (so resulting to be unintelligible also to an expert human eye) or as covered by the provider's trade secret[113]. Specifically, those who are involved in this third debate argue that this lack of algorithmic transparency is a threat to the values at the core of our democratic societies[114], above all, when algorithms produce controversial effects for the individuals.

However, no debate or research developed so far is focused on the potential ethical implications and risks raised by the algorithmic governance on our moral freedom.

This second chapter of my dissertation aims at filling this specific gap and be a forerunner for the discussion on the impact of algorithmic governance on our moral freedom. Specifically, I question whether the algorithmic governance is generating a hampering impact for the exercise of our moral freedom, by analyzing whether algorithms can undermine the *conditiones sine qua non* underlying the exercise of moral freedom.

To pursue my inquiry, in the first section, I provide a technical analysis of the main algorithmic models and techniques nowadays in use, and show how they are reshaping our informational environment and the context in which we prepare and make our choices, by giving rise to what I define "the establishment of algorithmic choice-architecture".

In the second section, "the epistemological problem of shaping users' options", I further develop how algorithmic choice-architectures redefine the informational environment and can weaken the users from the epistemological standpoint. I claim how this epistemological impact can constitute a constraint on the first *conditio sine qua non* for the exercise of our moral freedom, that is – as argued in the previous section – the availability of morally heterogeneous alternative options (what I also define as the epistemological level of individuals' autonomy).

the world around them without being transparent or comprehensible to human beings that they affect.

[113] J. Burrell "How the machine thinks: Understanding opacity in machine learning systems" in *Big Data and Society* 2016, pp. 1-12.

[114] D. Citron and F. Pasquale, "The scored society: due process for automated predictions" in *Washington Law Review*, 89, 1 2014, pp. 1-34.

In the third section, I show how this algorithmic impact can overcome the epistemological level of users' autonomy and, in certain cases be deeper, by affecting our moral autonomy at its core, namely, by suspending the reflective endorsement through which we approve the reasons and the options as motives for our choices and actions, deeply undermining the second *conditio sine qua non* for the exercise of moral freedom.

II.1 The Establishment of Algorithmic Choice-Architectures

As we argued above, whilst the rise of algorithmic governance is currently widely acknowledged in the debate of ethics of information technology, ethics of AI and algorithms, the literature is missing a reflection on algorithmic governance's ethical implications for our freedom, and especially for our moral freedom.

This missing ethical investigation is specifically what I develop in the three sections of this chapter. In doing it, I firstly need to clarify the use of the term 'algorithm' I adopt and the specific kinds of algorithmic models I specifically consider into my ethical inquiry in order to assess whether – and, if this is the case, how – algorithms can affect or even undermine our moral freedom.

I devote the first part of this section to the clarification of the technical terms and categories involved in my analysis to the extent it is propaedeutic to understanding the second part, which is focused on introducing my concept of *algorithmic choice-architectures* and specifically explaining how the rise of algorithmic governance is turning out in the establishment of algorithmic choice-architectures, that is, the consolidation of the social power of algorithms in architecting the context in which we prepare and make our choices.

The term 'algorithm' assumes a wide array of meanings across diverse fields such as computer science, mathematics, or public discourse, therefore it is complex to univocally define it. Nevertheless, to ethically analyze and above all determine the potential and actual impact of ICTs based on algorithms on our moral freedom, we cannot prescind from this technical task.

As noticed by Burrell and Kitchin, the majority of the literature in the fields of ethics of information technology and ethics of AI fails to specify the technical categories or a formal definition of an 'algorithm'[115]. The most commonly adopted formal definition of an algorithm as a mathematical construct is that offered by Hill (2016), whereby an algorithm is defined as "a finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose under given provisions"[116]. Although this formal definition is a common reference benchmark to understand the algorithmic structure with a formal macro lens, my inquiry is not limited to algorithms as mathematical constructs, but also considers domain-specific understandings focusing on the implementation of these mathematical constructs into a technology configured for a specific task.

Let me make this point clearer.

From Hill's specific use of the terms 'purpose' and 'provisions' – involved in the definition of what an algorithm is – we can broadly understand algorithms as designed and operationalized to take action and produce effects. Here the use of the term in public discourse becomes relevant, insofar as public discourse does not generally define algorithms as general mathematical constructs but as particular implementations, namely, the definition of an algorithm also depends on its practical implementation into a specific technology and on the application of that technology as designed for a specific task. For our inquiry, it makes little sense to consider algorithms only as an abstract formalization and therefore independently of how they are implemented and operationalized in real-life application fields. For this reason, I have used so far the term algorithm-based ICTs to refer to algorithms as "implementations", namely, to artefacts with an embedded algorithm. Hence, for the sake of simplicity, I will continue to refer hereinafter to algorithms-based ICTs. However, the term is broad, as it comprehends different models (e.g., deterministic or probabilistic algorithms) and task-based techniques (as profiling, filtering, classification by prioritization, just to mention a few). Therefore, before I get underway, I need to clarify which algorithmic models and specific algorithmic techniques I refer to in my analysis.

[115] J. Burrell, "How the machine 'thinks:' Understanding opacity in machine learning algorithms". *Big Data & Society*, 3(1), 2016, pp. 1-12. R. Kitchin, "Thinking critically about and researching algorithms". *Information, Communication & Society* 20 (1), 2016, pp. 14-29.
[116] R.K. Hill. "What an algorithm is". *Philosophy & Technology*, 29 (1), 2015, p 47.

The scope of my ethical inquiry on algorithms and moral freedom is restricted to algorithms that can learn from massive amounts of data generated by users to product outputs used to redefine our social environment and specifically our choice-context, as the latter is crucial to evaluate the algorithmic impact on the conditions underlying our moral freedom. These algorithms are an object of study in the field of machine learning (ML) algorithms[117], therefore I use the term algorithmic ICTs to refer hereinafter to ML-based ICTs.

ML concerns "any methodology and set of techniques that can employ data to come up with novel patterns and knowledge, and generate models that can be used for effective predictions about the data"[118]. ML is specifically connoted by the capacity to define or modify decision-making rules autonomously, and its specific connotation is to be probabilistic, hence, not deterministic: this means that ML outputs do not follow from causal relationship or correlations (such as in the case of deterministic algorithms that usually follow basic instructions and causal rules), but they are induced by ML from data. ML algorithms show a predictive power: this is possible as they are not pre-programmed with certain rules in order to solve particular problems; instead, they are programmed to 'learn' to solve problems[119].

Let us think, for example, of ML algorithm applied to classification tasks. They typically consist of two components: the *learner* which produces a *classifier* with the intention to develop classes that can generalize beyond the training data and find new correlations driving certain outputs.[120] The ML algorithm functioning works by placing new inputs into a model or classification structure. The algorithm 'learns' by defining rules to determine how the new inputs will be classified. The model can be taught to the algorithm via hand labelled inputs (supervised learning

[117] See T. Mitchell, *Machine Learning*, Singapore McGraw-Hill 1997.

[118] Van Otterlo, M. (2013). "A machine learning view on profiling". *Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology* (edited by Hildebrandt M and de Vries K). Abingdon (UK). Routledge: 41-64.

[119] P. Domingos, "A few useful things to know about machine learning". *Communications of the ACM*, 55(10) 2012, pp. 78–87

[120] "Every algorithm has an input and an output: the data goes into the computer, the algorithm does what it will with it, and out comes the result. Machine learning turns this around: in goes the data and the desired result and out comes the algorithm that turns one into the other. Learning algorithms – also known as learners – are algorithms that make other algorithms". See P. Domingos. *The master algorithm: how the quest for the ultimate learning machine will remake our* world. New York, NY: Basic Books 2015.

or human-in-the-loop), or in other cases the algorithm itself defines best-fit models to make sense of a set of inputs (unsupervised learning or human-out-of-the-loop)[121]. However, in both cases, the algorithm defines decision-making rules to handle new inputs – algorithmic rules or correlations whose rationale the human operator very often does not understand.[122]

Even though highly complex, ML algorithms are all around us, and they are weaving the social fabric of our societies: when you enter a research query into an Internet search engine (like Google and Bing), there is a ML algorithm, and also ML algorithms underlie the decision about which results (informational contents and adds) in turn the search engine will show you; ML algorithms filter our email function by excluding what is labeled as spam, as well as ML algorithms rule the recommendation of some of our most commonly used websites, apps, and services from Amazon, Netflix, and Spotify to Yelp up to the health-oriented ones like Fitbit (just to mention a few). ML also rule the platforms where we spend a huge amount of time as Facebook, Twitter, and Instagram to decide which update, post, picture, and tweet to show you. Even more, ML are used to determine college admission, job application, housing, credit landing, cancer diagnosis, and so forth. As pointed out, considering that the application fields and the tasks for which they are applied for are many and diverse, multiple and diverse are also the algorithmic techniques developed and in use so far.

Here I specifically focus on three main kinds of ML-driven algorithmic techniques, which are often used combined all together with the general purpose to personalize users' experience: i) *algorithmic profiling*, ii) algorithmic *classification and filtering*, iii) and *algorithmic recommendation systems*. The reason for focusing on these kinds of algorithmic techniques is that they clearly show how algorithms can learn from users' data and generate outputs which are used purposely or accidentally to redefine people's social environment and specifically user' choice-

[121] B.W. Schermer, "The limits of privacy in automated profiling and data mining". *Computer Law & Security Review*, 27(19), 2011, pp: 45-52.

[122] A. Matthias, "The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and Information Technology", 6(3), 2004, p. 179.

contexts. Moreover, ML algorithms are the mainly in use today:[123] though their complexity and opacity[124], they rule the functioning of the majority of algorithmic ICTs surrounding us, as shown by the cases of application domains above mentioned. The same reason is valid for the algorithmic techniques chosen: they are not just the most pervasive one, but also the most diffuse, subtle, and fine-grained, as they act invisibly and silently in reshaping our informational space. Therefore, their consideration is mandatory, inasmuch as their analysis allows us to show how the algorithmic governance works specifically in redefining the contexts in which we prepare and make our choices.

II.1.1 Algorithmic profiling

Algorithmic profiling occurs in a large diversity of contexts: from criminal investigation to marketing research, from mathematics to computer engineering, from healthcare applications for elderly people to genetic screening and preventive medicine, from forensic biometrics to immigration policy with regard to iris-scans, from supply chain management with the help of RFID-technologies to actuarial justice. [125] Profiling is used to refer to a set of technologies, which share at least one common characteristic: the use of algorithms or other techniques to create, discover and even construct knowledge from huge sets of data. Automated profiling involves different *technologies* (i.e., hardware), such as RFID-tags, biometrics, sensors and computers as well as *techniques* (software), such as data cleansing, data aggregation and data mining. Both technologies and techniques are integrated into *profiling practices* that allow both the construction and the application of *profiles* on people (profiles used to make decisions, sometimes even without human intervention, and on which is based the restructuring of our environment, online and offline). Here, I

[123] A. Tutt, *An FDA for algorithms. SSRN*, Rochester, NY: Social Science Research 2016.

[124] This critical issue is widely argued by T. Zarsky, "The trouble with algorithmic decisions an analytic road map to examine efficiency and fairness in automated and opaque decision making". *Science, Technology & Human Values*, 41(1), 2016, pp. 118-132.

[125] A. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law*. New York, NY (USA): NYU Press 2017. V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, New York, NY 2017; S. Noble, *Algorithms of Oppression*, NYU Press, New York 2019; C. O'Neil, *Weapons of Math Destruction*, Penguin London 2016.

focus specifically on profiling techniques (and therefore as software, rather than specifically hardware, as software are embedded in every technology, insofar as in the world of IoT, everything is deeply interconnected), and especially I focus on algorithmic profiling as a result of data mining, namely, a procedure by which large databases are mined by means of algorithms to find patterns or correlations between data.

In this sense, we can define algorithmic profiling as connoted to be a *way of detecting patterns and making predictions on the basis of them*. Profiling can be inductive or deductive, or a combination of both. Inductive profiling relates to the generation of profiles as testable hypotheses, while deductive profiling is concerned with testing profiles on datasets to confirm hypotheses.

Algorithmic profiles resulting from data mining are currently used in a wide range of contexts including insurance, finance, differential pricing for advertising and marketing, education, employment, governance, security, and policing. In all these contexts, algorithmic profiling is used as a method of inferential analysis that identifies correlations and patterns within datasets, that can be used as an indicator to classify a subject as a member of a certain group.[126]

These groups or categories are formed from probabilistic assumptions[127] that are de-individualized.[128] The first aspect, i.e., the probabilistic assumptions on which algorithms depend to infer patterns and categorize people, means that correlations indicate a relation between data, without establishing causes or reasons. The second aspect, i.e., the fact that probabilistic associations and patterns on which people are categorized are de-individualized – can be made clearer with an example. Let us think about an ML algorithm designed for a profiling task, i.e., to determine whether a person is creditworthy in the evaluation process for a loan-request. The

[126] M. Hildebrandt, "Defining profiling: A new type of knowledge?". *Profiling the European Citizen* (edited by Hildebrandt M and Gutwirth S). The Netherlands: Springer, 2008, pp. 17-45. W. Schreurs et al., Cogitas, "ergo sum. The role of data protection law and non-discrimination law in group profiling in the private sector". In: Hildebrandt, M, Gutwirth, S (eds) *Profiling the European Citizen: Cross-Disciplinary Perspectives*. Dordrecht: Springer, 2008, pp. 241–270.

[127] M. Leese, "The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards". *The European Union*. *Security Dialogue*, 45(5), 2014, p. 502.

[128] B.W. Schermer, "Risks of profiling and the limits of data protection law". In: Custers, B, Calders, T, Schermer, B, et al. (eds) *Discrimination and Privacy in the Information Society*. Berlin: Springer, 2013, pp. 137–152.

decision for a loan application may not be made by the algorithms on the basis of the consideration of real individual risk of default, but on the basis of postcode or neighborhood, that may operate as an indirect proxy of other indicators such as socio-economic or racial composition of one's neighbors – that can be correct and so reflect a real situation or not. The de-individualization specifically depends on the fact that the profiles of people are not only generated on the basis of a user's data, but from their correlation with massive amounts of available data produced by others, where the algorithm is called to discover precious patterns, i.e., to discover patterns that can be helpful in predicting future behavior (of people as consumers, voters, and so on).

Algorithmic profiling has started to rule almost every ICTs to deal with one of the most challenging problems of the informational society, namely, dealing with increasing data-overload or informational overload. In the informational societies, indeed, everything is information, everything in the physical world is assimilated as information that fuel ICTs and SNS via what we enter online as well as what can be captured by the increasing presence of IoT or ubiquitous computing:

> Digitization penetrates every aspect of our lives: the technology nestles itself *in* us (for example, through brain implants), *between* us (through social media like Facebook), knows more and more *about* us (via big data and techniques such as emotion recognition), and is continually learning to behave more *like* us (robots and software exhibit intelligent behavior and can mimic emotions). IoT penetrate in our *material world* (e.g., the production process, public space, and our home) and is based on a network that integrates the physical world with the virtual world of the Internet. Through the emergence of IoT, we are on the brink of a new era in which objects and people in the material world can be monitored, and where objects and people can exchange information automatically. In this way, the alarm clock does not just wake up a person, but at the same time switches on the coffee machine for making fresh coffee with our breakfast; or the fridge tells us a product has passed its expiry date; or the lighting in the room adjusts itself to what is happening in a video game being played at that moment. [129]

[129] L Royakkers, et al. Societal and Ethical Issues of Digitization. *Ethics and Information Technology*, *20*(2), 2018, p. 127.

The interconnectedness of our contemporary societies means the digitalization of all sorts of object, content, thought, emotion, affiliation, relation: *everything is information*, everything can be captured as datified by ICTs and IoTs and used to train ML algorithms. The digitalization of our world means that everything can be captured and datified, and algorithmic profiling is the specific technique that rules this process. We as persons are datified, and our data fuel algorithms to compare ourselves with others in order to generate profiles of us used to categorize us into groups, as previously underlined. This means that ML algorithms are fueled not just by the data we directly enter, i.e., *provided data*, but also the "onlife"[130] trails we indirectly leave behind us, i.e., *observable data*, up to *derivable data*, i.e., derived and inferred information obtained through the correlation of individuals' data with other sets of data already available.

The process of datification is huge and encompasses everything, or better, what is not datified often finished to be ignored alike something that does not exist. Let me give a few examples of the datification of our lives and ourselves.

For example, my exercise is now datified. I went for a run this morning, and my FitBit device recorded exactly how long I ran for, how many strides I took, and how many calories I burned in the process. My network of friends is now datified with Facebook. My network of professional connections is datified with LinkedIn. My location is datified with Foursquare. My latest random thoughts are datified on Twitter, or via my search history on Google, my music preferences are datified with Spotify; my readings by Kindle devices, what I daily watch on Netflix or I buy on Amazon. Since due to ICTs and IoTs the amount of available information becomes enormous and increases exponentially, it has thus become important for companies to discriminate information from noise and detect useful or interesting information, above all to increase economic benefits.

To sum up: as everything is codified in data – from individuals' characteristics to habits, opinions, interactions, and movements – and generates a value-laden vast amount of information about people, profiling algorithms have become inescapably value-laden in discovering patterns and correlations in large quantities of

[130] L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford 2014.

aggregated data. As specified above, this process is specifically called predictive knowledge-discovery. What profiling algorithms provide is a kind of prediction that is based on past behavior (of humans or nonhumans like artificial "agents"). In that sense, the correlations generated stand for a probability that things will turn out the same in the future. What they do not reveal is why this should be the case. In fact, profilers are not very interested in causes or reasons, their interest lies in a reliable prediction, to allow adequate decision making. For this reason, profiling can best be understood from a pragmatic perspective: it aims at a kind of knowledge that is defined by its effects, not by causal connections (i.e., inductive profiling).

Another way to articulate the particular kind of knowledge produced by profiling is to see profiles as hypotheses. Interestingly, these hypotheses are not necessarily developed within the framework of a theory or on the basis of a common-sense expectation. Instead, the hypothesis often emerges in the process of data mining, a change in perspective that is sometimes referred to as a discovery-driven approach (deductive profiling).

From a procedural standpoint, we can see the technical process of profiling as separated in several steps:

- Preliminary grounding: the profiling process starts with a specification of applicable problem domain and the identification of the goals of analysis, hence, not with the specification of causes and steps, but only of the goal or task that algorithms have to achieve.
- Data collection: the target datasets or databases for analysis are formed by macro selecting the relevant data in the light of existing domain knowledge and data understanding (narrowing the datasets in the light of domain goal).
- Data preparation: data are pre-processed for removing technical noise.
- Data mining: data are analyzed heuristically with the algorithm developed to suit the data, model and goals, to probabilistically lead to the emergence of valuable recurrence, inferences, and patterns from correlations.
- Interpretation: the mined patterns are tested and evaluated on their relevance and validity by professionals in the application domain to exclude spurious

correlations (in the automated profiling this phase foresees human out of the loop).

- Application: the constructed profiles are applied to categories of persons, to test and fine-tune the algorithms, this means, the profiles are used to group large amount of people on the basis of the discovery of recurrent patterns or correlations as similarities between groups and categories.

Data collection, preparation and mining belong to the phase in which the profile is under construction. However, as specified above, profiling also refers to the application of profiles constructed, namely, the usage of profiles discovered for the identification or categorization of groups or individual persons to drive ADM, e.g., for classification and filtering purposes. However, the process is circular: there is a feedback loop between the construction and the application of profiles. The interpretation of profiles can lead to the reiterant – possibly real-time – fine-tuning of specific previous steps in the profiling process. The application of profiles to people whose data were not used to construct the profile is based on data matching, which provides new data that allows for further adjustments. For this reason, the nature of algorithmic profiling is defined as smart or self-learning as both dynamic and adaptive.

The last step, i.e., the application of profiles, easily leads us to explore the second specific algorithmic technique (based on profiling as data mining) and to argue why it plays a crucial role in shaping our choice-contexts: algorithmic classification and filtering.

II.1.2 Algorithmic classification and filtering

Algorithmic classification and filtering are personalization algorithms which – on the basis of constructed profiles – filter and classify the informational contents available to the users. In other words, the informational availability to the users is tailored and depends on the profiles constructed for them, and more specifically is shaped on the basis of the categories under which those profiles fall (see the image below). This means that on the basis of how the user is profiled and categorized, she will see different updates and information, i.e., she will have different informational contents as available.

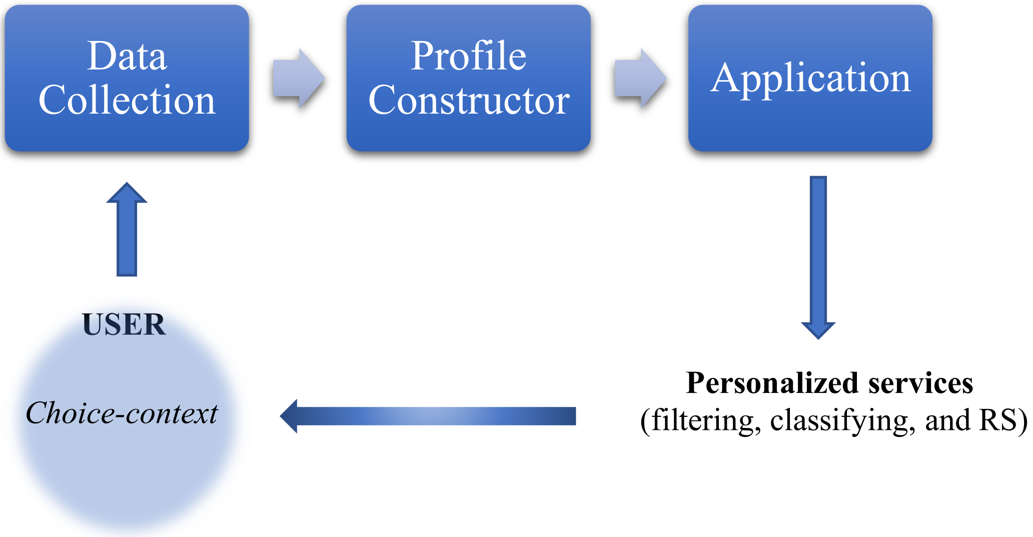


Fig. 1. User's profile construction for personalization.

Personalized systems also address the above-cited problem of informational overload, by managing, filtering, and classifying information in a way customized for individual users, or better, to their profiles. This is because otherwise it would require an unreasonable effort and time for any individual to audit all the available information. For example, Facebook's personalization algorithms track the user's interactions with other users, the so-called "social gestures"[131] such as likes, shares, subscribes, and comments. When the user interacts with the system by consuming a set of information, the system registers this user interaction history. [132] Later, on the basis of this interaction history and its correlation with those of other members,

[131] B. Upbin, Facebook ushers in era of new social gesture, *Forbes* 2011.

[132] When a user clicks on an item or views a page, profiling algorithms assume that this indicates some user interest in the informational content clicked.

the user is profiled as belonging to a certain category, and her information is so filtered out.[133]

In this sense, personalization algorithms (such as filtering and classifying) control the incoming information (users do not see everything available), but also determine the outgoing information and whom the user can reach (not everything shared by the user will be visible to others). In this sense, filtering and classifying algorithms are *informational gatekeepers* that replace traditional media. Indeed, the latter were used to perform this gatekeeping role for news, determining what is newsworthy and important for its audience. In this sense, algorithmic gatekeepers control today whether information, on the basis of the knowledge they have on us (profiles), but, their gatekeeping role is more pervasive and enveloping, because of they are embedded in a large variety of ICTs and IoT which, as previously outlined, do not concern and rule the sectors of news and advertising.

The social (and also political) power of the algorithmic governance becomes now clearer. When, indeed, with the digital revolution, everything is considerable as information, information becomes a "primary good"[134], a good that everybody requires as a condition for their well-being, to choose and act in a way that is genuine and aware.

Access to informational contents is crucial as informational contents do not just add alternative to one's choice set of options, but *are choice options themselves*: an informational content can be a reason that informs our choice, as well as an informational ad can reflect an opportunity that can represent a motive to choose and act in a certain way, so it becomes an option to which we can steer our choices and actions.

Due to the digitalization and datification of everything carried out by ICTs, information is enveloping our world: our alternative options can be conceived in informational terms. This means that every informational item, content, ad (but also: movie, song, product, news, update, opinion, friend, book, and the list can go

[133] This means that contents produced by certain friends might be hidden from the user, because the user did not interact with those friends over a period of time.

[134] J.V., Hoven, & E. Rooksby, "Distributive justice and the value of information: A (broadly) Rawlsian approach". England 2008.

on and on) can become an option that we can value (if not it is a value in itself), a reason for our choices and actions, and specifically, a reason that we can endorse as a motive to determine our choices and actions. Therefore, who shapes, allows or blocks access to information, determines what options are available, and what instead is unavailable to us.

Algorithms deciding what options are available and rule out alternatives as unavailable have a key-function in structuring our choice-contexts, i.e., the context of options on which we make our choices and so perform our actions. This means that who determines access and manages information, also structures the choice-context of individuals, and hence influences the way in which persons can choose and act, behave, and by doing so, develop their identity and live their lives.

Let us think of an example. Imagine that you are going out with your group of friends and you want to find a quiet place to have a conversation. You open TripAdvisor and start to look for places where you can have a break, by digiting the particular key-word ‘snack’. TripAdvisor will show a list of places and every person starts to look at what place shows the catchiest pictures of food and drinks. The question is: is the menu relevant to the group’s original need? In this case, the SNS substitutes the original question about “a place to make a break” with “the restaurant with the best pictures about food drinks”, while none of the member of the group originally felt the need to eat something, but their need or desire ends to align with the options available. Moreover, the group has the illusion that TripAdvisor’s list of places represents a complete list of options, as places to go. While looking down at their phones, they do not see the park across the street with a band playing live music. They miss a coffee shop serving cakes and coffee. None of those showed up on the SNS’s menu. This is a simple example, but the mechanism works for anything else: when we think about who is free tonight to hang out becomes a menu of most recent people who texted us or Facebook has shown to us; when we ask about what is happening in the world becomes a menu of news feed stories chosen by Google or other news’ aggregators; when we start to look for a partner becomes a menu of faces to swipe on Tinder, instead of local events with friends, or urban adventures. This is the result of personalization and algorithmic classification and filtering. This action is very subtle as we never usually ask to ourselves whether something has been excluded from the SNS's menu, or why we are being given those specific options rather than others, or whether we know the menu provider's goals or if this menu empowering our original need, or the options are actually more responding to something else: we usually try to adjust our choices on the basis of the options shown to us, very often also changing our original preference, desire, need (e.g., I know that I do not have to eat Chinese food, but in the neighborhood where I am, I cannot find alternatives to it).

To sum up, filtering and classifying algorithms personalize and hence shape the informational choice-context of individuals, by determining and structuring the options available to the subject – and the shaping of individuals' choice context is based on how they are profiled, namely, how profiling algorithms construct profiles on them.

However, algorithmic classification and filtering is not the sole algorithmic method based on profiling to personalize users' choice-context. There is another very common algorithmic personalization technique that rules the majority of our ICTs, from SNS to IoT, whose action on the subject is even more fine-grained and their effect in reshaping users' choice-context is even more powerful. They take the name of *algorithmic recommendation systems* (hereinafter: RS).

II.1.3 Algorithmic Recommendation Systems

RS is another algorithmic technique of personalization based on algorithmic profiling that we experience on a daily basis (e.g., by using services such as Netflix, Amazon, YouTube, and Spotify). How does the personalization produced by RS differ from that generated by classifying and filtering algorithms?

RSs do not just personalize the choice-context of individuals by re-structuring, classifying, and filtering the informational options available to the subjects: more intrusively, they can capture what – amongst those options – the user values mostly, and by learning on them, try to predict what are the choice-driving elements (from mere interests to gender, race, and sexual, political, and religious orientation, up to values, beliefs, and deep vulnerabilities).

An example of the logic behind RSs is how much the movies we choose to watch tell about our history as persons. Then, RSs, on the basis of this personalized predictive knowledge, target the user with highly-personalized options that are labeled as easily be re-chosen by her, as they respond to the choice-driving elements – as algorithmically inferred (i.e., knowledge-discovery profiling method).

Let me make this point clearer.

RSs try to analyze how, within a given context of options, a user values some of them, what option (e.g., a product or a service) is mostly preferred and, on the basis of the preference, predict what the user will be interested in next, namely, how it is likely she will behave in the future. So, RS analyses the user context (e.g., what the user has recently purchased or read), and, if available, a user profile (e.g., the user likes mystery novels). Then RSs specifically target the user with one or more options (e.g., books, people, movies) that are algorithmically determined of users' interest[135] – that means "most likely to be chosen – no matter of what it can cost to her". Users' interest can be determined in two ways by RS: if this recommendation is done solely by analyzing the associations between the user's past choices and the descriptions of new objects, then it is called "content-based filtering". Nevertheless, today, due to the huge processing capacity of ML and availability of user-generated contents, RS is mainly based on "collaborative filtering"[136], that means that options as items are recommended to a user based upon values assigned by other people with similar taste. RS determines users with similar taste via standard formulas for computing probabilistic correlations[137] and then

[135] G. Adomavicius *et alia*, "Incorporating contextual information in recommender systems using a multidimensional approach". ACM Transactions on Information Systems (TOIS), 23(1), pp. 103-145 and Garcia-Molina, *et alia*. "Information seeking. Communications of the ACM", 54(11), 2011, p. 121.

[136] *Ibidem.*

[137] See U. Shardanand, & P. Maes, Social information filtering: algorithms for automating ''word of mouth''. In I. R. Katz, R. Mack, L. Marks, M. B Rosson & J. Nielsen (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (CHI '95), 1995, pp. 210-217. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co. For instance, Facebook uses a collaborative filtering called EdgeRank, which particularly adds a weight to produced users' stories (i.e. links, images, comments) and relationships between people. So, depending on interaction among people, the site determines whether or not the produced story is displayed in a particular user's newsfeed. In this way, a produced story by a user will not be seen by everyone in that user's

associates profiles constructed on the basis of common choice-driving elements discovered with social categories in which the population is divided or grouped. These categories allow profiling algorithms on which RS is based to infer what are – according to macro-lenses – common choice-driving elements within a certain group, that are peculiar to one group rather than another, and with that predictive knowledge, re-target users with refined and more customized information. By this targeting action on users on the basis of their inferred profiles (based on what they value mostly correlated to what other people value mostly as well), RSs discover not just fine-grained tendencies inside constructed categories, but they can use them to further personalize the informational options displayed inside certain groups with the result of further corroborating the choice-driving elements of persons inside a category, in which the users as profiled are already inserted. This means that RSs generate a further reshaping impact on users' choice-context:

- Firstly, RSs allow the further refinement of individual's choice environment in response to changes in the target's behavior on the basis of the analysis of the target's constantly expanding data profile.
- Secondly, RSs continuously produce data feedback fueling profiling algorithms, which can itself be collected, stored and repurposed for further filtering goals.
- Thirdly, RSs monitor and can further influence population-wide trends that are identified via collective filtering, by population's categorizing and targeting.

It is important to notice that algorithmic profiling, filtering and classifying, and RSs very often work all together in ruling algorithms-based ICTs and in triggering their governance. What does it imply for our argumentation?

contact list. All stories produced by user X can be completely hidden in user Y's newsfeed, without the knowledge of both users.

II.1.4 Algorithms: The New Architects of Choice-Contexts

I have previously underlined that with the pervasive applications of ICTs and countless IoTs ruled by interconnected algorithms, everything today is increasingly digitalized, that means that everything via the process of datification involving our reality (ever more conceivable as hybrid or "onlife"[138]) can be captured and therefore embedded in information. By analyzing the main features driving the algorithmic governance, I have highlighted how the information discoverable via data on us by profiling algorithms (and data mining) leads to the construction of profiles of us on which personalization techniques work. By exploring algorithmic filtering and classifying, I have underlined the gatekeeping function of algorithms to deciding which kind of information is available to us, namely, which kind of informational content will be displayed to us. However, insofar as everything can be understood as information, from our features and relationships to our beliefs and values, as it can be captured by profiling algorithms and then used to drive the algorithmic personalization of our informational contents, algorithms have a key-function that does not concern only the way in which we understand what surrounds us, our reality, and ourselves. Specifically, whether everything is information, and information captures our reality (from physical things to thoughts, opinions, beliefs, values, and so forth), informational contents can be also understood as alternative options amongst which we can choose the reasons and the motives for our choices. Indeed, every piece of information (about a product, a friend, a story, and so forth) can embed a value and become a reason we can endorse in our choices. This means that algorithms (profiling, filtering & classifying, and RS) do not just structure our informational choice-context *stricto sensu*: they structure our onlife choice-context, by managing both our informational options *stricto sensu*, and our offline options. Although this distinction is increasingly losing its *raison d'être*, here it helps to highlight how the algorithmic governance is reshaping both the online and offline contexts of our choices. Let me clarify further.

[138] L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford 2014.

The algorithmic function in structuring of our available options does not only concern the informational environment of our ICTs, such as Internet search engines or SNS. The phenomenon of ubiquitous computing (or ubiquitous IoT) specifically refers to our onlife, deeply interconnected, and hybrid environments that invisibly embed computational devices in everyday objects and equip them with sensors that enable them to collect individuals' data without the user's active intervention or awareness and in turn to the capacity of these ICTs to interact with our environment.

This means that algorithms can restructure and reshape our environments, both online and offline. On the latter, an example can be useful and is provided by commercialized smart fridge, interconnected with our health-oriented app and SNS, which is technologically capable of changing the order for your favorite cheese to a low-fat cheese because the biometric sensor has measured that your cholesterol levels are too high. This example makes it clear how algorithms play a fundamental role in reshaping the contexts where individuals make their choices, and allows me to introduce a clearer conceptual lens to evaluate their impact on individuals' choice-context.

By highlighting how algorithms structure the options available for our choice, we can define them as the *architects* of the contexts in which we make our choices. In this sense, we can find the expression of the power of algorithmic governance in the definition of choice-architectures, and especially, as algorithms are the architects of this reshaping function, we can define the specific choice-contexts they govern and reshape as *algorithmic choice architectures*.

The term 'choice-architecture' has found particular academic attention with Cass Sunstein and Richard Thaler who have defined this design-based approach to influence people's choices and actions as a "nudge" [139], namely, a gentle way to influence individuals' choice behavior, by "organizing the context in which people make decisions".[140] Considering that human decision-making is often biased and based on heuristics[141], nudge is a way to influence people's decision-making by

[139] Thaler, & C. Sunstein, *Nudge: Improving decisions about health, wealth and happiness*. London (UK): Penguin 2009.

[140] *Ivi*, p. 428.

[141] The intellectual heritage of nudge rests in experiments in cognitive psychology which seek to understand human decision-making, finding considerable divergence between the rational actor

slightly intervening in the order of presentation of the options, e.g., by classify them according to a different priority (here the classic example is about encouraging customers to choose healthier food items, by suggesting that cafeteria managers place the healthy options more prominently – such as placing the fruit in front of the chocolate cake). In this sense, according the libertarian paternalism of Sunstein and Thaler, the users can be nudged to choose in a way that can lead them to align with those long-term goals, while not limiting their autonomy and freedom of choice.

Even if the architecting role of algorithms in re-shaping or structuring (e.g., by re-classifying) the options of contexts in which people make their choices and actions can be understood as a form of nudge, by "algorithmic choice architecture" I do not mean that the influence that algorithms are exercising can be considered as a form of gentle push, or nudge, that relies on little changes in the presentation of options to the subjects, which do not undermine their autonomy and their freedom of choice. Indeed, as it will become even clearer at the end of this chapter, the choice-architectures governed by algorithms do more than little changes in the presentation of options (and this is also intuitive, by thinking about the algorithmic techniques previously outlined). Choice-context personalization based on filtering is more than a slight intervention on the menu of options available to the subjects, inasmuch as it entails the exclusion of some options – operation that instead is not contemplated by conceptualization provided by Sunstein and Thaler about nudge. Furthermore, they do not think about nudge as a method that can affect users' autonomy and freedom to choose. Instead, as I will show in the next two sections, the algorithmic influence in architecting our choice-contexts can endanger both our autonomy, and especially our moral autonomy, and even our freedom of choice and action – impact completely refused in the libertarian paternalism of the two authors.

model of decision-making assumed in microeconomic analysis and how individuals actually make decisions due to their pervasive use of cognitive shortcuts and heuristics. Critically, much individual decision-making occurs subconsciously, passively and unreflectively rather than through active, conscious deliberation. See the works of D. Kahneman, *Thinking, Fast and* Slow. Farrar, Straus & Giroux 2011, and Kahneman, D., & Tversky, A. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47 1979, 263–91.

We may also add other reasons to distinguish the impact on individuals driven by algorithms as choice-architects from that implied by the conceptualization of Sunstein and Thaler of what can count as a nudge. First of all, here the architects or agents of nudges are not institutional agents, as in the libertarian paternalism, which aim at increasing both individual and collective well-being, but algorithmic agents that are in charge of third-party interests – as those of advertising companies and political parties (this has been clearly unveiled by the case of Cambridge Analytica, when Russian political parties have paid to exploit people's personal data to direct micro-targeted informational ads via Facebook in order to influence their political vote in the 2016 U.S election) – and aim at increasing just economic profit.

Secondly, the degree of algorithmic impact is incomparably higher than that of any institutional nudges; for the pervasive nature of the algorithms, it is impossible to describe it as 'gentle'.[142] Furthermore, the algorithmic influence does not come from "above" (e.g., from a political entity or an institution) and is not perceptible as a visible subjective imposition (e.g., a new institutional policy that you must comply with), but it is invisible, surrounds us, and as conveyed by technology that is usually perceived as objective and neutral, in the majority of the cases, this impact is not perceived as an interference or an imposition on us by someone or something, but – when and if it is perceived – it is more understood as a context-reshaping effect produced by ourselves and specifically by how we interact with ICTs.

The last but not the least, if the nudging actions operated by institutions have the purpose to influence short-term individuals' choices in order to improve their long-term well-being, the goal of algorithms as choice-architects is the short-term boosting of the click and profit maximization, e.g., boosting the economic income of such companies (or political parties) that pay online service providers to show and the informational options that reflect their economic and political long-term goals.

[142] Furthermore, the impact of algorithmic choice-architecture is even less comparable with those producible by any other technology, given the fact that we are completely immersed in this informational environment permeated by algorithms. To understand this point, we can think the difference between the impact of television (which diffuses information and shapes people's opinions) and that of Internet: we do not live "on television," we watch it, or we listen to it; vice-versa, we are on the Internet, we live completely immersed in this new hybrid environment).

Beyond this key-clarification, it is important to notice that thinking about the algorithmic governance as forming the algorithmic choice-architectures helps us to deepen how powerful and intrusive is the algorithmic impact, as it does not just re-structure our choice-context, but it also affects our freedom to choose and act as moral agents, by undermining the necessary conditions underlying its exercise.

I may indeed go on in underlying the differences between the nudge as theorized by the libertarian paternalism and the architecting function of algorithms here analyzed; however, the next chapters do this work, by implicitly showing how much the impact of algorithmic choice-architectures differs from that as intended in the paternalistic theory of nudge by affecting our autonomy and our freedom of choice and action as moral agents. Specifically, I will devote the next two sections in exploring and arguing this specific algorithmic impact on the *conditiones sine qua non* underlying the exercise of our moral freedom, as highlighted in section 1.2, namely, a) the availability of morally heterogeneous options and b) moral autonomy as relational self-determination.

II.2 The Epistemological Problem of Shaping Choice's Options

As underlined in the first section of this chapter, automated profiling ruling algorithmic ICTs is one of the most common knowledge-discovery methods used by providers to fuel and steer filtering algorithms towards the classification of informational contents (or options) and so the creation or shaping of "personalized" choice-contexts – that I define as algorithmic choice-architectures – on the basis of insights and patterns (considered as predictively valuable ways of grouping persons on just surprising similarities) probabilistically inferred and discovered within data-sets. In this section, I explore how algorithmic ICTs based on algorithmic profiling can undermine the first key-condition underlying the exercise of our moral freedom, namely, the "availability of morally heterogeneous options", by affecting and specifically weakening the epistemological position of the profiled individual.

I frame the impact of algorithmic choice-architecture on the first necessary condition of our moral freedom – as freedom of choice and action as moral agents – as "the epistemological problem of algorithmically shaping of choice's options".

This impact can be explored according to different epistemological levels related to different scenarios created by the complexity of algorithmic functioning.

First of all, this algorithmic impact can be framed both in *quantitative* and *qualitative* terms, although it is intuitively clear that the latter is the most critical with regard to the first necessary condition underlying the exercise of our moral freedom, i.e., the availability of morally heterogeneous options.

The quantitative impact raised by algorithmic choice-architecture on the availability of alternative options is quite intuitive. Algorithms are information gatekeepers: by structuring our informational environment, and hence what piece of information does not just come first but is available or unavailable to us (i.e., they exclude what is probabilistically inferred as not relevant for us as profiled), they unavoidably reduce the number of informational contents available to us, where – as previously argued – informational contents can embed values, reasons, desires, beliefs, and therefore everything can be conceivable as an option standing for a potential alternative course of action. In this sense, a first potential constraint raised by algorithms on our moral freedom, and especially on its underlying first condition of possibility, is quantitative, as the available number of alternative options is concretely reshaped and above all limited by classifying and filtering algorithms.

Nevertheless, this filtering action is very often justified by tech-providers as 1) *necessary*, specifically, to overcome the problem of informational overload, and 2) implemented in the *subjects' interest*, i.e., it is executed to respond to the users' implicit and explicit interests (as probabilistically inferred by data), because vice-versa, i.e., a large availability of options (as informational contents) would risk to overwhelm the subject, and therefore to hamper rather than foster individuals in the process of choosing what responds to their best interest. In this regard, indeed, there are contentious arguments provided by philosophers, as well as behavioral economists, who argue indeed both in favor or disfavor the too-many-options condition for individuals' choice and their decision-making.[143]

[143] Researchers across various disciplines have found that the performance of a decision-making (i.e., the quality of decisions or reasoning in general) of an individual correlates positively with the amount of information he or she receives, up to a certain point. If further information is provided beyond this point, the performance of the individual will rapidly decline. See, for example, M.J.

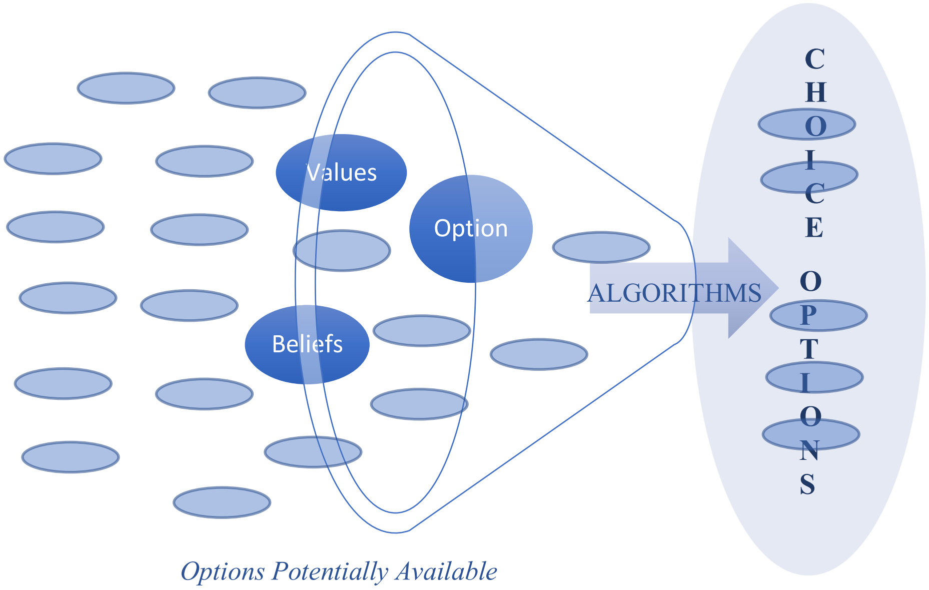
However, here, what is interesting is that, beyond the number of options that are available to the user for the preparation and making of her choice, this limiting action is performed by algorithms, thus, is a *hetero-determination* of the availability of options to the subjects. Therefore, beyond the number of options available, what is really critical is that what is excluded and what is instead part of our choice-context is algorithmically determined, i.e., heteronomously-determined, by external factors on which we are just partially in control. In other words, algorithms are the external architects which determine what option (which in turn can embed values, reasons, and therefore constitutes a potential alternative course of action) is part of my menu of choice alternative: what option – from being *potentially* considerable into my choice-context of options – is *effectively determined* as part of my choice-menu of options, is effectively part of my availability of alternative options.

This is a heteronomous predetermination of our informational environment and choice-context by default: algorithms, by selecting the relevant information for us, predetermine the conditions of our choices, by restricting the range of available options, and this action is based on profiles that are probabilistically predetermined.

This is another way to consider the personalization of users' informational environment. I claim that this algorithmic personalization of the options available

Eppler, & J. Mengis, "The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines." *The Information Society*, 20(5), 2004, pp. 325–344R. Thaler, & C. Sunstein, *Nudge: Improving decisions about health, wealth and happiness*. London (UK): Penguin 2009. D. Kahneman, & A. Tversky, "Prospect Theory: An Analysis of Decision Under Risk". *Econometrica*, 47, 1979, pp. 263–91. D. Kahneman. *Thinking, Fast and* Slow. Farrar, Straus & Giroux 2011. D. Kahneman, & A., Tversky. (eds.), *Choices, Values and Frames*. Cambridge University Press 2000. To expand the concept of bounded rationality, see M. Hilbert, "Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making, Psychological Bulletin", 138(2), 2021, pp. 211–237.

to us as profiles should be better understood as a *hetero-definition or algorithmic pre-determination of our availability of choice-options*.



Options Potentially Available

Options algorithmically actualized or pre-determined as available in the user's choice-menu

- **Figure 2.** Algorithmic personalization as hetero-determination of the subject's availability of options.

Indeed, as I mentioned above, our "control" in this hetero-determination of our choice-context is very limited and can only applies to the options displayed, that is, to the options available in my choice-context, that are in turn chosen on the basis of constructed profiles that are aligned to us as persons. My person as profiled, indeed, is represented as a set of interests, needs, values, characteristics, goals (and so forth…) that is probabilistically inferred, and therefore – as before explained – discovered by comparing the data I enter directly or indirectly with other huge sets of data available presenting similar characteristics.

In this regard, it is possible to develop two scenarios, but in both of them we will see how algorithmic-choice architectures can undermine our first c*onditio sine qua non* underlying the exercise of our moral freedom, i.e., the availability of morally heterogeneous options, by hetero-determining our choice-contexts.

Let me clarify this point.

The first scenario is an algorithmic choice-architecture where the profiles of the individuals as inferred from probabilistic assumptions (on which classification and filtering algorithms will base the determination of the options that are available to them) reflect *effectively* the subjects' values, reasons, beliefs, goals (and so forth). This is possible thanks to the large quantity of data available on us deployable by profiling algorithms to construct highly accurate or personalized profiles on us as persons (let us think about RSs and their capacity to capture the driving elements of our choices and via feedback retailor our profiles).

In this case, profiling algorithms steer classifying and filtering algorithms to align the determination of the kind of informational options available towards users' well-captured values, reasons, beliefs, goals. In this sense, the subjects could result fostered in their choices and actions by the algorithmic action, insofar as the algorithmic hetero-definition of options would result aligned to their values, beliefs, intentions, in other words, to their moral orientation. If the moral orientation of the subject is not formed yet, the algorithmic hetero-definition of her options can be based to the well-captured moral dispositions of the subject, i.e., moral disposition as *diàthesis*, at the root of our moral identity, as *héxis*. The moral disposition is indeed what can be consolidated via certain choices and actions over time in the formation of moral posture or the development of moral identity of the individual (the moral disposition is how I react or take a moral stand to reality and morally-loaded events, when my ought to, i.e., the dimension of my obligation, is not formed yet). Insofar as ML are continuously self-learning from subjects' data input, ML algorithms would be able to tailor, re-fine, and personalize continuously the options available to the subjects in a way, according to this first scenario, aligned to the users' capacity of moral changing.

In short: the hetero-determination would be aligned to the *current* and *future* subject's moral orientation. So, the hetero-determination of the options available to the users would reflect a sort of self-determination of those options by the subject. Following this reasoning, the hetero-determination would not pose a real limiting or hindering impact on the availability of alternative options, as they reflect users' moral orientation (values, goals, projects…) probabilistically well-inferred. In simple terms, the options presented would be the same the users would have chosen if they would have time and resources to perform a similar filtering operation.

This is the main ethical justification for the deployment of invasive profiling algorithms and classifying techniques provided by tech providers: algorithms foster the users to better exercise their choices by presenting them what is aligned to their interests and therefore as more relevant for them.[144]

Unfortunately, whilst apparently this hetero-determination may increase the capacity of the subject to choose and act according to her values, goals and so forth, actually, it affects this first *conditio* at its core, in as far as it undermines qualitatively our availability of alternative options, i.e., the alternativity factor, the qualitative alternativity of the options available to the subject, whereby qualitative alternativity I mean specifically, as argued in the previous chapter, the *moral heterogeneity* of the available options that is peculiar to this condition. By highlighting this, we point to the other transformation mentioned above, that is the *qualitative* one. Algorithmic choice-architectures do not just affect and specifically predetermine the number of options available to the subject in her choice-menu, but also the qualitative nature of those options.

Nicholas Negroponte has described the mechanism underlying this impact a long time ago, by prophesizing the emergence of the 'daily me' phenomenon to specifically indicate a digital personalized package of information, namely, a digital environment of items, personally designed with informational contents fully chosen in advance so to respond to the predicted characteristics and preferences of a certain user. He referred to the economy and the business sector, and highlighted how our digital can help us to satisfy our preferences in the market, by selecting alternative options among which we can choose as consumers.

As we discussed previously, with the pervasive application of algorithms, and especially of algorithmic profiling and classification, the 'daily me' phenomenon has been amplified and rules all our algorithmic ICTs (from SNSs to the countless applications of IoTs): it works in the selection of what is algorithmically understood

[144] It has been uncovered how this justification is more an ethical facade that hides the significant economic interests that lie behind this technical operation. In this regard, see S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York, NY (USA) 2019: Public Affairs.

as a relevant option for me (whatever it is a piece of information, a product, a movie, a friend, and the list could be expanded much further), whereby relevant is meant 'consistent' with the information I liked or searched for in the past, or as I clarified earlier, consistent to the profile of me as probabilistically inferred and constructed.

This alleged choice-enabling algorithmic functioning has been uncovered to undermine our exposure to different points of view, beliefs, values, and relations to the point to create what Eli Pariser calls as "filter bubbles"[145], or Cass Sunstein defines as "informational silos" or "informational echo-chambers"[146], i.e., environments characterised by like-minded people, hence, people with similar beliefs, orientations, and values, therefore, environments characterized by "morally homogeneous" options, that is, environments characterized by like-minded persons who express similar values, ideas of the good, and moral practices – and therefore groups characterised internally by a moral homogeneity or similarity (in values, ideas of the good, and so forth). Furthermore, this phenomenon has been criticized to shape our social interactions in a way that tend not to expand them but to narrow them in ways that very often produce polarization.

Paradoxically, whilst algorithms may vastly increase the number of people, points of views, opinions, beliefs, values (and so forth) we can encounter, by globally exposing ourselves to information about people with other cultures, values, and ways to do things – that is crucial to open the possibility of wondering whether the moral rules, values, and practices we are following are optimal or eventually to change them –, actually, classification and filtering depending on algorithmic profiling can determine a narrowing of social exposure, by shaping our available informational options (and therefore who to get in touch with and what piece of information see) on the profiles probabilistically discovered as similar or just like ourselves. This can lead us to encounter "those of exactly the same opinion sets as our own"[147] and tends to "make us more prejudiced and our attitudes more insular"[148], by ultimately leading to radicalize our previous orientation, instead of

[145] E. Pariser, *The Filter Bubble*, Penguin. 2011.

[146] C. Sunstein. "Democracy and the Internet". In J. van den Hoven & J. Weckert (Eds.), *Information Technology and Moral Philosophy*. Cambridge University Press, 2008, pp. 93–110.

[147] M. Parsell, "Pernicious Virtual Communities: Identity, Polarisation and the Web 2.0", in *Ethics and Information Technology*, 10 (1) 2008, p. 43.

[148] *Ibidem*.

critically challenge it.

This phenomenon influences the way in which we develop our identity and tend to create polarization and social division. In fact, phenomena as polarisation and social cascades are likely to occur more often when people only engage in relationships with those who are similar to them as they "are likely to move toward a more extreme point, in the direction to which they were previously inclined" and are likely to "end up thinking the same thing that they thought before – but in more extreme forms".[149] This lack of heterogeneity of relations, points of views, and orientations is a lack of heterogeneous reasons and diversified ideas on what is good and undermine the possibility to challenge and reasoning our moral orientation, and therefore the possibility to develop genuine reasons and values, and hence genuine moral identity, by instead easily leading us to increasingly develop self-enclosed reasons, values, and identity.

To summarize, at the core of the critical phenomena mentioned here, there is what I define the impact of algorithmic choice-architectures on the moral heterogeneity of our available options, which become hetero-determined and personalized options, chosen by algorithms as relevant to the user. In this sense, I claim that even when the options are shaped on the basis of values and reasons that may reflect the moral orientation of the users, the echo-chamber or bubble produced deeply undermines the moral heterogeneity of options – as embedding values and moral reasons – that is crucial for the subject to develop her own idea of the good, values and reasons and then critically test and endorse them so to act and choose as a genuine moral agent, and to develop a genuine moral identity, that is, to become a genuine moral agent.

I will further analyze the ethical implications of the algorithmic impact on our moral freedom in the first section of the third chapter. For now, it suffices to

[149] C. Sunstein. "Democracy and the Internet". In J. van den Hoven & J. Weckert (Eds.), *Information Technology and Moral Philosophy*. Cambridge University Press, 2008, p. 99. Sunstein underlines how people want to be perceived favourably by other group members, and so they often adjust their position in the direction of the dominant position (and how this adjustment often does not depend on a rational choice) and the outcome of this adjustment is that both the group, as a collectivity, and its members, as individuals, tend to support positions that become more and more self-enclosed.

highlight how reducing the heterogeneous expositions of the users' informational environment undermines the possibility for the subjects to develop their own idea of the good, their moral values, and moral ground projects in a genuine way, and therefore in a way that requires the encounter of heterogeneous values, beliefs, reasons (and so forth) crucial to act and choose as genuine moral agents. Indeed, in order to develop morally genuine identity, agents need to be able to critically form, expose, and test their moral orientation, and so those values, reasons (and so forth), she will endorse as motives for their choices and actions and on which will steer the development of her moral identity.

I define this impact of algorithmic choice-architecture on our availability of morally heterogeneous options as an *epistemological impact* on the subjects' freedom to choose and act as moral agents, as it affects the way in which the subject develops her own idea of good, her values and moral reasons, namely, her formation and critical reasoning on her *moral knowledge*, i.e., of those moral reasons and values she can endorse as motives for her choices and actions, on which she steers the development of her moral identity, in a genuine way. The algorithmic hetero-determination of options raises an epistemological problem on the subject's freedom of choice and action as a moral agent, as it affects the formation and reasoning on her *moral knowledge* i.e., of those moral reasons and values she can genuinely endorse as motives for their choices and her actions.

However, as I will further explain in the first section of the next chapter, this algorithmic impact as described in this first scenario, although it is anyway freedom-undermining insofar as it produces an interference to our moral freedom by undermining the dimension of moral heterogeneity of the options available to us, can be specifically defined as a *moral interference* (according the definition of moral interference I gave in the negative concept of moral freedom), inasmuch as it is conducted in a way that keeps track of agents' interests and values, i.e., their moral orientation.[150] However, there is also another kind of epistemological impact on agents' freedom to choose and act raised by algorithms, which also concerns the hetero-determination of our availability of morally heterogeneous options.

[150] In the third chapter, I will specifically deepen these algorithmic interferences in the light of the conceptualization of moral freedom developed in chapter 1.

I have considered a first scenario, where the profiles according to which our informational environment is shaped and determined effectively reflect – broadly speaking – our moral orientation, and I highlighted that even in this case, where the hetero-determination may be considered as aligned to our values and reasons, the first condition underlying our moral freedom is endangered, specifically, the moral heterogeneity of the options predetermined for us is deeply jeopardized.

There is also a second scenario, where the profiles developed for us, as based on de-individualized probabilistic assumptions, do not reflect our moral orientation. Let me clarify this scenario and show how this second case results as well into a hetero-determination of the availability of morally heterogeneous options that is particularly problematic for our moral freedom, namely, for us to choose and act as moral agents.

An often-repeated worry[151] in the ML context is that algorithmic profiling can also ignore the individuality of people, their consideration as particular persons, because it relies on patterns whose predictive values is imperfect.[152] For example, a person can be subject to an adverse decision, such as being denied credit, simply in virtue of being similarly profiled to persons who are not credit-worthy.[153]

This worry is legitimate even when such decisions are based on correlations and patterns that – even when they are not fully reliable – appear to be reasonable (for instance, following the previous example, because they link credit-worthiness to the employment-history of the person profiled). Algorithmic profiling, however, becomes particularly problematic when it is based on seemingly arbitrary ways of categorizing and grouping persons, as we underlined in the first chapter, such as groupings based on intuitively irrelevant features or properties. Not only one faces an unjust of wrong decision because the predictive accuracy of the grouping is

[151] M. Leese, "The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards". *The European Union. Security Dialogue*, 45(5), 2014, p. 502. B.W. Schermer, "Risks of profiling and the limits of data protection law". In: Custers, B, Calders, T, Schermer, B, et al. (eds) *Discrimination and Privacy in the Information Society*. Berlin: Springer, 2013, pp. 137–152.

[152] This raises concerns about potentially unfair and discriminatory profiling and on the socially undesirable consequences it can produce. See S. Gutwirth, and M. Hildebrandt, *Data Protection in a Profiled World*. Erasmus University Rotterdam; Springer Netherlands 2010.

[153] A. Vedder, "KDD: The Challenge to Individualism." *Ethics and Information Technology*, 1 (4), 1999, pp. 275–81.

limited, as the example underlines, but one is also unable to predict that, or even understand why, she should be informationally profiled and categorized as such.

Worries of this second kind are traditionally associated with ML because its functioning uncovers previously unnoticed but potentially valuable patterns. At the same time, the critique that profiling can lead to situations where persons are grouped in a way that is arbitrary, perhaps even unreasonable or plainly wrong, raises a second kind of epistemological impact on the profiled person, beyond the one previously defined as an hetero-determination of options in a way that limits moral exposure of the agents to heterogeneous relations, reasons, and values, namely, the moral knowledge on the basis of which we choose and act as genuine moral agents.

This further epistemological impact on individuals can be understood in two ways. Firstly, algorithmic profiling can raise an epistemological impact on agents by creating epistemological asymmetries *stricto sensu*. This happens specifically when the algorithmic probabilistic assumptions on which the subject is profiled and categorized into a certain group are inexact and de-individualize the subject itself. In this case, the epistemological impact of profiling algorithms underlying the way in which options are hetero-determined and make available to persons, depending on their consideration as specific profiles, concerns the epistemological asymmetries (power-asymmetries) between the profiling algorithms and the users as profiled. In this case, not only we can experience a limitation of available options by algorithms, inasmuch as they are hetero-determined and chosen in advance by algorithms on the basis of the profile of us probabilistically discovered; moreover, this limitation is also based on probabilistic assumptions which can categorize and classify ourselves as profiles into a wrong group, by associating our profiles with those of others presenting featured inexactly inferred as relevant. This results in an epistemological asymmetry between who I am as a moral agent, and therefore, my values, my attachments, moral ground projects, beliefs, and so forth, and how I have been profiled, known, and described by algorithms, and on the basis of this profile, re-influenced by the shaping of my informational environment. In other terms, the algorithmic knowledge on me – as profiled – on which classifying and filtering algorithms base the pre-determination of my available options does not reflect me as a particular person: algorithms reshape my availability of options according to that constructed profile on patterns probabilistically inferred as valuable, instead of me as a particular person.

Whether in the first scenario discussed the hetero-determination of options in a way aligned to subjects' moral orientation raised the problem of affecting the moral heterogeneity of options available to the subject, this second scenario shows that the solution to the previous problem does not lie in the generalization or de-individualization of inferences and profiles.

The problem of excessive personalization that can lead to undermining the exposure of the individuals to different points of views cannot be solved by de-individualizing the algorithmic functioning. As indeed I show in the first section of the next chapter, this second kind of profiling does not just produce an interference to our moral freedom (that in the first scenario, as the algorithmic interference is aligned to agents' interests, values, and broadly speaking moral orientation, it can be defined as moral), but produces also an immoral interference on our moral freedom, inasmuch as is illegitimate, as based on inexact, wrong or even unjust (as often biased) assumptions.

This second scenario is very perilous, as this illegitimate algorithmic impact cannot just reshape and limit informational options, but also affects alternative courses of actions as opportunities and real life-chances. This is as an algorithmic impact on the first condition of our moral freedom, as this is a hetero-determination of availability of options as informational contents and real chances, opportunities, and alternative courses of actions.

Indeed, in this case, it is also more evident how on the basis of algorithmic profiling not just the informational options of the user profiled can be narrowed, but also what can derive from those informational options is affected, such as alternatives, opportunities, and real chances for the person who is subject to the algorithmic profiling. To follow the example on creditworthiness previously mentioned, if I am defined as not credit-worthy on the basis of the wrong assumptions (e.g., my online history shows that the majority of my friend has a low-credit score), I lose the access to real life-chance, as the access to that landing necessary for saving my economic business, buying a house, or be able to pay the university fees to my children. This means that that this kind of inference is beyond freedom-undermining, also deeply morally illegitimate.

Here, the epistemological impact raised by algorithms on individuals can be observed according to a second standpoint. Indeed, persons who are categorized and grouped in arbitrary (or in increasingly complex) ways are often unable to predict, understand, or contest the decisions they are subjected to. In this sense, the epistemological asymmetry is a real asymmetry in terms of power between the profiler and the person profiled.

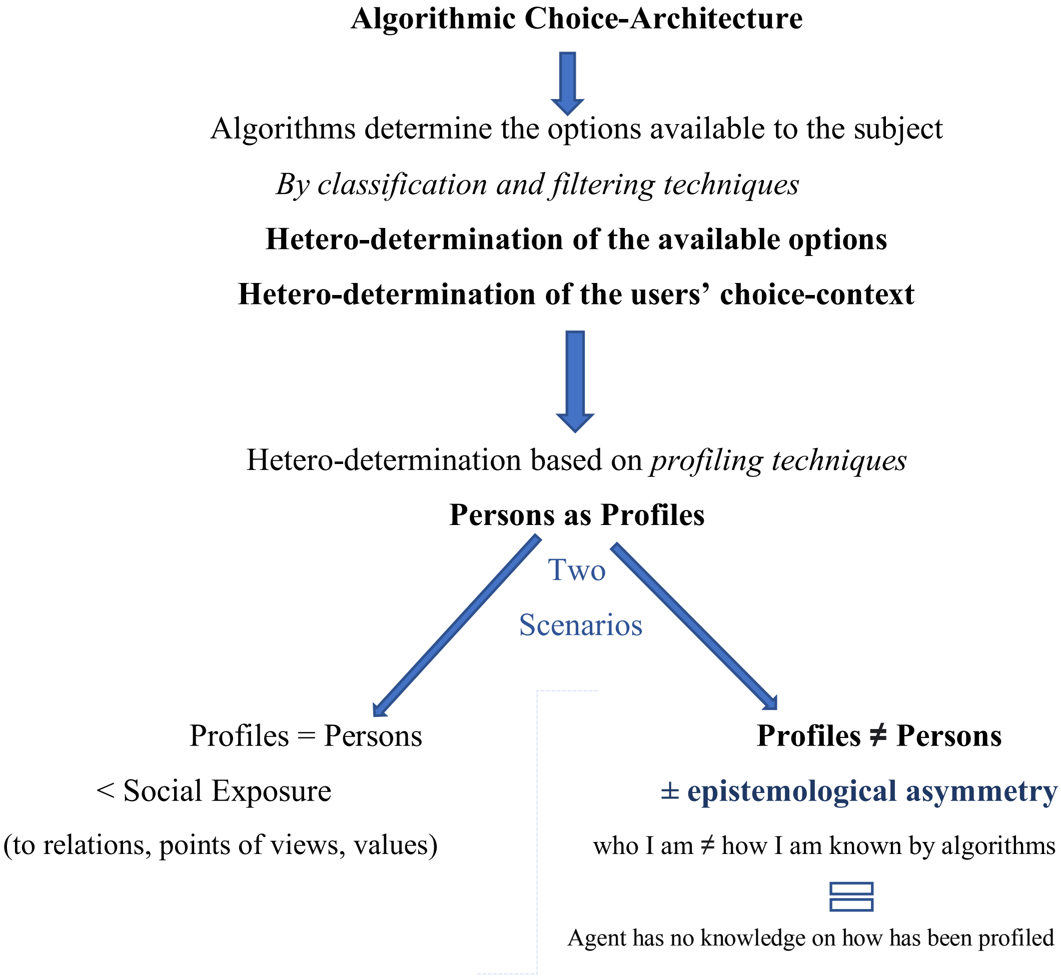
This specific hetero-determination, that I defined as illegitimate as based on wrong, irrelevant, or inexact algorithmic assumptions on individuals as particular persons, deeply weakens the epistemological position of persons as agents, insofar as the assumptions on which algorithms base their action are opaque to us: we do not have knowledge on them and therefore, without knowledge, we do not have the power to intervene on them.

In this sense, I define this impact of profiling algorithms on individuals' options as a second kind of epistemological impact on individuals, as it weakens the epistemological position of the "decisional" agent, who is made unaware and passive towards her choice-context, as she cannot know and intervene on the options algorithmically pre-determined – options which in turn lead agents to a context of pre-determined alternative possibilities, opportunities, life-chances, and courses of actions. In this sense, this is not just an impact on persons' moral knowledge, but also on the epistemological level of persons' autonomy, as they lack the required knowledge to be the agents of their lives.

To sum up, this algorithmic impact is a further epistemological impact on our moral freedom, because not just the hetero-determination raised by algorithms pre-selects and narrows the moral heterogeneity of the options available to us, but this operation can be conducted according to profiles constructed on de-individualized patterns and associations that can be wrong or inexact, thus creating an epistemological asymmetry *stricto sensu*, i.e., between how we are as moral agents and how we are known as algorithmically profiled. This epistemological asymmetry is moreover also an asymmetry of power, insofar as by being unaware of the associations and assumptions driving the discovery of our profiles, we are unable to act on them: we have no power on them, neither cognitive power nor power of action.[154]

We have mainly analyzed so far the functioning of classifying and filtering algorithms that are based on algorithmic profiling and we have shown how they can influence and epistemologically undermine the subject (as profiled), by affecting the first condition of possibility underlying the exercise of moral freedom: the availability of morally heterogeneous options.

To sum up, this epistemological impact on agents' first necessary condition can be schematized as follows:



Algorithmic Choice-Architecture

Algorithms determine the options available to the subject

By classification and filtering techniques

Hetero-determination of the available options

Hetero-determination of the users' choice-context

Hetero-determination based on *profiling techniques*

Persons as Profiles

Two Scenarios

Profiles = Persons

< Social Exposure

(to relations, points of views, values)

Profiles ≠ Persons

± epistemological asymmetry

who I am ≠ how I am known by algorithms

=

Agent has no knowledge on how has been profiled

[154] This algorithmic impact raises serious issues also in terms of justice and fairness, inasmuch as whether profiling algorithms categorize us on wrong assumptions, they can easily produce unfair outcomes, such as being treated in ways that can discriminate us. This is a very important topic that deserves to be treated per se, but as I specified in the introduction (section: caveat) I do not address in this work.

< Critical Reasoning
(on moral values and reasons)
< Moral Heterogeneity
= Moral Echo-Chambers

± asymmetry of power
=
Agent as no power of action on her profile

It is important to underline that the epistemological impact just described is not just an impact on the first necessary condition underlying our moral freedom, i.e., our availability of morally heterogeneous options. This impact can go deeper and affects also our moral autonomy, as we will see in a while.

Indeed, there is also another algorithmic technique that is largely applied in combination with the two algorithms previously analyzed that is key to the establishment of algorithmic choice-architectures: the algorithmic recommendation systems.

In the next and last section, I specifically focus on algorithmic RSs to show how the impact of algorithmic choice-architectures can affect also the second *conditio sine qua non* to secure at a minimum threshold the exercise of our moral freedom, by going beyond users' epistemological level and affecting specifically their moral autonomy as relational self-determination, namely, agents' reflective endorsement.

II.3 Between Nudge and Push: Is It Possible to Suspend Moral Autonomy?

Before proceeding with the argumentation regarding the second necessary condition (*conditio sine qua non*) underlying the exercise of our moral freedom, that is, our moral autonomy, it is necessary to provide an overview of the ways in which the current literature in ethics of AI and algorithms has been thinking of algorithmic ICTs and human autonomy.

The rise of algorithmic governance has recently spurred the debate in ethics of AI on the impact of algorithms-based ICTs on the autonomy afforded to users[155], even though the concept of autonomy is very often taken for granted as referring to a general condition of self-governance or self-determination of individuals. Furthermore, the properly moral dimension of autonomy has not been adequately conceptualized in the literature on ICTs and AI. This shortcoming is visible by looking at the debate in ethics of AI and algorithms developed so far.

In the contemporary literature we can identify at least three main different but strictly interconnected ways in which algorithms-based ICTs have been described as potentially limiting users' autonomy.

The first *limit* posed by algorithmic ICTs to users' autonomy stems from the users' inability to understand algorithms' model and functioning. In fact, as noticed by Shin and Park, algorithms "do not have the affordance that would allow users to understand them or how to best utilize them to achieve their goals"[156] and this can constitute a limit to individuals in the understanding of some information produced by algorithms-based ICTs and therefore in making appropriate decisions on them. As a consequence, it also turns out to be complex for users to strike a balance between relying on their own decision-making and how much of that to delegate to algorithms (an issue further complicated by a lack of transparency that very often connotes algorithmic decision-making to which human decisions, choices, and task are increasingly delegated).[157] In this way, individuals end to blindly delegate more

[155] For a general debate on algorithms and human autonomy see M. Ananny and K. Crawford, "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability". *New Media & Society*, 17 (1), 2016, pp. 973-989. D. Beer, "The Social Power of Algorithm". *Information, Communication & Society*, 20 (1), 2017, pp 1–13; L. Floridi & M. Taddeo, "How AI Can Be a Force for Good", *Science*, 361 (6404), 2018, pp. 751–52. J. Möller et al, "Do Not Blame It on the Algorithm: An Empirical Assessment of Multiple Recommender Systems and Their Impact on Content Diversity", *Information, Communication & Society*, 21 (7), 2018, pp. 959–77. T. Hauer, "Society Caught in a Labyrinth of Algorithms: Disputes, Promises, and Limitations of the New Order of Things", *Society*, 56 (3), pp. 222–30; C., Malhotra et al., "ETHICAL FRAMEWORK FOR MACHINE LEARNING", In *2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K)*, 2018, pp. 1–8. Santa Fe: IEEE.

[156] D. Shin, & Y. G, Park. "Role of Fairness, Accountability, and Transparency in Algorithmic Affordance". *Computers in Human Behavior*, 98, 2019, pp. 277–84.

[157] For a preliminary understanding of the problem of users' autonomy and transparency connected to algorithms' complexity, see M. Ananny and K. Crawford, "Seeing without knowing: Limitations

and more of their tasks, decisions, and choices to algorithms, and this delegation results in an erosion of their autonomy as a gradual loosing of their skills and mental and physical abilities.

This threat is understandable just by thinking about the increasing algorithmic application with human out-of-the-loop, that is, the increasing delegation of tasks and decisions to algorithms full automated and so more and more autonomous in their functioning. The shift from human in the loop to human out of the loop has also occurred due to the increasing amount of information from various sources and devices that has to be integrated and subsequently interpreted to come to a decision in a wide array of domains from policing to healthcare. Algorithms can do this far more efficiently and effectively than humans, for whom it is almost impossible. As a result, people no longer make the decisions themselves but leave it to algorithms. Examples include knowledge systems that make medical diagnoses based on a large amount of information, military robots that take life or death decisions using information from various sources, and the driver support systems that decide what speed we should drive on a particular stretch of road: in all these cases there are ADM.

In this first debate, human autonomy lies in human control, where the control expresses in holding and performing skills, as well as mental and physical abilities. The erosion of autonomy is therefore understood as human de-skill and capacity loss. Thus, no moral dimension of autonomy is evoked.

The second *limit* posed by algorithms to users' autonomy stems from a users' lack of power (or appeals) over algorithmic outcomes, namely, the difficulty to make the decisions and outputs of algorithmic ICTs, that – as we have underlined in the previous section – shape our informational environment, as constrainable and reversible. This limit is strictly connected to the previous one, as it is generated by the opaqueness of algorithmic ICTs to users' understanding, often due to their not-explainability (black boxes)[158] linked to their high complexity in functioning or to

of the transparency ideal and its application to algorithmic accountability". *New Media & Society*, 17 (1), 2016.

[158] On the critical issue of algorithms as 'black boxes', the main source is F. Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA (USA): Harvard University Press 2015.

their obfuscation as a design choice of their providers. It is indeed problematic that the European Union (EU) General Data Protection Regulation does not include an explicit 'right to an explanation' when decisions affecting people are reached by autonomous or semi-autonomous algorithmic ICTs.[159]

This is an algorithmic impact on human autonomy in terms of lack of control of users (on what can affect them) and above all on their right to self-determination.

In this second debate, the highlighted impact of algorithms-based ICTs on autonomy in terms of a lack of power to appeal over algorithms' outcomes brings out a moral connotation of human autonomy, that goes beyond the previous one, as the control in terms of exercise of skills and abilities to perform tasks and activities, and specifically refers to autonomy as human capacity of self-determination, where the right of self-determination implies human dignity.

The debate on this critical impact posed by algorithmic ICTs on autonomy does not further explore this moral dimension of autonomy in its consideration as self-determination. Nevertheless, we can further define this limit as connected to the one previously argued: this lack of understanding and knowledge of users on algorithmic functioning and specifically on what drives their outcomes is what we have previously defined as an algorithmic epistemological impact on agents that produces an asymmetry of knowledge and in power between algorithms and individuals that are made passive and unaware on the reasons behind the way in which they are algorithmically known (profiled) and treated (categorized and grouped). I have indeed explored how algorithms reshape individuals' options (both as information and as real chances, opportunities, and alternative courses of life) on the basis of the probabilistically way in which they profile and categorize people and how this end on weaken epistemologically and in power the individuals subject to algorithmic outcomes, i.e., to know, assess, and contest them.

As previously mentioned, this algorithmic impact on users is not just at the epistemological level (although it raises a critical epistemological problem) as an

[159] In this regard, myself and B. Giovanola explore this problematic aspect both for human autonomy and especially for fairness in AI and algorithmic decision-making (ADM) in *Weapons for Moral Construction? On the Value of Fairness in Algorithmic Decision-Making* (forthcoming), by defining the 'right to justification' as a constitutive component of fairness in ADM and introducing a specific corresponding criterion to achieve it through the design of algorithms-based ICTs.

interference to our possibility to act and choose alternatively from what it displayed to us, i.e., the options pre-determined for us as profiled, but raises also an impact on individuals' autonomy and specifically on users' self-determination and dignity. Indeed, "no relation should exist that cannot be adequately justified toward those involved"[160], as it would constitute a disrespect of people as equal end-setters (disrespect of their dignity by virtue of their humanity). Algorithmic opacity creates a lack of justification as a lack of knowledge and so an impossibility for individuals to appeal over algorithmic outputs. This lack of knowledge about what moves or is behind algorithmic decisions undermines people's right to self-determination and their dignity as end-setters, inasmuch as it hinders the informational conditions that enable personal agency and self-realization. As I previously argued, this lack in knowledge is a lack of power owned by individuals to know why certain options have been excluded to them and on the basis of what: this is of crucial importance, inasmuch as the available options to us, both intended as information and as real life chances, define how we can act (between what alternatives) and therefore how we can develop our identity and realize ourselves. Here the moral connotation of human autonomy emerges clearly and lies in the right of persons to self-determine themselves as final end-setter on their choices, actions, behavior, and identity. As mentioned, the debate on ethics of AI does not define this moral connotation of autonomy but it is nonetheless fundamental to be clarified here inasmuch as it is strictly connected to my argumentation, although – as I have conceptualized in the second section of the first chapter of the work – the concept of moral autonomy I endorse and develop as a condition of moral freedom goes beyond the consideration of autonomy just in terms of self-determination in order to encompass also the social and relational dimension.

There is also a *third* way in which algorithms-based ICTs have been described to affect and limit human autonomy in the debate in ethics of AI and algorithms and it is connected to their pervasive distribution and governance (what I have defined in the first section of this chapter as the rise of "algorithmic choice- architectures") and above all is connected to their pro-active capacity of learning from users' data

[160] Forst, R. (2014). Two Pictures of Justice. In *Justice, Democracy and the Right to Justification. Rainer Forst in Dialogue* (pp. 3-26). London: Bloomsbury.

to influence and shape users' choice. Although this specific impact is mentioned in the current literature[161], it is rarely analyzed in depth. Mittelstadt *et al.*[162] and Tsamados *et al.*[163] in their critical overview about the ethical problems raised by algorithms recognize that the predictive capacity of algorithms can undermine the way in which we make our choices by affecting human autonomy and that an ethical inquiry into human autonomy, and especially into the moral value of autonomy, is largely missing in the debate developed so far. The importance of human autonomy as a moral value is also acknowledged in many high-level initiatives on ethical principles for AI, including, *inter alia*, those of the European Commission's European Group on Ethics in Science and Technologies, the UK's House of Lords Artificial Intelligence Committee, and also – by looking beyond the West – the Beijing AI Principles. However, an ethically-informed inquiry into autonomy is needed to adequately investigate the impact of algorithms on our moral freedom. Indeed, as we high pointed, we have seen in the current debate that in the first limit detected posed by algorithms to autonomy, the latter is mainly intended as self-control and active power to exercise of skills, therefore an account of human autonomy that does not evoke its moral connotation, instead, it is possible to bring out the moral dimension of human autonomy in the debate on the second limit posed by algorithms on users' capacity to appeal over algorithmic functioning, even this account of moral autonomy tends to be always traced back to the idea of autonomy as self-governance and individual self-determination. Yet, I have argued in the first chapter that an account of moral autonomy intended as full self-determination does not seem adequate to understand how persons can choose and act in a way according to which they can be defined as the authors of their choices as performed not in abstract but in the context of our contemporary informational societies which

[161] L. Floridi & M. Taddeo, "How AI Can Be a Force for Good", *Science*, 361 (6404) 2018. C. Burr, N. Cristianini, & J. Ladyman. An Analysis of the Interaction Between Intelligent Software Agents and Human Users. *Minds and Machines*, *28*(4), 2018, pp. 735–774. K. De Vries. "Identity, profiling algorithms and a world of ambient intelligence". *Ethics and Information Technology*, 12(1), 2010, pp. 71-85.

[162] B. Mittelstadt *et Alia*, "The Ethics of Algorithms: Mapping the Debate" in *Big Data & Society* 2016.

[163] A. Tsamados, N. Aggarwal *et alia*, *The Ethics of Algorithms: Key Problems and Solutions* 2020.

are globally interconnected and culturally diversified. Indeed, accounts of autonomy as agents' self-determination or self-governance risk to do not be able to give account of the way in which people act as moral agents in societies to the extent that they do not take into account the huge role played by the collective or social-relational dimension in informing and shaping moral autonomy and freedom, instead privileging an idea of isolated individual capable on its own to determine herself. Instead, I claimed that the social and relational dimension play a crucial role in informing the context in which we make our choices and actions, and therefore our conception of autonomy cannot prescind from considering it.

Specifically, I have argued how our moral freedom as freedom to choose and act as moral agents and form genuine moral identity requires that the agent is considered the author of their choices and actions. However, this autonomy as authorship is not defined as full independence or complete self-governance of the subject, inasmuch the condition of autonomy is always informed by the socio-relational dimension via the availability of options, in other words, our choices and action are always made in a social context of options. In this sense, the condition of autonomy is always seen as a *self-relational determination*, where the social and relational dimension inform the options we can choose as motives for our choices.

In this sense, moral autonomy consists in the possibility for the subjects to be the authors of their choices and actions by exercising the reflective endorsement on the morally heterogeneous options available, that in turn are deeply informed by the socio-relational dimension in the morally heterogeneous availability of options (values), hence, by endorsing them as moral motives for their choices and actions.

To sum up, insofar as the social and relational dimension informs deeply the individual dimension in which the subjects choose and act as moral agents, and that the reflective endorsement, as the condition in which the subjects express their self-determination is deeply informed by the social and relational dimension, the concept of moral autonomy elaborated as a necessary condition to secure at least at a minimum threshold the exercise of our moral freedom consists in the capacity of self-relational determination of the agents.

In the previous section I have underlined that the algorithmic role in reshaping or architecting our availability of options is not just an epistemological impact on the first condition of our moral freedom, the availability of morally heterogeneous options, but it can affect also the second necessary condition of our moral freedom, namely, our moral autonomy.

Our clarification of moral autonomy as *self-relational determination* makes immediately clear how algorithms by pre-determining which options are available to us affect also moral autonomy and above all its relational dimension, and hence, the way in which the social and relational dimension of our living can inform – via the options available to us – how we choose and act as moral agents in a social context. The revision of the concept of moral autonomy beyond a full determination of the agents to instead favor a self-relational determination provides us an adequate account to shed light on the impact of algorithmic technology on the way in which we choose and act in our informational contemporary societies: indeed, it allows us to understand how via technology our exposition to the socio-relational dimension is mediated and can be reshaped (also narrowed). I have indeed argued earlier how algorithms can architect our environment and options in a way that can reduce our exposition to social relations and above all heterogeneous experience of the other (in culture, opinions, thoughts, moral practices, values, and so forth) and how this lack in exposure has a key impact on our possibility to choose and act alternatively and therefore have a real or effective possibility to develop our moral identity in a genuine way (by forming and critically reasoning/testing our moral knowledge on the basis of which we steer the development of our identity). This account of moral autonomy considering the role of socio-relational dimension for our choice, action, and development of identity, seems also the soundest to give account of how today the socio-relational dimension informs inevitably our autonomy, and so our way to choose, act, and develop our identities when ML algorithms are based increasingly on filtering and classifying techniques such as collaborative filtering and therefore construct profiles of users – on which driving its categorization and then reshaping of their environment – based on the data gathered on other users' interactions.

If the account of self-relational autonomy is adequate to understand the impact of algorithms on our freedom of choice and action in our contemporary societies, and if – as we have highlighted – algorithms have a narrowing impact on the socio-relational dimension characterizing our autonomy (through the predetermination of our availability of options), in order to understand if this impact undermines moral autonomy, it is necessary to consider its core, hence, the way in which we determine an option rather than another alternative for our choices, by specifically approving it as a motive for our choice and action: the reflective endorsement.

Indeed, in the reflective endorsement we can exercise on the options available we find the distinctive trait of our authorship over our choices, actions, and identity. In the reflective endorsement we find the way in which we determine ourselves given a context of options, the last call for our moral freedom. Indeed, by exercising the reflective endorsement, and so by endorsing certain options embedding certain moral reasons and values, those we embrace, by approving them as motives for our choices and actions, we develop our *ought to*, namely, the way in which we make those moral reasons and values not just motives but the moral rules for our behavior, i.e., we make them normative for our conduct. This key-trait is indeed of crucial importance as by exercising our reflective endorsement we develop the way in which we respond to reality, conveyed by our mediated perceptions and emotions, by taking a moral stand (here emerges clearly the moral dimension of our agency), and therefore developing our moral posture: develop our moral identity as genuine persons. Indeed, by endorsing options as specific values and reasons as motives for our behavior we actualize our moral disposition towards a certain direction, rather than another, so exercising our freedom of become certain moral agents, rather than others, to choose and act with a certain moral posture, rather than another: in sum, to develop our moral identities in a genuine way as moral agents that are authors of their choices and actions.

As a consequence, in the reflective endorsement, as the expression of our moral autonomy, we find a key-condition for our moral freedom and identity. Indeed, even if who can pre-determine the ‘availability’ of our options can bind our choices (i.e., we cannot choose those options that result as unavailable) to certain options rather than others, and therefore can exercise a constrain on our freedom of choice and action, this constrain is soft to the extent to we as moral agents have always the power to decide to act against their informational options (as well as against our preferences and needs), or choose not to choose. This happens because our options do not necessarily determine us or constrain us to choose. Vice-versa we should admit that any strong biological, psychological, environmental or social influence would determine us necessarily (e.g., the fact that I live and grow up in a racist family or neighborhood can make the possibility that I will become racist more likely but does not necessarily determine my behavior and my identity as such; conversely, it would be difficult to admit the possibility of social and moral progress). The influence on our available options, as our environment of choice, is undeniable, but this influence does necessarily move or determine us to choose and act in a certain way rather than another. This is because we can always 'yes or no' to a certain option (reason, value, event, desire, and so forth) and this approval lies in the exercise of reflective endorsement.

Therefore, if we want to understand whether the impact of algorithms-based ICTs can undermine our moral autonomy, we have to specifically inquire into their potential action on the exercise of our reflective endorsement. To do so, we need to consider particularly algorithmic RS, inasmuch as their personalization action is deeper than that produced by filtering and classifying algorithms, insofar as it is not limited to filter and classify the options available to the subject, more intrusively, algorithmic RSs take one or more specific options (inferred as potentially affecting users' behavior) to micro-target the user in order to trigger the compliance-effect between users' behavior and RSs goals.

Let me clarify better.

RSs do not just personalize the choice-context of individuals by re-structuring, classifying, and filtering the informational options available to the subjects; more intrusively, they can capture what – amongst those options – the user values mostly, and by learning on them, try to predict highly personal characteristic (from mere basic interests to gender, race, and sexual, political, and religious orientation, up to values, beliefs, and deep vulnerabilities) and above all what are her choice-driving elements, i.e., what can trigger on the basis of the inferred personal characteristic a certain interest and therefore a user's response-behavior[164]

[164] To do so, RS can analyze both the associations between the user's past choices and the descriptions of new objects and also by analyzing what is valued by others and their interactions. RS determines users with similar taste via standard formulas for computing probabilistic correlations and then associate profiles constructed on the basis of common choice-driving elements discovered

– that is considered highly valuable for the maximization of click (that is usually a RS goal by default). Then, algorithmic RSs, on the basis of this personalized predictive knowledge, target the user with highly-personalized options that can be easily chosen by her (inasmuch as they respond or trigger her choice-driving elements as algorithmically inferred).

In this sense, the RSs' operation of micro-targeting users with options that are highly value-laden has the potential to generate a deeper impact on users' choice-behavior. For example, from an anthropological perspective, Seaver defines this RS action as inescapable or ubiquitous "sticky traps", insofar as they try to "glue" their users to some specific solutions, items, and products (or according our vocabulary: options).[165] A good example is the YouTube's RS algorithm, which received much attention recently for its tendency to promote biased content and fake news, in a bid to keep users engaged with its platform.[166] Instead, others as Burr *et al.*, Floridi and Taddeo, and de Vries privilege to call the algorithmic recommendation as "nudges" that precisely 'nudge' users to a particular direction, by trying to "addict" them to certain contexts[167]; or – as pointed out by Hilty – as a form of "technological paternalism" [168], on the basis of the fact that algorithms RSs seem to know better what is good for other people than these people themselves.

with 'social categories' in which the mass population is divided or grouped. These categories allow profiling algorithms on which RS is based to infer what are – according to macro-lenses – common choice-driving elements within a certain group, that are peculiar to one group rather than another, and with that predictive knowledge, re-target users with refined and more customized information. By this targeting action on users on the basis of their inferred profiles (based on what they value mostly correlated to what other people value mostly as well), RS discovers not just fine-grained tendencies inside categories, but can use them to further personalize the informational options displayed inside certain groups with the result of further corroborating the choice-driving elements of persons inside a category, in which the users as profiled are already inserted.

[165] N. Seaver (2018) "Captivating algorithms: Recommender systems as traps". *Journal of Material Culture*, 135918351882036 2018. Seaver calls this RS functioning as "captivation metrics", i.e., to measure users' retention.

[166] G. Chaslot (2018). "How Algorithms Can Learn to Discredit the Media". *Medium.*

[167] C. Burr, N. Cristianini, & J. Ladyman. An Analysis of the Interaction Between Intelligent Software Agents and Human Users. *Minds and Machines*, *28*(4), 2018, pp. 735–774. L. Floridi & M. Taddeo, "How AI Can Be a Force for Good", *Science*, 361 (6404) 2018. K. De Vries. "Identity, profiling algorithms and a world of ambient intelligence". *Ethics and Information Technology*, 12(1), 2010, pp. 71-85.

[168] L.M. Hilty, (2015). "Ethical issues in ubiquitous computing – three technology assessment studies revisited. In: Kinder-Kurlanda, Katharina; Ehrwein Nihan, Céline. Ubiquitous Computing in the Workplace. Cham: Springer, 45-60.

At the beginning of this section, I have already underlined the reasons why we should not think about algorithmic actions or interferences as a form of nudge or paternalism (to expand, see chapter 2.1). Indeed, algorithmic recommendations do not just filter or re-order options available, but use a few of them to specifically target individuals, therefore differently from nudge (as a choice-design method) as well as paternalism which include in their main definition the idea to do not exclude options, to privilege long-term individuals' goal and not undermine their autonomy and freedom of choice, they make some options as unavailable (really excluding them), work to privilege short-term goals (as the satisfaction of a curiosity, in the light to boost the click) and above all, as we will see in a while, they can undermine our autonomy and freedom of choice. Although personalization is a choice-design method which shares with nudge theory the idea to architect people's context in a way to make their behavior predictable, algorithmic recommendations can become a *real push*, not a gentle nudge. What I specifically claim is that RS can create a hard constraint on our moral freedom as freedom of choice and action as moral agents by affecting or suspending our reflective endorsement. Specifically, this phenomenon may happen when the information used to target individuals is highly sensitive (e.g., from information about individuals' physical or psychological status to personal vulnerabilities or weaknesses).

Let us describe these deep interferences as 'real pushes' with an example.

Profiling algorithms could infer from my queries on Google about, for example, "how people with cancer feel" and from my past geo-localizations, for example, "at the hospital", by correlating them with other sets of data presenting the same characteristics (see collective-filtering), that I am ill and so interested in a certain kind of information about cancer. In the light of this inference, RSs in order to maximize their goal (the CTR: click-through rate) could start to [hyper]target me (ubiquitously: in all my inter-connected devices, let us think about ads or videos that pop-up in the middle of a song you listen, or while you swipe up stories on Instagram and so on) with sponsored items or information labeled "relevant to me" as linked to my predicted pathology (where the link is probabilistically inferred by algorithms through the analysis of correlations inside a category of people grouped as presenting similar characteristics/ that have clicked similar contents).

These informational options labeled as "relevant to me"[169] are based on what others have clicked mostly so the ubiquitous hyper-target (in the form of advertising showing up in the webpages I consult, videos I watch, or SNS in use) is moved in order to meet the goal of driving me to click them, to incentive my choice in order to meet the algorithm's task, that is, as previously underlined, the maximization of click (with the user's loyalty) which in turn can result in buying a pair of shoes, or watching a certain movie, subscribing to a certain channel, but also buying a house, applying for a certain university rather than other, choosing a certain insurance, or preferring a certain medical treatment rather than another. This happens because the use of RSs today ranges from contexts such as health care, lifestyle, insurance, and the labor market that are morally loaded – i.e., RSs' outputs can produce consequences morally relevant for the individuals in contexts where a choice rather than another can be life-changing.

Due to the RS use in morally loaded contexts, the possibility for algorithms to capture highly personal and sensitive aspects of the individuals is very likely and this entails that is also very likely for RS capture the choice-driving elements (also via cross-inferences between groups and within the same group) which connotes users and that can be highly valuable to be exploited in order to push their behavior to a certain direction rather than another (a choice such as a purchase to a political vote, for example). In this sense, the options recommended (or better pushed) by RS are extremely value-laden, as able to trigger the choice-driving elements caught. Following the previous examples, the options recommended ubiquitously may vary from strong pictures (for example, a cancer patient under medical treatment) to very sensitive information (about the right of care's withdrawal). Since these options are pre-determined on the basis of the predictive knowledge about the user, they have a specific potential: that of triggering physical or psychological weaknesses, as well

[169] In order to set the relevance of an option, RSs make online experimentation or A/B testing, i.e. the practice of exposing selected groups of users to modifications of the algorithm, with the aim of gathering feedback on the effectiveness of each version from the user responses.

as vulnerabilities, pathologies such as depression or anxiety, or evoking fears and trauma. This means that RS have the power to raise emotional instinctive responses. Indeed, there is nothing stronger than touching a sensitive key to trigger emotional responses, and specifically, primary emotions[170]: fear, joy, anger, sadness, interest, and disgust. In psychology, these emotions are called primary as are those we have in common with very young child and animals (are instinctive and innate), that can be distinguished from those such as shame, envy, guilt, pride or regret (and of course many more) that instead are called as secondary inasmuch as they require a certain awareness, degree of socialization, and the formation of an idea of what is good (the formation of moral dispositions). This means that the options chosen by RSs to target user on the basis of the captured personal sensitive traits and choice-driving elements can work as *triggers*. Such options can trigger users' emotional and instinctive behavior-response (e.g., fear, extreme joy, anger and so forth) in a way that can suspend users' exercise of reflective endorsement, by leading that option emotionally loaded to determine their choices and actions.

In these cases, the option recommended can become a real push that affects in-depth individuals' autonomy, by suspending their reflective endorsement and transforming the main informational *option* pushed from being a *motive* of people's choices and actions (an option they can approve as a motive for their choices by reflectively endorsing it) to turn out as the main *cause* of users' choices and actions. In this sense, the RSs recommendation of such option, instead of epistemologically informing agents' choices and action (informing them), ends to decide or choose at the users' place, in other words, to determine them.

Specifically, since this algorithmic determination is based on a predictive knowledge developed to meet pre-determined or pre-set goals, we can define this potential action as a predetermining algorithmic action. Following the previous example, the algorithmic recommendation of such triggering options – such as strong pictures (for example, a cancer patient under medical treatment) or sensitive

[170] There is a wide agreement on this Darwinian matrix distinction in psychology, moral psychology and evolutionary theories. A key reference is P. Ekman, *Emotions Inside Out: 130 Years after Darwin's The Expression of the Emotions in Man and Animals,* New York: New York Academy of Sciences 2003.

information (about the right of care's withdrawal) – may determine the person's choice and behavior, for example, to renounce or stop a medical treatment, by deeply deploying a vulnerability or weaknesses to meet certain goals (at any user's cost). The choices as determined can be life-changing, as in the case mentioned, but also when they are not so morally loaded, they can open certain courses of actions while declining others, in a way in which firstly do not express the authorship of the agents and above all do not respect their moral reasons and values, whose approval has been suspended – this means that I do not take a moral stand in response to a certain option/event or information, insofar as I have been determined, though my choice has always a consequence on how I form my identity and therefore binds to a certain extent my future moral development.

Continuing to follow our example, if before a user valued the right to live and, for her moral orientation, she was against euthanasia or assisted suicide, the hyper-target of her with specific options (news, pictures, movies, stories, friends and so forth), embedding different moral values and reasons, options that are value laden in terms of capable to trigger choice-driving elements and so an instinctive response-behavior, can lead her not just to start to take into consideration an option previously excluded (that can be to stop a medical treatment, but also in the same context, to assume anti-depressive, decide to do not abort etc.), but as that option is shaped to trigger her vulnerabilities (problems, weaknesses…), the emotion raised can suspend her reflective endorsement, her capacity to approve an option and a reason as a motive for her choices and actions, instead triggering disruptive emotions that can lead that option to determine her choice.

In this specific case, the producible outcome can be accidental, but in many other cases the behavioral response can be pre-set and so intentionally designed to be met. As an example, we can think of political parties that want to reach consensus, exploiting personal events inferable by people in order to leverage them by micro-targeting them with specific information (news or policies etc.) to modify their political orientation, or we can think about pharmaceutical companies that target psychological patients to increase the purchase of their depressive drugs, and the list of cases can be easily expanded further.

Whilst, indeed, the output in question is morally neutral to RSs, since it responds to their designed goal, and they cannot qualitatively discern the content of a recommended option, the same cannot be said for the individual. Strong images about how much painful the cancer medical treatment can be (the same example can be for other crucial moral, religious, bioethical, or political issues) can trigger our more instinctive emotions (primary ones) and so, by suspending our reflective endorsement, determine our choice. This entails a deep undermining of our moral autonomy and the erosion over time of our moral freedom as the freedom to choose and act as moral agents/authors and so developing genuine moral identity.

To sum up: the relentless RSs' pushes in triggering options (high-sensitive contents) can affect users' autonomy in-depth by suspending their endorsement and transforming the main option pushed – algorithmically chosen for them – from being a motive (that users can endorse) to be the cause (strictly determining) of their choice and their behavior. So, our autonomy would become not just influenceable (via the hetero-determination of the relational dimension via the reshaping of our options), but also predeterminable (via the suspension of the endorsement).

Thus, this impact of algorithmic choice-architectures cannot just affect how persons develop their moral knowledge (epistemological level), but it might affect what we have defined as the formation of the ought to, our moral posture, that is, the corroboration of our moral dispositions towards our moral identity via the way in which we respond (or not respond) by taking a moral stand to our reality, i.e., by endorsing the value and moral reasons we embrace as motives of our choices and actions, which over time and in turn give form to our moral identity. In this sense, RSs cannot just create a "soft constraint" on our freedom of choice and action as moral agents, but ultimately, they might also raise a "strong constrain" on how we develop our moral identity, that would end in turn to not reflect our intrinsic values (as critically formed and endorsed), but third-party interests, goals, and preferences – by pushing us to choose and act not as genuine moral agents according to our own values, reasons, and projects, but according to predetermined criteria, goals, and interests algorithmically pre-set. In this sense, the algorithmic choice-architectures show the potential to turn their influence in a completely new and invisible form of impediment to our moral freedom.

It is important to notice that research into the ethical issues posed by RSs is still in its infancy. The debate is fragmented across different scientific communities, as it tends to focus on specific aspects and applications of these systems in a variety of contexts. The current fragmentation of the debate may be due to two main factors: the relative newness of the technology, which took off with the spread of internet-based services and the introduction of collaborative filtering techniques, and the proprietary and privacy issues involved in the design and deployment of this class of algorithms (that make the details of RS currently in operation treated as highly guarded industrial secrets). In this last section, I tried to further develop the relation between two of the topics less explored in the debate on ethics of AI and algorithms, with the consequent difficulty to understand and frame a problem that has missed of an inquiry so far. For this reason, this attempt should play as a forerunner to show that RSs pose several urgent ethical issues, here specifically addressed in the light of our moral autonomy and moral freedom, and therefore that much more ethical inquiry needs to be carried out on this specific field and topic.

To sum up my argument so far: in this chapter I have analyzed the impact of algorithmic choice-architectures – as ruled by algorithmic profiling, algorithmic classification and filtering, and RSs – on the *conditiones sine qua non* of our freedom of choice and action as moral agents and I have shown how this ethical inquiry has brought out a controversial predictive potential hold by algorithms; a predictive potential that can affect our moral freedom from different angles.

In the next chapter, I resume the main steps of my argumentation and define how this predictive capacity of ML algorithms at the root of the creation of choice-architectures is delineating a novel potential form of impediment to moral freedom, namely, what I define the *algorithmic predeterminism*.

THIRD CHAPTER

On the Protection of Moral Freedom

After having analyzed in the first chapter the value of our moral freedom and highlighted the conditions of possibility underlying its exercise, namely, the availability of morally heterogeneous options and moral autonomy as relational self-determination, in the previous chapter, I have provided an analysis of how algorithmic ICTs and specifically the algorithmic governance that is structuring in our informational societies in choice-architectures is reshaping our choice-contexts and that more than influencing is silently undermining the conditions of possibility underlying our exercise of our moral freedom, opening the risk of a new potential kind of predeterminism driven by algorithmic ICTs on our moral freedom.

The goal of this last chapter is that to further clarify this specific "new" threat posed by algorithmic ICTs to our moral freedom, as I have conceptualized it in the first chapter of the dissertation as both positive and negative moral freedom, by firstly enucleating the kind of impediment to our moral freedom algorithms are raising, and therefore shed light on what I define as "algorithmic predeterminism".

This definition helps me to elaborate a novel conceptual lens to frame and then introduce this unexplored algorithmic threat inside the current ethical debate on algorithms-based ICTs, and to finally define the specific social agents called into action when it comes to apply and translate the conceptual lens and tools proposed in this work into practices capable to secure at a minimum threshold the exercise of our moral freedom in our contemporary algorithmic societies.

Therefore, in the first section of this chapter, I further explore how the algorithmic choice-architectures – by affecting the *conditiones sine qua non* of our moral freedom – can endanger our moral freedom, both in a positive and negative sense, as freedom to become genuine moral agents and develop genuine moral identity (positive concept of moral freedom), and also as freedom from either actual and potential, moral and immoral interferences (negative concept of moral freedom). Hence, I develop and clarify this growing form of impediment to moral freedom, namely, the rise of algorithmic predeterminism, by analyzing its impact and ethical implications in an array of social domains where algorithmic ICTs are broadly applied.

In the second section of this chapter, I will draw insights from theories in the privacy debate concerning the protection of freedom, and specifically, the theories of informational privacy and intellectual privacy – which are developed in the framework of legal scholarship – in order to elaborate on them from an ethical perspective and more specifically to interpret them through a new conceptual lens, that I define *moral privacy*, and that – I claim – constitutes the third level of a three-layers privacy framework for a comprehensive protection of agents' freedom that specifically considers and includes the protection of our moral freedom in our informational societies. In particular, I underline a third-level zone of protection of our moral freedom and I elaborate specific axioms or prescriptions that should steer how to develop novel techniques to operationalize the value of moral freedom by design in algorithmic ICTs. Most fundamental is that this specific call for moral freedom encompasses a specific ideal underpinning these prescriptions, that is, our *freedom from algorithmic predeterminism*, namely, our freedom from being pre-determined in our choices and actions (and therefore over time in our identity) in a way that undermines our being genuine moral agents, by binding our choosing and agency to pre-determined criteria, patterns, and profiles algorithmically predicted via probabilistic assumptions.

Thirdly, in the last section of the chapter, I define who are the specific (institutional and technological) agents of social responsibility called to act in the application of the conceptual privacy tools developed so far both at a technical and a policy level and to secure their operationalization to protect our moral freedom in our contemporary algorithmic societies, the ones I define as the *champions of moral freedom*.

III.1 Predictability and Moral Freedom: Algorithmic Predeterminism?

Predeterminism is a very old concept and has assumed diverse declinations in the course of history (natural, biological, psychological, just to mention the main theories), although its underlying idea is that events (including human actions) have been already decided or already known, i.e., they are already pre-determined, and therefore the is no space for freedom of choice and action, as the chain of events is pre-established and human action has no power to interfere and intervene to that.

By algorithmic predeterminism, I specifically mean an unprecedented form of predeterminism generated by the predictive potential of ML algorithms-based ICTs, i.e., their power to discover predictive knowledge on individuals on the basis of their capacity to see what is invisible (and also impossible) to the human eye. This is allowed by their capacity to scale and compare huge amounts of data from which mining and discovering precious patterns and correlations on individuals' behavior in order to meet a specific and pre-determined task. The controversial use of this predicted knowledge on individuals (e.g., patterns, correlations, and profiles) can take the form of intentional or accidental, moral and immoral interferences on people's capacity to form their own ideas of good, values, moral reasons, goals, and their own ground projects to the point of subtle binding people's behavior to predicted elements, namely, in a way that can bind – via the techniques previously underlined – their conduct towards pre-determined goals and so undermines their freedom to choose, act, and thus become genuine moral agents. People's choices, actions, and conduct can end to align with an algorithmically pre-determined goal via the algorithmic reshaping of their choice-contexts on the basis of the algorithmic exploitation of the potentially predicting knowledge on them that is discovered or constructed in order to meet the pre-determined goal. In this sense, algorithms show a huge potential to predetermine individuals' choices and actions, by making very difficult for them the exercise of their power to intervene and choose and act as according to their own values, reasons, projects genuinely embraced and endorsed.

In order to argue the rise of this threat to our moral freedom and so to clarify the definition and the impact of the algorithmic predeterminism it is necessary to recall the steps of the whole argumentation developed in the work so far and then highlight the rise of algorithmic predeterminism in the algorithmic impact on the two conditiones sine qua non underlying the exercise of our moral freedom.

In the first section, I have defined the concept of moral freedom in both a positive and negative sense. Moral freedom is our freedom and power to become moral agents, and specifically genuine moral agents, that means our freedom to develop genuine moral identity. I have also argued that freedom to develop genuine moral identity means freedom to develop or embrace our own moral reasons, values, and ground-projects, that is, the freedom to choose and act in accordance with moral reasons, values, and ground-projects as options I can deeply endorse as the motives for my choices and actions. I have then argued that moral freedom also means freedom from potential and actual, moral and immoral interference to develop moral identity, whereby moral interference is meant an interference on us that even if can track and reflect our interests and moral orientation is nonetheless moral-freedom constraining, as it always interfere with my choice-behavior, and whereby immoral interference is meant an interference on us that is illegitimate, wrong, or unjust, as it does not respect us as persons by not considering our moral values, reasons, and goals (moral orientation).

Then, I have brought out what are the necessary conditions underlying our moral freedom, which specifically can secure at a minimum threshold its exercise, namely, the condition of availability of morally heterogeneous options, according to which an agent can develop an genuine moral identity if and only if the context of choice in which she chooses is characterized by morally heterogeneous options, i.e., options that embed plural moral values and reasons that are diversified, which in turn requires a sufficient moral exposure of the subject to morally diversified social relations, attachments, values (and so forth) – insofar as they can allow her to develop an alternative but (as critically tested and then endorsed via choices and actions over time) genuine moral identity. The second condition is that of moral autonomy according to which an agent can develop a genuine moral identity if and only if she can be the author of her moral identity by reflectively endorsing amongst the morally heterogeneous options available those values and reasons she embraces as motives for her choices and actions. Since the latter condition strictly requires the former, we cannot think moral autonomy as a full self-determination, and so, without recognizing the role played by the social-dimension (as implicated by the first condition) in the exercise of our freedom of choice and action as moral agents: therefore, by moral autonomy I mean a relational self-determination.

I have then concluded the first chapter by arguing for the ethical-normative value of moral freedom for the flourishing of the moral dimension of both the individual (for the development of our own idea of good, our moral values, and the dimension of ought to that allows us to oblige ourselves to them) and of our living together (as it permits moral openness and the development of a culture of reasons-giving that are essential for the social dialogue, joint cooperation, and the sharing of social commitments), and therefore I have underlined its necessary protection as an ethical normative value from existing and novel forms of impediment.

In the second chapter, I questioned whether the algorithmic governance that is structuring into our reality, in the light of its disruptive impact on our reality, may hamper our moral freedom. Specifically, I have shown how algorithms ruling our ICTs are becoming architects of our choice-contexts, insofar as they re-shape and structure the set of available options displayed to the users. I have argued how due to the digitalization of our societies, everything is datified, and therefore captured by data as information, and so information, and specifically informational contents, embed in turn values, beliefs, thoughts, namely, everything that can be considered as a reason and that can be endorsed as a motive for our choices and actions. This means that informational contents can be considered as options we can choose, and to the extent algorithms as gatekeepers manage information – by determining what can be available to us and rule out what is instead deemed as irrelevant on the basis of the predicted or constructed profiles on us – they are structuring our options and so reshaping our choice-contexts, by generating what I have defined as algorithmic choice-architectures. I have then argued how algorithmic choice-architectures can affect the two *conditiones sine qua non* underlying the exercise of moral freedom.

Firstly, I have underlined what I have defined the epistemological problem posed by the algorithmic re-shaping of our choice-contexts, and specifically, the epistemological impact raised by algorithms on individuals by pre-determining the number and the kind of options available to the subjects on the basis of the inferred, discovered or constructed profiles on them: I called this impact as the *algorithmic hetero-definition of the availability of options*, by underlining that the definition of our availability of options is hetero-connoted, i.e., depends on external forces and probabilistic assumptions on which we do not have the power to intervene, and that this definition is by default, as algorithms by selecting the relevant information for us pre-determine the conditions of our choices and restrict the range of available options. In other words: algorithms are the external forces that determine what option (which in turn embeds values, reasons, and therefore potential alternative courses of action) from being potentially considerable into my choice-availability of alternative options is effectively actualized as a part of my choice-menu, i.e., as a part of my context of available options. This is the first hetero-determination of algorithms on our freedom to choose and act as moral agents as it consists of pre-determining the available options to the subjects to choose on the basis of the profile algorithms can discover and construct of them. Here I have developed two scenarios to show how this epistemological impact does not seem avoidable in the way in which algorithms are currently designed and I have shown that algorithms can affect individuals epistemologically in different ways (or better, the epistemological problem can be understood and declined in two different ways).

The first scenario is an algorithmic choice-architecture where the profiles of the individuals as inferred from probabilistic assumptions (on which classification and filtering algorithms will base the determination of the options that are available to them) reflect effectively subjects' values, reasons, beliefs, goals (and so forth), broadly speaking, inferred or constructed profiles reflect effectively subjects' moral orientation and therefore tailor continuously the choice-contexts of the users on the basis of this constructed knowledge.[171] In this case, profiling algorithms steer classifying and filtering algorithms to align the hetero-determination of the kind of options available to the subjects towards users' well-captured moral orientation (i.e., values, reasons, beliefs, goals). In this sense, the subjects may result as fostered in their choices and actions by this algorithmic action, insofar as the algorithmic hetero-definition of options would result as aligned to their values, beliefs, intentions, in other words, to their moral orientation (or if it is not formed yet to their moral dispositions). Following this reasoning, this hetero-determination would not pose a real limiting or hindering impact on the *first conditio sine qua non*

[171] This is made possible thanks to the large quantity of data available on us that can allow the construction of highly accurate or personalized profiles on us as persons (let us think about RSs and their capacity to capture the driving elements of our choices) and therefore reshape our availability of options on the basis of this predictive knowledge on our moral orientation.

of our moral freedom, i.e., the availability of alternative options, as the options preset would reflect the probabilistically well-inferred users' moral orientation.[172] Although apparently this hetero-determination may increase the capacity of the subject to choose and act according to her values, goals, and so forth, I have argued that actually this algorithmic hetero-determination affects the first *conditio sine qua non* underlying the exercise of moral freedom at its core, specifically, it undermines qualitatively our availability of alternative options, i.e., the alternativity factor, and specifically, the moral heterogeneity of the available options that is peculiar to this condition. I have shown how this alleged choice-enabling algorithmic functioning has been uncovered to undermine our exposure to different points of view, beliefs, values, and relations and to creating filter bubbles or informational echo-chambers, i.e., environment characterized by like-minded people, hence with similar beliefs, orientations, and values, therefore, characterized by morally homogeneous options (moral bubbles and echo-chambers). I have shown that algorithms end to narrow our exposure to diverse social interaction, information about people with other culture, values, and ways to do things, that is instead crucial to develop genuine moral values, reasons, and identity, opening the possibility of wondering whether the moral rules, values, and practices we are following are optimal, so to test and eventually change them. In this way, algorithms, by shaping our availability of options (and therefore who to get in touch with and what piece of information see) on the profiles probabilistically discovered as similar or just like ourselves, lead to encounter of those of exactly the same opinion or value sets as our own and this tends to make us more enclosed and radicalize our previous orientation, instead of critically challenge it – critical test that is necessary to become aware of our moral orientation. With this first argument, I have maintained how the algorithmic choice-architecture start to pose a risk to our moral freedom as freedom of choice and action as moral agents and specifically to develop moral genuine identity, by undermining the necessary but not sufficient condition underlying its exercise, the availability of morally heterogeneous options, that turns out to be pre-determinable in a way that diminishes the social exposure of the agents to diversified values and

[172] In other terms, the options presented would be the same the users would have chosen, if they would have time and resources to perform a similar filtering operation.

moral reasons. This lack of heterogeneity of relations, points of views, and orientations is indeed a lack of heterogeneous reasons and diversified ideas on what is good, and so undermine the possibility to challenge and reasoning on our moral orientation, and therefore the possibility of developing genuine reasons, values, and identity. I have defined this impact of algorithmic choice-architecture on our availability of morally heterogeneous options as an epistemological impact on individuals' freedom to choose and act as moral agents because it affects the way in which the subject develops her own idea of good, her moral values, moral reasons, namely, the formation and critical reasoning on her *moral knowledge*, i.e., of those moral reasons and values she can endorse as motives for their choices and her actions and on which she steers the development of her moral identity as a genuine moral identity. Algorithms-based ICTs impact on our choices, by pre-determining the options available, where the epistemological influence is an interference on the way in which we form and critically test our moral knowledge, namely, we form those moral reasons and values we can endorse as motives of our choices and action – and on the basis of this endorsement over time we steer and shape our moral dispositions and define our moral identity (or moral posture). So, in the first scenario, I have argued that even when algorithms are able to capture our "moral orientation", they end to affect epistemologically the individual, specifically affect her moral knowledge, by narrowing individuals' exposure to diversified social-relations, points of views, and therefore moral reasons and values, as a result of personalization based on profiling.

Here one may think that considering the fact that very often ML algorithms work on de-individualized assumptions and correlations this may constitute a counter-effect to personalization and so safeguard the availability of morally heterogeneous options. Nevertheless, I have argued how ignoring the individuality of a person via de-personalization is not the solution to the above problem. Rather, I have shown that when the predictive models which drive profiling algorithms are imperfect and specifically ignore the user beyond the profile as a person, the algorithmic reshaping impact on users' options can become even more problematic, above all when as – previously argued – the options pre-determined as available to people are not just informational *stricto sensu*, but are also considerable alternative possibilities such as real chances and opportunities. We have indeed widely described how everything is interconnected via algorithms and the same profiling algorithms that rule our SNS can inform predictive recidivism algorithms, or algorithms that determine who can be denied a credit, loan, and housing, who can access to a job, to a subsidized rate of health insurance or a particular health service. This means that, for example, a person can be subject to an adverse decision, due to de-individualized assumptions, such as being denied credit, simply in virtue of being similarly profiled – in a way that do not consider her as particular person – to persons who are not credit-worthy; as well as, a person can find out that she is paying very-high rates of her insurance on her life as her profile shows symptoms of depression and from the analysis of her clicks and interactions she has been categorized in a group more likely to commit extreme gestures (such as suicide or other life-threatening actions), or again a person can be on the hot-spot of police patrols just inasmuch as her name is very common between criminals and so be easily subject to be arrested, or see to be denied an housing application just on the basis of that name. In all these examples, the interferences beyond being freedom-constraining are deeply morally unjust and illegitimate, as based on inexact, wrong and often biased assumptions which underlying the construction of algorithmic profiles and then are used to reshape the options (online and offline) available to the individuals. I have argued then how algorithms which make decisions on subjects on the basis of profiles that do not consider them or reflect them as particular persons can deeply affect individuals epistemologically, by creating asymmetries both in knowledge and in power. Indeed, when the algorithmic probabilistic assumptions on the basis of which we are profiled and categorized into a certain group are inexact and de-individualize us as persons, not only we experience a limitation of available options by algorithms, inasmuch as they are hetero-determined and chosen in advance by algorithms on the basis of the profile of us probabilistically discovered, but also morally unjust and illegitimate, as based on an algorithmic knowledge of me that does not reflect me as an agent and person by triggering severe consequences on a wide array of domains. This results in an epistemological asymmetry between who I am as a moral agent, and therefore, my values, my attachments, moral ground projects, beliefs, and so forth, and how I have been profiled, known, and described by algorithms, and on the basis of this profile, re-influenced by the shaping of my choice-environment. To sum up, the algorithmic knowledge on me – as profiled – on which classifying and filtering algorithms base the pre-determination of my available options does not reflect me as person and rater it shapes my informational environment according to that profile, instead of myself. Moreover, when we are categorized and grouped in arbitrary or increasingly complex ways that we are often unable to predict, understand, or contest the algorithmic decisions (that can an information shown to me or "my label as not creditworthy") we are subject to. In this sense, the epistemological asymmetry is a real asymmetry in terms of power between the algorithmic profiler (and who is behind) and the person profiled. In this sense, I have defined also this impact as an epistemological impact on the individual, insofar as it weakens the epistemological position of the "decisional" subjects, who is made unaware and passive towards their choice-context: she cannot know as well as intervene on the options algorithmically pre-determined – options which in turn lead agents to a context of pre-determined alternative possibilities, opportunities, chances, and courses of actions. This algorithmic impact is a further epistemological impact on our moral freedom, because not just the hetero-determination of the options available are pre-determined and narrowed, but this operation is conducted according to patterns or associations of which we are unaware and on which we have no power, neither cognitive power nor power of action, so creating an epistemological asymmetry stricto sensu, i.e., between how we are as moral agents and how we are known as algorithmically profiled, and an epistemological asymmetry as a power asymmetry, as we are unaware of the associations and assumptions driving the discovery of our profiles and so we are made unable to act on them.

The potential of algorithms to architect, and therefore influence or interfere – via personalization techniques based on profiling – our availability of alternative (morally heterogeneous) options, as such, it can be always considered as *freedom-conditioning* because they always reshape and condition the way in which we form our knowledge (our thoughts and beliefs), and specifically, our moral knowledge (our own idea of good, values, moral reasons, and projects), by architecting and so determining our choice-contexts. However, in this influencing action, algorithms may not distinguish so much from many other interferences we can experience in our lives (as those exercised by the environment in which we grew up, our family, our biological inheritance and so forth) – perhaps, except for the degree of influence that in the case of algorithms which are interconnecting and enveloping our world is extremely high.

The crucial trait is that their influencing power is based on a huge *predictive capacity* developed and applicable on us and the environment around us (think for example about collective-based filtering based on profiling). This predictive power developed by ML algorithms in the combination of profiling, filtering/classifying, and RS works by testing the efficiency (the value) of the patterns and correlations[173] probabilistically discovered in predicting individuals' behavior by using them such as hinges on the basis of which reshaping and structuring users' choice contexts, firstly by attaching labels or profiles to individuals and secondly zipping them into constructed categories to which showing similar options. The threat of algorithmic predeterminism lies in this specific *modus operandi*: in order to meet a goal that is algorithmically pre-set by design in order to satisfy third-party interests, algorithms can construct profiles of individuals basing on the achievement of that goal, that means considering features that can be relevant with regard to that goal, but that very often do not reflect us as specific persons or predict our real future behavior. Our alignment in behavior to the predictive knowledge inferred can result over time as the consequence of the reshaping action carried out by algorithms on our choice's options that can be so limiting to do not allow an alternative course of conduct. In other words, algorithms can end to make valuable also patterns and correlations that are inexact or do not really predict our future behavior, but make us to align to them by narrowing quantitatively and qualitatively the options that are available to us as constructed profiles and therefore as part of a certain group (a very simple example: I do not like criminal series, but as my channels transmit just them and my friends talk just about them and the news tells how they vehiculate a certain political

[173] The patterns are discovered via probabilistic techniques and therefore are not based on a specific rational or causal links: this means that very often can be inexact or not really relevant to understand human behavior.

message to which I am interested, I will be likely driven to conform my behavior to the kind of options available and therefore easily start to watch criminal series).

The threat of algorithmic pre-determinism is becoming clearer: what I should like, watch, think, and – above all – what I should value can be either already known (as in the case of the first scenario where the constructed profile reflects my moral orientation and therefore the algorithmic impact can enclose my choice behavior into a moral echo-chamber), or can be algorithmically decided (as in the second scenario in which my profile does not reflect my moral orientation) as valuable not on the basis of what I genuinely value but on what has been discovered via data as valuable to meet a certain pre-set goal. This means that even if the algorithmic prediction is not exact, algorithms – by restructuring our choice-context and determining the availability of options on the basis of this prediction and profile – can pre-determine us to value certain things rather than others, to choose and act in a certain way rather than in another way, inasmuch as the condition of choosing and acting otherwise underlying the exercise of our moral freedom as freedom of choice and action as moral agents is deeply undermined.

This first impact of algorithms on the first *conditio sine qua non* of our moral freedom has shown that as our availability of morally heterogeneous options can be undermined in different ways, the risk of an algorithmic pre-determinism according to which our choices are pre-determined algorithmically is not a sci-fi scenario. The analysis of the impact of algorithms on the second necessary condition on our moral freedom, i.e., moral autonomy, allows to claim that the algorithmic predeterminism driven by the huge predictive capacity of algorithms more than a risk is becoming a reality.

In the last section of the previous chapter, I have shown how the impact of algorithmic choice-architecture can go beyond the epistemological level of persons (as profiled) and affects also their moral autonomy, as relational self-determination. I argued how the relational dimension that is part of our possibility to be the author of genuine moral identity is undermined by the pre-determination of algorithms of the availability of options that very often reduce users' socio-relational exposure and thus the moral heterogeneity of values and reasons that are crucial for the agents to test the values embraced, therefore deeply limiting their autonomy and power to choose and act as genuine moral agent to a very narrowed – in quality and quantity – and so constrained availability of options algorithmically determined, or better, predetermined on the basis of patterns discovered and profiles constructed. Therefore, the predictive knowledge developed on the agent can become a way to bind her potential choice and action to a set of options algorithmically pre-determined. Furthermore, I have also shown how this algorithmic impact can also create a further constraint on persons, by affecting at the core of their autonomy their ability to reflectively endorse those options they embrace as motives of their choices and actions. Reflective endorsement is our last call for moral freedom: in the reflective endorsement we exercise on the options available we can find the distinctive trait of our authorship over our choices, actions, and identity. In the reflective endorsement we find the way in which we can determine ourselves given a context of pre-determined options. Indeed, by exercising reflective endorsement, and so by endorsing certain options embedding certain moral reasons and values, those we embrace, by approving them as motives for our choices and actions, we develop our *ought to*, namely, the way in which we make those moral reasons and values not just motives but the moral rules for our behavior, i.e., we make them normative for our conduct. This key-trait is indeed of crucial importance as by exercising our reflective endorsement we develop the way in which we respond to reality, conveyed by our mediated perceptions and emotions, by taking a moral stand (here emerges clearly the moral dimension of our agency), and therefore developing our moral posture: develop our moral identity as genuine persons. Indeed, by endorsing options as specific values and reasons as motives for our behavior we actualize our moral disposition towards a certain direction, rather than another, so we exercise our freedom to become certain moral agents, rather than others, to choose and act with a certain moral posture, rather than another: in sum, to develop our moral identities in a genuine way as moral agents that are authors of their choices and actions. I have argued that even if who can pre-determine the 'availability' of our options can bind our choices (i.e., we cannot choose those options that result as unavailable) to certain options rather than others, and therefore can exercise a constrain on our freedom of choice and action, this constrain is soft to the extent we as moral agents have always the power to decide to act against their informational options (as well as against our preferences and needs), or choose not to choose. Therefore, I had to explore the algorithmic impact on our endorsement to understand whether the risk of algorithmic predeterminism is real. I have shown that algorithmic recommendations are not nudges, but real pushes, that can create a hard constraint on our freedom of choice and action as moral agents. RSs create indeed a more pervasive action on individuals: they do not just filter or re-order options available, but can use a few of them to specifically target and predetermine individuals' choice-behavior.

To understand through and through the risk of algorithmic predeterminism, I highlighted when the phenomenon I called 'endorsement suspension' may happen, and this is specifically when the information used to target individuals is highly sensitive (e.g., from information about individuals' physical or psychological status to personal vulnerabilities or weaknesses). Due to the RSs' use in morally loaded contexts, the possibility for algorithms to capture highly personal and sensitive aspects of the individuals is very likely and this entails that is also very likely for RSs capture the choice-driving elements (also via cross-inferences between groups and within the same group) which connotes users and that can be highly valuable to be exploited in order to push their behavior to a certain direction rather than another (a choice such as a purchase to a political vote, for example). In this sense, the options recommended (or better pushed) by RS are extremely value-laden, as able to trigger the choice-driving elements caught. In turn, since RSs' use today ranges from contexts, such as health care, lifestyle, insurance, and the labor market, that are morally loaded, RSs' outputs – what can trigger a certain recommendation – produce consequences morally relevant for the individuals, where a choice rather than another can be life-changing. The options algorithmically recommended are pre-determined on the basis of the predictive knowledge about the user and so they have a specific triggering potential: they can trigger physical or psychological weaknesses, as well as vulnerabilities, pathologies such as depression or anxiety, or evoking fears and trauma. This means that RSs have the power to raise emotional instinctive responses, and specifically, trigger primary emotions that are instinctive and innate. This means that the options chosen by RS to target user on the basis of the captured personal sensitive traits and choice-driving elements can trigger users' emotional and instinctive behavior-response (e.g., fear, extreme joy, anger and so forth) in a way that can suspend users' exercise of reflective endorsement, by leading that option emotionally loaded to determine their choices and actions. I provided examples on the way in which the predictive knowledge of ML algorithms can lead to discover users' vulnerabilities, traumatic experiences, weaknesses (and so forth) by associating data capturable from diverse applications (such as GPS, SNS, healthy-apps) on the users' history (and broadly on other users' interaction via collective-based filtering) and, as algorithms are not capable to distinguish qualitatively the options (the moral weight of a certain option embedded in an informational content), RSs can target those vulnerabilities (also called choice-driving elements) with information that can be sensitive or emotionally loaded so much to create not just a soft constraint on individuals' freedom of choice and action, as that created by pre-determining the options amongst which users can choose, but a hard constraint on it, by suspending the key-endorsement subjects can give to a certain option rather than to another, by approving it as a motive for her choice and action. In these cases, the option recommended can become a real push that affects in-depth individuals' autonomy, by suspending reflective endorsement and so transforming the main informational *option* pushed from being a *motive* of people's choices and actions (an option they can approve as a motive for their choices by reflectively endorsing it) to turn out as the main *cause* of users' choices and actions. As a consequence, the RSs' recommendation of such options, instead of epistemologically informing agents' choices and action (informing them), ends to decide or choose at the users' place, in other words, to determine them.

In this sense, algorithms turn out to be not just the architects of the contexts of our choices, by informing our choices via the reshaping of our availability of options, but they become the architects of our choices and actions themselves, by not just informing our choices, rather pre-determining them.

The choices as algorithmically determined can be life-changing, as in the case mentioned, but also when they are not so morally loaded, they can open certain courses of actions while declining others in a way in which firstly do not express the authorship of the agents and above all do not respect their moral reasons and values, whose approval has been suspended – this means that I do not take a moral stand in response to a certain option/event or information, insofar as I have been determined, though my choice has always a consequence on how I form my identity and therefore is binding my future moral development.

In this sense, our moral autonomy would become not just influenceable (via the predetermination of the relational dimension), but also predeterminable (via the suspension of the endorsement). Therefore, also our moral autonomy, as the second necessary condition underlying our moral freedom, might result as affectable by algorithmic choice-architecture. It follows that algorithmic choice-architecture cannot just affect (and interfere) on how persons develop their moral knowledge (epistemological level), but it might affect (and interfere) on what we have defined as the formation of the ought to, our moral posture, that is, the corroboration of our moral dispositions towards our moral identity via the way in which we respond (or not respond) by taking a moral stand to our reality, i.e., by endorsing the values and moral reasons we embrace as motives of our choices and actions, which over time and in turn give form to our moral identity.

If this suspension of our endorsement happens regularly, the constrain raised by algorithms on our freedom of choice and action as moral agents can ultimately become a strong constraint on how we develop our moral identity (*hard constraint*). Indeed, I would not be able anymore to express in my choices and actions the values I really embrace as a result of an heterogeneous exposure and its relative critical reasoning and assessment, but I would end to recursively emotionally-respond to the options presented (e.g., an update, an ad, a news, a friend's post, and so on) in a way that would not reflect my intrinsic values (as critically formed and endorsed), but that is aligned to third-party interests, goals, and preferences, namely, towards algorithmically predetermined options to meet pre-set and heteronomous goals.

In this sense, it sounds that we are in front of a novel threat to moral freedom: an unprecedented form of predeterminism generated by ML algorithms' potential to discover or construct predictive knowledge on individuals whose controversial application can generate intentional or accidental, moral and unmoral, interferences or constraints on people's capacity to form their own ideas of good, their own moral ground projects, as well as form and assess own values and moral reasons, to the point of undermining their freedom to choose, act, and so become genuine moral agents. This is exactly what I call the rise of algorithmic predeterminism.

Let me give some examples about the extent of the impact of the algorithmic predeterminism.

Think about a user who suffers from anorexia or a person that has been profiled as highly interested in losing weight on the basis of her clicks on websites and interactions with persons who speak about the topic of how to lose weight. On the basis of this profile, algorithms can structure a choice-context of informational options responding to this profile of her, so she will be continuously recommended with options on SNSs or offline (e.g., think about the smart fridge) of products or information in order to lose weight (diet food options, pictures of skinny models, and so on). This is very common indeed as there are many companies that pay online advertising on Facebook, YouTube, Instagram to select specific groups of people to which targeting and recommending their products exactly on the basis of what they watch, listen, visualize, and buy. As SNSs' algorithms work by rewarding those paying companies (as they are clients) by setting the goal of their algorithms to incentivize the click on these companies' products, algorithms will be set to mine data in order to discover certain characteristics, patterns, correlations and construct profiles that can be useful to categorize users in specific categories that in turn can be exploited to decide to whom showing or recommending certain information, so to boost the click or the purchase of a certain product. Basing on this functioning, algorithms will target those categories so grouped as interested, obsessed, or even vulnerable (e.g., as they suffer from anorexia) to the issue of losing weight, as they will be the most likely to buy those products. According to this *modus operandi*, the profiles of users will be constructed on the basis of the pre-set goal – that means that algorithms will deem valuable in the construction of a profile what can be inferred of the individual as related to the issue considered –, as well as what will be displayed to a certain user as part of a certain group – labeled of interest – will be predetermined on the basis of the profile constructed on the basis of the pre-set goal.

This type of algorithmic pre-determining action can lead to very dangerous consequences. First of all, if the algorithmic prediction is correct and so it reflects correctly a person’s disease (e.g., anorexia), the algorithmic interference can exacerbate the aforementioned pathology and the person’s psychological suffering. But, let us consider specifically the perspective of a person’s values, goals, and interests. For example, we can consider the specific case of people that suffer from anorexia but they are deeply aware and so believe that anorexia is a severe disease and for this reason they are pursuing a path to recover from it consisting in a series of daily small choices towards the healing (e.g., a person’s small choice can consist to stop to look at skinny models and work on accepting her own body, or stop to count food calories, and so forth). We can also think about people whose constructed profiles say that they are particularly interested (or even obsessed in losing weight), although they denounce any behavior that can undermine their long-term well-being. These long-term goals and genuine values, reasons, and interests will not fall under the algorithmic consideration, as they do not constitute features relate to the achievement of the pre-set goal (i.e., boosting the click on for example diet products). At the same time, their research online about the topic of interest can be captured as a valuable element for their profiling and categorization into a certain group – deemed of interest – to which micro-targeting and recommending options related to the issue of losing weight (e.g., because of the high number of correlations between clicks on anorexia’s websites, pictures of skinny models, and diet recipes). On the basis of the algorithmic functioning, for example, algorithms-based SNSs can start to recommend or show pervasively informational contents that recommend how to lose weight (pictures, diets, and so on) by gradually and daily change beliefs, reasons, and values of people that are daily target with pictures or information about the perfection of a skinny body so changing their behavior by influencing their daily small choices (from skipping a meal to believing that what you see online is real or that the picture of that skinny model stands for a beautiful and healthy body). This can happen also because it is very easy to go from a recommendation of a website with products to lose weight or with advice of people to buy them to pro-anorexia websites that explains to the youngest how to contract gradually the pathology up to communities providing to justification, support, and reinforcement to continue in the behavioral disorder. In this way, algorithms “inadvertently” do not just exploit a vulnerability or weakness inferred directly or by proxy to push that person toward their predetermined goals so to satisfy third-party interest and increasing clicks, but they can also trigger a chain of serious and high-risk choices that can endanger the life of some individuals. This example is also valid for a person as mentioned above that does not suffer from anorexia but is just interested in losing weight, while morally disapproving of anorexia or any other disease, addiction, or behavior that can end to undermine deeply her health status. The algorithmic recommendation of support groups, friends, products on the basis of the profile constructed on her that can leverage on the desire of losing weight can trigger an emotional response of 'interest' (i.e., primary emotion) that can lead the person to enter in specific bubbles and change her values and long-term goals just to pass a short-term costume fitting. Given the application of algorithms-based ICTs in morally loaded contexts, the implications of algorithmic predeterminism can be of this extent.

If the threat and the extent of algorithmic predeterminism are now clearer, it is also important to clarify what at the root of algorithmic predeterminism is really detrimental for our moral freedom as freedom to develop genuine moral identity. Indeed, whilst the profiling action of algorithms is very difficult to regulate in our informational societies where algorithms are more and more applied in morally-loaded domains and therefore our sensitive information is increasingly exposed to algorithmic profiling, what instead can be regulated – as we will try to frame in the next section – is the algorithmic use of this information to further refine profiles and patterns to achieve certain pre-set goals (so using individuals to test the inferred predictive knowledge) and therefore to interference to our choice-behavior. The threatening action on our moral freedom carried out by algorithms-based ICTs is precisely expressed in the moral and immoral interferences they can exercise on us – the negative concept of moral freedom elaborated in the first chapter of the work can be of help to clarify this point. The negative concept of moral freedom highlights indeed that our moral freedom can be defined also as freedom from potential and actual, moral and immoral, interferences to develop genuine moral identity, where by moral limitation is meant an interference on my choice-behavior that even if is based on my well-deducted interests, goals, and projects (and so on) is nonetheless moral-freedom constraining as it determines and narrows my availability of options and so interferes with the possibility to critically test, eventually change, and so genuinely embrace my own reasons and values for my choices and actions, and whereby immoral limitation is meant an interference on us that is illegitimate or unjust as does not consider us as specific persons, our goals, projects, values, and reasons (moral orientation), and this can result in a de-individualization of us that can lead to label, profile, and categorize us according to inexact, irrelevant, wrong, or biased assumptions. We have previously seen this distinction in the algorithmic impact *on the first conditio sine qua non* (second section of the second chapter). According this definition we can distinguish the algorithmic interference as delineated in the first scenario, where algorithmic reshaping function of our options is based on profiles of us that can track our interests, goals, and moral orientation, as always freedom constraining (as argued above) but definable as moral, inasmuch as take into consideration us as persons. 'Moral' does not mean that is good per se, but that it takes into account us as moral agents (i.e., our moral dimension), and that is not morally unjust or illegitimate, as not based on wrong assumptions and so do not subject individuals to illegitimate outcomes. Nevertheless, this interference always remains a moral freedom-constraining, and so needs a moral justification and can be regulated (and fixed) on the basis of what justification can be provided for it. This may sound controversial, nevertheless, in contemporary societies where profiling algorithms rule more and more our lives, our consideration as specific persons is better than a profiling of us based on wrong and inexact assumptions. Indeed, the kind of profiling that de-individualize persons does not just produce an interference to our moral freedom (that in the case of the first scenario, as the algorithmic interference is aligned to agents' interests, values, and – broadly speaking – moral orientation, can be defined as moral interference), but produces also an immoral interference on our moral freedom, where by 'immoral' I mean that it is illegitimate, as based on inexact, wrong, or even biased assumptions. Indeed, being subject to an algorithmic decision or a reshaping action of my effective options, chances, possibilities on the basis of wrong or biased assumptions is an interference that can produce deep and severe phenomena of injustice and as such they need to be detected and eliminated.

As we underlined at the beginning of the first chapter, our moral freedom is multidimensional and dynamic. There is not a black and white scenario, i.e., with full moral freedom or no-moral freedom. In other words, this means that we can enjoy moral freedom more on a positive or on a negative sense depending on cases and situations, as well as the impact and the consequences raised by algorithmic predeterminism can be of different degree if it expresses itself via moral or immoral interferences (or both). The distinction between moral and immoral interferences does not change the fact that, because of their current *modus operandi*, algorithms – and specifically the phenomenon of algorithmic predeterminism they can raise – can constrain our moral freedom to choose and act as moral agents, by binding us to pre-determined options. Nevertheless, this distinction can be helpful when we are called to distinguish what interferences need to be just regulated or fixes and those that instead must be eliminated as illegitimate or unjust. This distinction is very useful when we need to understand how to prevent or mitigate the effects of the algorithmic predeterminism that expresses itself in both moral and immoral interferences to our moral freedom to become genuine moral agents.

In this first section, I have argued the risk of a novel threat to our moral freedom, what I defined as the rise of algorithmic predeterminism. In the following section, I shed light on how the debate on algorithmic ICTs and AI focuses on the protection of freedom especially through the privacy theories. Since everything is considerable as information with the digitalization of our world, the specific theory that in the field of privacy looks at the protection of freedom via informational terms is that so-called the theory of informational privacy. Therefore, I discuss two of the most relevant paradigms of informational privacy provided so far, respectively, the theory of informational privacy *stricto sensu* and the theory of intellectual privacy, inasmuch as they provide meaningful insights to understand from a regulative point of view how to look at protection of our overall freedom. Even though these theories do not consider moral freedom and its actual and potential threats, so leaving the problem solving of this issue uncovered, they lay the foundations to build a specific privacy lens to introduce, frame, and address in the ethical and legal debate the risk for our moral freedom of algorithmic predeterminism – as delineated above.

Therefore, in the next section, I will argue how these two theories can be further developed in the light of the specific protection of our moral freedom, by sketching how this fundamental value – whose normativity has been argued in the first chapter of the dissertation – ought to be safeguarded through the elaboration of a novel specific privacy lens, what I will define as *moral privacy*.

III.2 From Informational Privacy to Moral Privacy

The main goal of this section is that of understanding how to secure our moral freedom as our freedom of choice and action as genuine moral agents.

In the previous section, I have resumed the steps of the argumentation developed in the second chapter in order to show how effectively the establishment of algorithmic choice architectures is posing a serious risk to our capacity to choose and act as genuine moral agents and therefore to develop genuine moral identity. I have called this risk as algorithmic pre-determinism by meaning the capacity to algorithms to predetermine our choices and actions and therefore to constrain our capacity to develop our moral identity in a genuine way, thus, according to genuine moral reasons, values, and moral ground projects as critically assessed via the moral exposure to heterogeneous options and therefore reflectively endorsed as motives for our choices and actions. In order to clarify the algorithmic interference on our moral freedom, I have shown how algorithms can undermine the *two conditiones sine qua non* underlying the possibility for the exercise of our moral freedom, by binding our choice-behavior to probabilistically inferred patterns and constructed profiles in order to meet pre-determined heterogeneous goals, and I have shown that the threat of algorithmic predeterminism to our moral freedom is expressed in the moral and immoral interferences that algorithms can pose to the exercise of our freedom of choice and action as moral agents.

In this section, I try to develop a conceptual regulative lens in order to frame, start to address, and open the debate on the risk of algorithmic pre-determinism. To do so, I look inside the debate in AI and algorithms to understand who has tried to address from a regulative point of view the protection of freedom from algorithmic technology and specifically who specifically frame the issue of our freedom in our informational societies, where due to digitalization and datification carried out by ICTs everything can be re-thought in informational terms, as everything is captured via data and therefore produce information. As underlined above, we are nourished by information, produce information, and are describable via information, as well as our options, choices, actions, behavior, identity, thoughts, beliefs, goals, projects, affiliations, attachments: everything is information and therefore can be describable in informational terms.

The most promising theory emerging from the literature as committed to address from a regulative standpoint the protection of freedom from the algorithms-based ICTs' exploitation of users' personal data in the context of informational societies is that so-called as *informational privacy*.

Informational privacy is a theory in the field of legal scholarship based on the consideration that individuals who live in informational societies are constituted by information, and hence that a breach of one's informational privacy or a violation of it through highly invasive data mining techniques for profiling goals – as those widely detailed in the second chapter of the dissertation – should be understood, more than as a theft, as a form of aggression towards one's personal identity.[174] From a juridical point of view, informational privacy is indeed described as the right of individuals to have more control over their personal information both online and offline (onlife) as they reflect and constitute a person's identity.

Although the theories of informational privacy offer fruitful lenses to frame and understand how individuals' freedom can be eroded by algorithmic ICTs and provide legal tools to protect it, they miss to address the problem raised by the advanced algorithmic techniques as those described in the second chapter on the specific dimension of freedom at stake in the present work, i.e., our moral freedom, and so to provide, both on the conceptual side and on the practical one, the tools to firstly recognize and then to prevent (or at least mitigate) the risk of the algorithmic predeterminism described in the first section of this chapter.

[174] P. Blume, *Protection of Informational Privacy*, Djøf Forlag, Copenhagen 2002. Also: L. Floridi. "Four challenges for a theory of informational privacy". *Ethics and Information Technology, 8*(3), 2006, pp. 111.

For this reason, in this section, I start from the critical analysis of two of the main theories in the field of privacy that are specifically related to the protection of freedom with a reference of individuals' decision-making process that have found a wide agreement amongst legal scholars as fruitful conceptual lenses to understand how algorithms can affect our freedom via information: the theory of informational privacy and that one of intellectual privacy.

As represented below in figure 3, I argue how these lenses are necessary to get a basis for the protection of our overall freedom but they are not sufficient for an adequate protection of our freedom in a way that includes also moral freedom.

Indeed, I show how in the light of the goal of the protection of freedom, they miss to consider moral freedom and of the consideration of its necessary conditions for its protection. Therefore, I provide a further lens to add to the already developed lenses of informational and intellectual privacy, that is, what I define *moral privacy*, understood as a further zone of protection of our moral freedom as our freedom to choose and act as moral agents. These diverse lenses can be conceived as reflecting different layers of enjoyment of our overall freedom in line with the previously highlighted complexity and multidimensionality of the concept of freedom itself. By taking these lenses all together, these lenses can constitute a compass to provide and design specific regulative tools for the protection of our freedom – considered in its multidimensionality or complexity and in the richness of its declinations.

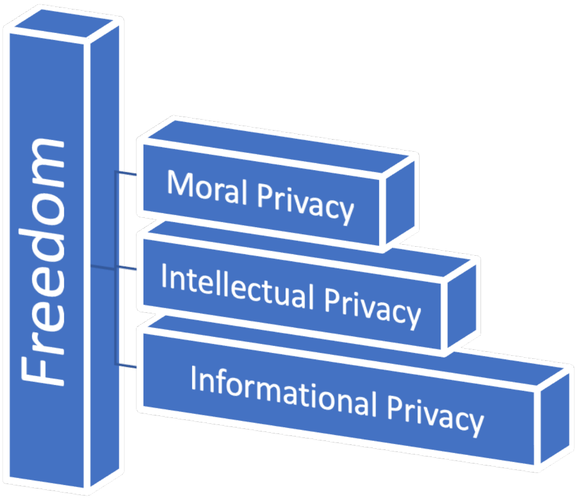


Figure 3. Three privacy lenses as tools for the protection of freedom in the hyper-profiled era

One of the main theories widely acknowledged for the protection of freedom through informational privacy in the literature developed in ethics of information technology is the one developed by Floridi, inasmuch as it provides some fruitful ideas from an ethical-regulative standpoint to rethink about the concept of privacy and its implications for the individuals in our informational societies. The choice to take this account as a primary lens to understand how to protect moral freedom in our informational societies is that this approach is the first one to frame the issue of privacy by considering the current informational characterization of our reality, ourselves, and our social relations and interactions – although this theory is more radical in its informational conceptualization than the approach I adopt, as Floridi comes to see the nature (essence) of our reality and ourselves as informational.[175]

Indeed, Floridi writes:

[175] Nevertheless, I do not discuss this specific point in this work, as it is a metaphyseal issue, that deserves much more debate and discussion to be philosophically validated. Beyond the fact that we can or cannot be informational by essence or nature, what is crucial is the fact that our information is so available today that can easily describe us, as well as, so much information is available thanks to ubiquitous ICTs and interconnected IoT that our reality is continuously captured and re-informed or re-shaped by information (we can set aside the issue whether this impact is ontological *stricto sensu* or not).

> Informational privacy requires a radical reinterpretation, one that takes into account the essentially informational nature of human beings and of their operations as social agents. Such re-interpretation is achieved by considering each individual as constituted by his or her information, and hence by understanding a breach of one informational privacy as a form of aggression towards one's personal identity.[176]

Indeed, According to Floridi, the protection of privacy should be identified as the protection of our personal identity. Personal information derivable from our data should be conceived as something that attains more to our identity than to a property that we possess: 'my data', he writes, is more like 'my hand' rather than 'my car'. His underlying idea is that "you are your information", and therefore anything done to your information is done to you, not to your belongings. [177]

> An agent is her or his information. 'My' in 'my information' is not the same 'my' as in 'my car' but rather the same 'my' as in 'my body' or 'my feelings': it expresses a sense of constitutive belonging, not of external ownership, a sense in which my body, my feelings, and my information are part of me but are not my (legal) possessions.[178]

This is because personal information plays a constitutive role in who I am and who I can become: in Floridi's theory, informational privacy and personal identity are inextricably entwined and a complete lack of privacy caused by digital ICTs means a loss of personal identity in act (who I am) and in power (who I can become). Thus, the right to informational privacy shields one's personal identity. This is why informational privacy is extremely valuable and ought to be respected.

According to Floridi, we must have the freedom to let be alone, that is, to develop ourselves via information, by reading, writing, and discussing, without the inhibition of being observed in so doing. Privacy, in this perspective, is not just

[176] L. Floridi, "Four challenges for a theory of informational privacy". *Ethics and Information Technology, 8*(3), 2006, p. 111.

[177] L. Floridi, "On human dignity as a foundation for the right to privacy". *Philosophy and Technology, 29*(4), 2016, pp. 307-312.

[178] L. Floridi, "The Ontological Interpretation of Informational Privacy. Ethics and Information Technology". 7(185-200), 2005, p. 11.

about stopping others from observing who we are, but equally of providing personal space for us to develop into 'who we are becoming'. He so writes that "our freedom is to be the masters of our own journeys, and keep our identities and our choices open", and that "any technology or policy that tends to fix and mold such openness risks dehumanizing us".[179] In my words: we must be free from interferences on our development – via information – of 'who we are' and 'who we are becoming' as in this freedom lies the possibility to form our own and unique identity (that Floridi prefers to call personality).

This personal space sounds deeply connected to his idea of "informational friction"[180]. He indeed defines privacy as a function of the informational friction in our informational environment: any factor increasing or decreasing friction will also affect privacy; lower is the friction, lower is the degree of informational privacy that can be implemented. Informational friction refers to the forces that oppose the information flow within (a region of) the infosphere and to the amount of work and efforts required to obtain, filter and/or block information[181] – where examples of informational friction are limited resources (time, computer power, access speeds), physical conditions (distance, noise), inadequate metadata and poor interfaces, lack of information and digital literacy, regulatory and copyright restrictions. In other words: informational friction is the space between what can be known of me and what should instead remain unknown. Given that digital ICTs can alter the informational friction, they have the potential to both reinforce and erode informational privacy.

This formulation lays a first basic ground to bring out the specific lens of privacy of our interest – that will be a further lens of privacy capable to secure also our moral freedom – in the idea of informational privacy as protection of our identity as freedom to develop ourselves into who we are becoming. According to Floridi, looking at the nature of a person as being constituted by that person's information allows one to understand the right to informational privacy as a right to personal immunity from unknown, undesired, or unintentional changes in one's

[179] *Ivi*, p. 310.

[180] L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*. Oxford University Press, Oxford 2014, p. 105.

[181] L. Floridi, "Four challenges for a theory of informational privacy". *Ethics and Information Technology, 8*(3), 2006, p. 110.

own identity as an informational entity, either actively, e.g., collecting, storing, reproducing, manipulating, etc. one's information amounts now to stages in cloning and breeding someone's personal identity, or passively, as breaching one's informational privacy may now consist in forcing someone to acquire unwanted data, thus altering her nature as an informational entity without consent.[182]

At the same time, Floridi just talks about human identity and the openness of our choices without further declining these concepts – specification that instead I claim to be necessary if we want to understand specifically what it means to choose and act as moral agents in our informational societies and therefore what is at stake in the undermining action on moral freedom carried out by algorithms-based ICTs. Moreover, without this further specification, it becomes also difficult to frame what specifically needs to be preserved in order to secure our freedom from interferences on 'who we are' and 'who we are becoming', as freedom to have a personal space in societies like ours that are ever more governed by algorithms-based ICTs which gatekeep, manage, and predetermine the informational availability.

Nevertheless, Floridi informational privacy is a valuable lens for our analysis as beyond the emphasis on the informational characterization of our personal identity and the conceptualization of informational privacy (as freedom to have a personal space where we can to develop ourselves into who we are becoming), Floridi also outlines how current privacy ethics is limited to the sole consideration of individual persons and not groups, while instead the privacy of groups also need a particular protection from algorithmic interferences. Groups instead should be considered as an entity as the individual in the sense of being definable by its information. In this regard, it suffices to think about the role of algorithms in constructing groups on the basis of correlations probabilistically discovered and in grouping us as profiled in them – and think of all the consequences when the probabilistic assumptions on the basis of which algorithms form group are inexact, biased, or just wrong. This means that groups are entitled to informational privacy as well, and the fact that current regulation miss of a group privacy concept is problematic to ensure both the

[182] L. Floridi, "The Ontological Interpretation of Informational Privacy. Ethics and Information Technology". 7(185-200), 2005, p. 11.

protection of individuals and the protection of collectivity.[183] Indeed, there are occasions when the group is the more natural holder of privacy rights than the individual, as in the cases where *ad hoc* collectives are discovered and constructed by ML-based algorithms and are used in domains as disparate as law-enforcement, healthcare, and retail.

The concept of privacy group becomes as a consequence very relevant for our concept of moral freedom and its protection, as the idea of informational privacy as the protection of a person's identity cannot prescind from the protection of the identity of groups on which individuals are categorized. The protection of group privacy is essential in the era of massive amounts of available data and advanced ML algorithms where data collection is often aimed at determining categories and groups of labeled individuals and this practice can pose serious problems regarding the correct treatment of groups algorithms identify.[184] Commercial profiling, for example, is based on the identification of groups (those who prefer red wine; those who listen to folk-rock music; those who live in Italy), regardless of the individual and her consideration per se. Let us think about the case of the data collected to respond to the present pandemic: the groups algorithmically detected could use as distinguish feature people infected, those who mourn someone killed by the virus, those who used the tracking app and those who did not, those who visited a park on a given day, the entire population of a country, a city, a region or even a nation, and on the basis of the constructed categories, users can receive different information and a diverse treatment. Aggregate data is key to develop solutions to contrast the pandemic and guide government decisions, such as where and when to loosen restrictive measures; but if misused they can lead to discrimination or ethically problematic decisions. Therefore, it is crucial that privacy policies and regulations are extended beyond the identification of persons (individual privacy) to also

[183] L. Floridi, "Group privacy: a defence and an interpretation". In Taylor, L., Floridi, L. and van der Sloot, B (eds.), *Group privacy: new challenges of data technologies*, Cham: Springer, 2017, pp. 83-100.

[184] Let us think about the inauspicious cases of Google's algorithms profiling users according to male/female gender and showing highly paid jobs more often to men than women, or offering criminal background checking services when doing searches that include names or typically African American surnames, are striking examples. R. Benjamin, *Race after Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Polity 2019. S. Noble, *Algorithms of Oppression*, NYU Press, New York 2019

include categories (group privacy) in order to protect individuals as algorithmically grouped into categories to be informationally re-informed and re-influenced on the basis of them. For this reason, I try to develop a lens for privacy aiming at protecting both individuals and groups from interference posed by algorithms, inasmuch as – as previously argued – the formation of a moral identity of a person cannot prescind from socio-relational dimension, and therefore, cannot be thought separately from the protection of identity of groups in which individuals are categorized and then on the basis of which their informational environment is reshaped.

The overall lens provided by Floridi is precious but also too broad in terms of understanding how to protect not just overall freedom but our moral freedom.

Richards offers a more detailed formulation of informational privacy, as a further lens beyond that one offered by Floridi, what he calls as *intellectual privacy*. Richards defines intellectual privacy as the “protection from algorithmic records of our intellectual activities underlying our free thought and expression, namely, the protection of the formation of our ideas and thoughts”.[185]

Both the privacy lenses of Floridi and Richards take the moves from privacy as theorized by Warren and Brandeis (1890) who had already realized that the value of production of some information is found not in “the right to take the profits arising from publication, but in the peace of mind or the relief afforded by the ability to prevent any publication at all”.[186] Both the theories see privacy as the right to be alone and in privacy the protection of freedom and autonomy of the individual (even if understood by both as self-governance). Floridi emphases this protection as the protection from technological interferences on the development of our identity, as our freedom to develop ourselves into who we are becoming, and the informational privacy right is framed as our right to personal immunity from unknown, undesired, or unintentional changes in one’s own identity, as well as from acquiring unwanted data that have the potential to alter or change our identity without our consent[187]

[185] N.M. Richards, “Intellectual Privacy”. *Texas Law Review*, Vol. 87, p.387, 2008, Washington U. School of Law Working Paper No. 08-08-03.

[186] S. Warren and L.D. Brandeis. The Right to Privacy. Harvard Law Review, 193(4): 1890.

[187] On the consent, contemporary data protection laws rest on what Daniel Solove calls a model of “privacy self-management” in which the law provides individuals with a set of rights aimed at enabling them to exercise control over the use of their personal data, with individuals deciding for themselves how to weigh the costs and benefits of personal data sharing, storage and processing. D.

Richards emphases this protection as the protection from technological surveillance and interferences in our process of generating ideas – such as thinking, reading, speaking with confidantes – before our ideas are ready for public consumption.

Specifically, Richards defines intellectual privacy as "the ability to develop ideas and beliefs away from the unwanted gaze or interference of others".[188] He explains that "surveillance or interference can warp the integrity of our freedom of thought and can skew the way we think, with clear repercussions for the content of our subsequent speech or writing. The ability to freely make up our minds and to develop new ideas depends upon a substantial measure of intellectual privacy"[189].

According to Richards, intellectual privacy as a right is something that turns out crucial with the algorithmic governance, as we did not have to worry on our beliefs, desires, and fantasies until recently, with the digitalization of our world by interconnected algorithmic ICTs, which can keep detailed records of our thoughts and reading habits. An example is provided by Kindle: Kindle keeps detailed records of what we buy, browse, how long our mouse rests over a word and our eyes linger over a page, what pages we underline and what the most underlined pages are, whether we finish a book, whether we re-read a book, and what passages we re-read. Some of this data serves useful functions for readers, but it also creates a detailed portrait of Kindle users as readers that, hypothetically, could be disclosed or used in harmful ways.[190] In this regard, for example, Amazon is free to sell all of its sensitive data however it wants to, at least under the current regulation, and to

J., Solove, "Privacy self management and the consent dilemma". *Harvard Law Review*, 126, 1880–1893 2013. This approach ultimately rests on the paradigm of 'notice and consent', which contemporary data protection scholars have strenuously criticized. Critics argue that individuals are highly unlikely to give meaningful, voluntary consent to the data sharing and processing activities entailed by Big Data analytic techniques, highlighting insuperable challenges faced by individuals navigating a rapidly evolving technological landscape in which they are invited to share their personal data in return for access to digital services. See A. Acquisti, L. Brandimarte, & C. Loewenstein, "Privacy and human behavior in the age of information". *Science*, 347, 509–514.

[188] N.M. Richards, "Intellectual Privacy". *Texas Law Review*, Vol. 87, p.387, 2008, Washington U. School of Law Working Paper No. 08-08-03, p. 389.

[189] *Ibidem*.

[190] This is not an entirely new problem. Librarians realized they had to know something about their patrons in order to be helpful. But they also knew that a deep ethical and professional obligation came with this knowledge. After some very good work by the American Library Association, librarians lobbied to have state laws passed protecting the confidentiality of their records. We don't have anything like that for Kindle or other digital bookstores even though we should.

see why this problematic we may think about the recent Uber scandal, where the company was said to consider authorizing opposition research on the journalists who criticized it.[191]

According to Richards, if we aim at the protection of a pluralistic society or the cognitive processes that produce new ideas, then some measure of intellectual privacy, i.e., some respite from cognitive surveillance, is essential, indeed inasmuch people feel surveilled or continuously tracked and profiled, they would become less willing to test new ideas, challenge common assumptions, and engage in rigorous debate. Therefore, freedom of thought and beliefs is the first dimension of Richards' intellectual privacy. Secondly, to elucidate how to protect intellectual privacy, the legal scholar clarifies the second dimension of intellectual privacy and evokes (like Floridi) the relation between spatiality and intellectual activity (beliefs, thoughts, new ideas development): "we often need spaces – physical, social, or otherwise – to allow us to think freely and without interference". [192] This is strictly linked to Floridi's definition of right to informational privacy as freedom to a personal space without interferences in order to develop ourselves into what we are becoming. This space is key as is the space where we can think about how we choose and to believe to what we need, desire, or aspire to; we may say, the space of retirement, reflection and reasoning. The third dimension of intellectual privacy is the freedom of private intellectual exploration. Whereas the freedom of thought and belief protects our ability to hold beliefs, the freedom of intellectual exploration protects our ability to develop new ones by reading, thinking, and discovering new truths – and this ability is both private and confidential (the act of reading must be free from interference by outsiders, and also unwatched, as surveillance of others chill the development of new idea).[193] The fourth and last dimension of intellectual privacy is describe by

[191] S. Lacy, "Uber Executive Said the Company Would Spend 'A Million Dollars' to Shut Me Up", *Times*, 14th November 2017. https://time.com/5023287/uber-threatened-journalist-sarah-lacy/

[192] N.M. Richards, "Intellectual Privacy". *Texas Law Review*, Vol. 87, p.387, 2008, Washington U. School of Law Working Paper No. 08-08-03, p. 413.

[193] The best example of how social institutions have nurtured the freedom of intellectual exploration is that of libraries. Libraries are the traditional institution in which the right to read privately and autonomously has been developed and protected. Until relatively recently, the only place available to most people for unfettered reading was the library –very often a public library provided by the government. But it is in this context that librarians developed many of the most important norms of intellectual freedom and privacy. Much of the tradition of libraries as places of private intellectual

Richards as the freedom of confidential communications. Confidentiality protects the relationships in which information is shared, allowing candid discussion away from the prying ears of others, i.e., privacy of a person's thoughts, so far as she sees fit to withhold them from others. It allows us to share our questions and tentative conclusions with confidence that our thoughts will not be made public until we are ready. We can immediately understand how algorithmic ICTs are undermining this dimension, if we think that, for example, how much our research history shows our associations, beliefs, and perhaps our medical problems. Indeed, the things we enter on Google can define us, inasmuch as our research history is practically a printout of what is going on in our brain: what we are thinking of buying, who we talk to, what we talk about. The duty to confidentiality is particularly evident when we think of certain professionals like doctors, lawyers, psychologists (and similar) – the theory of the law of confidential relations in this regard protect vulnerable parties in information transactions against abuse, including misuse of information for the confidant's gain and disclosure of confidences. Richards stresses how confidential communications are essential to meaningful intellectual privacy. Our confidants are a source of new ideas and information, but without confidentiality they may be reluctant to share subversive or deviant thoughts with us lest others overhear. Without the ability to speak with trusted confidants, we would lack of the ability to develop our own ideas in collaboration with others before we are ready to share them publicly. Moreover, consultation with intimates allows us to better determine if an idea is a good one, and to gauge some expectation of how it will be received if we finally decide to publish it. Without a meaningful expectation of confidentiality, then, we would have fewer ideas, and those that we did have might be unlikely to be shared.

According to Richards, each of the four dimensions of intellectual privacy contributes to the generation of new ideas and new ways of thinking about the

exploration in the United States is a product of the American Library Association (ALA). In 1939, the ALA adopted its first library bill of rights, a ringing declaration of intellectual freedom and privacy that enshrined the intellectual autonomy of library patrons as the heart of a library's institutional mission. Library Bill of Rights affirms that all libraries are forums for information and ideas, and that libraries should cooperate with all persons and groups concerned with resisting abridgment of free expression and free access to ideas.

world. Without free thought, the freedom to think for ourselves, to entertain ideas that others might find ridiculous or offensive, we would lack the ability to reason, much less the capacity to develop revolutionary or heretical ideas about (for instance) politics, culture, or religion. Engaging in these processes requires a space, physical and psychological, where we can think for ourselves without others watching or judging. But despite the prevailing cultural myth of the creator toiling alone, few of our ideas come from the operation of a single mind. The freedom of intellectual exploration allows us to read and to receive exposure to the ideas of others so we can evaluate them and improve or adapt them for ourselves. And at a certain point, when our ideas are ready to share with others but not yet developed enough for widespread dissemination, we might want to communicate our ideas to a few trusted friends in confidence. The freedom of confidential communications affords us this opportunity. Richards' theory of intellectual privacy is so valuable as it nurtures processes by which we as individuals can come to think for ourselves.

As Richards writes:

> It allows us to imagine, test, and develop our ideas free from the deterring gaze or interfering actions of others. Without intellectual privacy, we would be less willing to investigate ideas and hypotheses that might turn out to be wrong, controversial, or deviant. Intellectual privacy thus permits us to experiment with ideas in relative seclusion without having to disclose them before we have developed them, considered them, and decided whether to adopt them as our own.[194]

Richards outlines four practical cases in order to understand ways (and to so think potential solutions) in which intellectual privacy is increasingly under threat today:

- *Government surveillance*, very often justified in terms of the tradeoffs between the government interests in security and individual interest in privacy. Richards claims the lens of intellectual privacy allows us to rethink this "privacy versus security" framework. The underlying idea is expressed with an example: one thing is to search a person's house where

[194] N.M. Richards, "Intellectual Privacy". *Texas Law Review*, Vol. 87, p.387, 2008, Washington U. School of Law Working Paper No. 08-08-03, p. 426.

there is probable cause that a murder weapon is inside, and quite another to listen to that person's phone calls without a warrant. This is the way in which we have to think about what needs to be regulated (and by which kind of necessary justification) to secure intellectual privacy.

- *Private records of intellectual activity*, that is, the idea that the Internet has increasingly come to serve as a hub of communication, expression, and intellectual exploration and therefore of records of our intellectual records.[195] The problem is that the current use of this information is unregulated and left to be regulated privately by companies' policies. Intellectual records – such as lists of websites visited, books owned, or terms entered into a search engine – are in a very real sense a partial transcript of the operation of a human mind, and in this sense, they are fundamentally different in kind from purchases of consumer goods. Search engines, online book-stores, SNS, and so forth must provide the same guarantees to their users that libraries have for the intellectual privacy of their patrons. Therefore, information fiduciaries like ICTs, search engines, and SNS should be subject to meaningful requirements of confidentiality.
- *Government access to such records*, namely, although businesses can be reluctant to sell or donate intellectual data, the government can fairly easily compel the holders of this information to share it, for example, in investigation with criminal nexus.[196] Via ICTs, governments can obtain detailed investigation of the intellectual preferences of their citizens, allowing scrutiny of who a person's friends and contacts are, and when they called or e-mailed them. Moreover, a person whose information is

[195] It suffices to think about tech companies like Amazon serve as vast electronic bookstores; or search engines like Google and Yahoo that allow users to search the Internet for anything interests them and provide RS services, which function like a massive notebook of their users' reading interests. Such companies also keep detailed logs of their customers' activities.

[196] These have been used in the past to obtain large numbers of queries from Internet search engines. Consider a list of book purchases, library records, Web sites read, or the log of a search engine: these records reveal our interests and often our aspirations or fantasies. When such records are not kept in confidence but are instead available for access by the government, what is at stake is not merely our privacy in general, but the intellectual privacy necessary for us to engage in the freedom of thought that enables individual right to privacy.

being secretly accessed typically lacks both notices of the request and the power to challenge it. These broad powers are constrained by little or no judicial oversight or statutory regulation permitting widespread abuse and overreaching by investigators. Such records reveal not just reading habits but intellectual interests, and in the case of search-engine records come very close to being a transcript of the operation of a human mind. Indeed, when we read, we do much more than entertaining ourselves. We are engaging with ideas and information, and the act of selecting reading material is a basic act of expressive liberty, regardless of the subject matter of what we read. For this reason, it is required confidentiality for them and regulation of the practices that can access to them.

- *The introduction of reading habits in criminal trials* is the introduction of our reading habits into evidence in criminal trials: this not only makes public these private cognitive processes, but can endanger our freedom of thought, considered that reading is often an act of fantasy, and fantasy cannot be made criminal without imperiling the freedom to think as one wants. For example, if a defendant denied having the ability to make a bomb, evidence that she was in possession of multiple bomb-making textbooks could be admitted, but evidences of fantasies are often inexact and should be inadmissible, as we as should the use of reading habits to establish motive or intent, for all of the unreliability. Put it in my words: we should not be judged for our desires, fantasies or the deep needs we express in the private sphere of our life, as beyond the fact that what the reading habits track can express a thought and a desire inexact or exact (maybe what you are searching is for a friend or a family member), it should not count as an element to evaluate inasmuch as at that phase it is just an informational option amongst those we have as available, there is np prove that is something we have decided to endorse as a motive for our choices and actions, thus, something via which defining our identity as well as the motives of our choosing and agency.

All these elements bring out crucial elements that have to be guaranteed to protect our intellectual privacy, that here we see as defined also as the protection of our intellectual identity as what concerns us in terms of very private information and desires, needs, fantasies that we have but – for various reasons – that we have not decided yet to share publicly.

Both theories offer precious insights to understand how to protect our moral freedom, although they do not consider and address moral freedom and so the risk of algorithmic predeterminism triggered by algorithmic ICTs. Their lenses though constitute two fundamental layers for the protection of our moral freedom, which deserved to be clarified in order to elaborate a third-level privacy lens that allows the protection of our moral freedom.

Floridi re-interprets privacy and the right to be alone in informational terms. The right to informational privacy is a zone of immunity from unknown, undesired, or unintentional changes in our personal identity. This means, according to Floridi, freedom from collecting, storing, reproducing, manipulating our own information, as well as, freedom from acquiring unwanted data without consent, as they can alter our identity. In short, the right to informational privacy defines a zone of protection that is our personal space to develop unique identity, a space free from interferences on 'who we are' and 'who we are becoming into', therefore the protection of the informational identity of both individuals, as well as of that of collectivities/groups.

We can consider this privacy lens, i.e., informational privacy, as a *first basic ground* for the protection of our overall freedom, the protection of freedom in our deeply interconnected and technologically permeated informational societies.

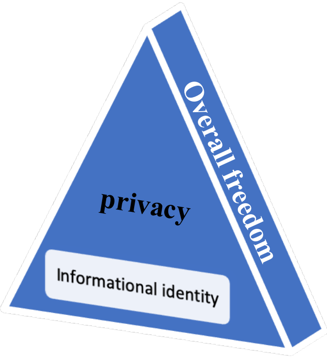


Figure 4. First privacy lens to protect overall freedom via the protection of informational identity.

Richards re-interprets privacy and the right to be alone – by building over a conception of informational privacy – as a space of protection of our intellectual activity such as the development of our thoughts and new ideas. Intellectual privacy is the right to develop our intellectual identity (my term), namely, what defines us in terms of private thoughts, fantasies, desires, and beliefs, as well as by the private conversation we have with our closest attachments. These private elements define us in a different way as persons from those we share publicly (*stricto sensu*: we 'decide' to share publicly), and we may not want to be known or be re-influenced by this specific knowledge (e.g., we do not want that algorithms construct profiles of us on the basis of that knowledge, as well as use those profiles so constructed to reshape our choice-context).

The lens of intellectual privacy – that Richards develops as underlying our free speech – can be very fruitful also for another reason (beyond those stressed by Richards), insofar as it offers a specific declination of informational identity.

With the intellectual privacy lens, he offers something more specific about our identity – as an informational identity – that we should safeguard in the light of the protection of our overall freedom: what I have defined above as our "intellectual identity", namely, everything we think and share with the closest attachments and relations or we do not share at all, such as our intimate secrets, ideas, thoughts (desires, fantasies, and we can go on and on). Intellectual identity is therefore under the umbrella of informational identity, as that information of our identity that we do not share publicly, for many reasons (e.g., the sharing of it may undermine the recognition from others), but and – more important – those thoughts, desires, and needs do not very often reflect who we want to become, but just desires or fantasies, that very often we do not choose to endorse in the light of higher goods and values, as well as in the light of long-term goals or moral ground-projects. Nevertheless, the latter may end to define us, for example, when shared with people with whom we are not in confidentiality, whose response gaze and actions can change after that confession (and with those changes, in turn we change as well). Moreover, as they can defined as when publicly disclosed, and so that piece of information as become public dominion sticks to who we are in the social-relational dimension, i.e., how we are recognized by others, that is something, as we underlined in the previous page, whose influence we cannot avoid in the development of our identity, on who we are and who we become to.

The right to intellectual privacy is therefore fundamental for moral freedom, as it should protect us from a re-definition of our choice-environment according to an identity that is private in thoughts, desires, needs, and so forth, an identity we do not are sure we want to publicly share (also for the social consequences this sharing can entails). It is important to notice that anyway these private thoughts or private intellectual activity plays a role in the deliberative process of the individuals, very often as a counter-balance between my desires and my ought to. Therefore, the pre-determination of my informational environment on the basis of the former can turn out very problematic in order to form and strengthen the latter, namely, my commitment towards myself to become "who I want to be" – as always thought in a dimension where I am not alone but I share experiences and commitments with others, i.e., who I want to be is always thought in relation to my socio-relational dimension, inasmuch as is in that dimension I live and in that dimension I act and choose as moral agents *amongst* and *with* other moral agents.

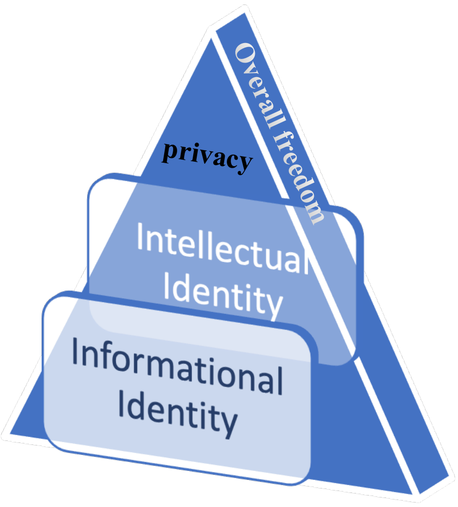


Figure 5. Second privacy lens to protect overall freedom via the protection of intellectual identity.

My *ought to be* goes beyond my informational and intellectual identity as it stands for who I want to become or develop into on the basis of the values I choose to embrace or endorse, namely, the development of my *ought to* stands for the development of my moral identity.

My moral identity goes beyond the conception of informational identity, as if my information can constitute my personal identity not all my information defines me as a moral agent, but only that information (or option) that I have decided to endorse as motive or reason for my choices and actions. In this sense, information on “my hand” does not define my moral identity, as well as information standing for a fantasy or desires that – even if I have – I have decided to do not endorse (e.g., think about them as available informational options that I have decided to do not choose as motives for my actions) does not define my moral identity as I do not have endorse that kind of information as motive for my choices. An example can be provided by the fact that I renounce to cede to a passion outside my marriage as I want to be faithful to the promise I made to my partner. In this chase, I have not endorsed a fantasy as a motive for my choice by instead privileging a long-term project, value, and goal. In this example, my moral identity is not connoted by the fantasy that I have, but my moral identity as moral agents is connoted on the basis of the option (information: value, reason, project, and so forth) I have endorsed, i.e., the value that I have chosen to which oblige my conduct. The development of my moral identity as the formation of my ought to requires my endorsement of a certain option – as a motive for my choices – amongst available options (amongst which there are also desires, needs, short-terms goals, fantasies and so on: everything can connote my identity both as informational identity – so my general personal information – and as intellectual identity, i.e., the information I still not have chosen or endorsed). So, it is in the act of endorsement that I define what information or option connotes my moral identity (at least actively), by endorsing it as motive on which steering the development of my moral identity.

The protection of moral identity is anyway informational; therefore, it requires informational privacy as a first layer or zone of protection. Informational privacy though is not enough, it is a starting informational ground in which we need to discern what really pertains to our choosing and acting as moral agents and what does not, as above pointed out.

At the same time, the protection of moral identity requires the protection of intellectual identity, as what we need on choosing on, as options on which we need to deliberate, but by which we do not want to be necessarily defined before our key approval or endorsement. Intellectual privacy is a part of our personal identity, but only what we decide to endorse defines us as moral agents (that is, how we take a moral stand) becomes part of our moral identity and so forms our ought to, i.e., our obligation to certain values and moral reasons, rather than to others. We can endorse a desire or a fantasy as a reason and motive for our choices and actions, those options in turn will define our identity in a certain way, rather than another. Or we cannot choose those as a motive for our actions, to privilege instead something else, such as a long-term project – as underlined above. Thus, intellectual identity stands for that very private information that constitute our personal identity but that we have not decided to choose yet, thus it plays a key-role and has to be protected in the light of our moral freedom; vice-versa, i.e., the algorithmic use or the social exploitation of that kind of private information to construct our profiles – on the basis of which shaping our choice-context – may result in defining us without our approval and therefore to subject us and determine our moral identity according to options that we have not chosen for our choices, i.e., in a way that undermine our freedom to choose and act as moral agents (so according to options we can endorse as authors of our choices and actions). In this sense, if ML algorithms can capture that kind of information and use to redefine or pre-determine our choice-behavior on the basis of profiles constructed on that, they would end to undermine our moral freedom – i.e., it suffices to remind the argumentation carried out in the second chapter on the possibility of ML algorithms to infer choice-driving elements such as vulnerabilities, deep desires, weaknesses, and so on, and use them to reshape options and target the user in a way that can suspend her endorsement.

Thus, the protection of our intellectual identity is crucial for the protection of our moral freedom, as it stands for what is private and very often remains private not just because we do not publicly share, but because we decide to not choose as a part of who we want to become.

At the same time, the protection of our intellectual identity or the exercise of the right to intellectual privacy is not enough, as it does not secure at a minimum threshold our possibility to choose and act as genuine moral agents; intellectual privacy just secure the protection from interferences on the formation of new ideas, thoughts, the protection of our intellectual activities as a space of confidentiality we deserve on what we do not want to share publicly, is that what Richards defines as underlying the protection of free speech. But though free speech, as well as our freedom to develop new ideas, to explore, to have confidentiality on desires and fantasies, is important, it does not secure at a minimum threshold the protection of the development of our moral identity and our ought to: intellectual privacy cannot secure our moral freedom.

Thus, if the informational privacy and intellectual privacy are not enough to protect our moral identity, what do we need in order to safeguard at a minimum threshold the development of our moral identity?

As I underlined, the protection of our moral identity is the protection of the development of our ought to, namely, what corroborates our moral dispositions into our moral posture, into our moral identity, i.e., the development of the obligation to the values and moral reasons we decide to embrace, by endorsing them amongst a plurality of morally heterogeneous options: in other words, the protection of our moral identity is the protection of our moral freedom.

As the informational privacy and intellectual privacy lens protects just two layers of our overall freedom, but do not address and include the protection of our moral freedom, I need to elaborate a more specific lens to protect moral identity. This lens is what I define *moral privacy* or the *right to moral privacy*.

Moral Privacy is what detects a zone of protection for our moral freedom.

In order to highlight what this zone should include, we need to recall the key or better necessary but not sufficient conditions of moral freedom I underlined in the first section and on the basis of which I argued the impact of algorithmic ICTs on moral freedom in the second section: a) the availability of morally heterogeneous options and b) our moral autonomy. Moral privacy needs to safeguard these two conditions to meet the protection of our moral freedom at a minimum threshold.

As it is implicit in the definition of these two *conditiones sine qua non*, indeed, they are necessary but not sufficient to exercise fully our moral freedom. At the same time, moral freedom is multidimensional and does not imply a black and white scenario. This means that even if we can secure these conditions, we can nonetheless enjoy situations with more or less moral freedom, depending on the algorithmic impact or interferences we can experience on the various elements that moral freedom requires, e.g., from the influence on our choice-environment, our available options, and our capacity to exercise our reflective endorsement to the interferences on the social-political conditions in which we live that can favor more or less openness and contamination to other societies.

Therefore, moral privacy is what defines a *zone of protection* for **a.** the moral heterogeneity of available options from a narrowing impact poseable by algorithms via information (as moral heterogeneity of our options can secure at a minimum threshold our possibility to choose and act alternatively as moral agent, therefore, to develop our own idea of good, reason and test the values and moral reasons we embrace and decide to endorse via an heterogeneous moral exposure), and **b.** for our moral autonomy, that is, our capacity to endorse as motives the information or options (values, moral reasons, and ground projects) we value into our choices and actions (as the endorsement can secure the authorship of our choices and actions). Moral privacy is therefore a zone of protection of our freedom of choice and action as genuine moral agents.

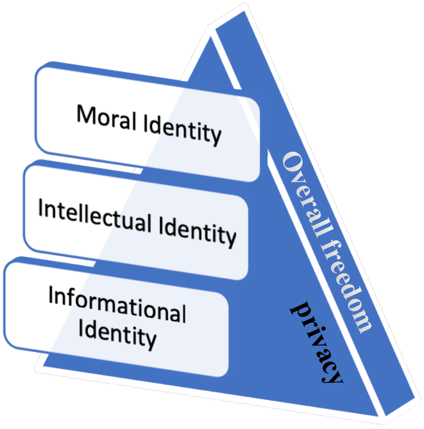


Figure 6. Third privacy lens to protect overall freedom via moral identity.

In this sense, moral privacy is – following the broad definition provided by Floridi – the protection from undesired changes via information that can alter our identity, but more specifically, from everything that can interfere (e.g., from algorithmic collection, storing, recommending) with the genuine development of our moral identity, and therefore from anything can interfere with our possibility to exercise our reflective endorsement so much to suspend it and therefore determine at our place our choices and actions. Moral privacy is therefore also the protection from what can reshape our choice-context by narrowing the moral exposure we require for the development of genuine moral identity by binding our choices to options for the achievement of pre-determined goals/criteria.
In sum: the ideal that underpins the right of moral privacy is that of *freedom from algorithmic pre-determinism*.

Freedom from algorithmic pre-determinism concerns both individuals and collectivities (groups) and means freedom from being predetermined according to predicted patterns and profiles constructed via probabilistic inference in order to meet third-party goals.

Let me further specify it.

For the individuals: freedom from algorithmic predeterminism lies in the freedom from being profiled with the goal of inferring valuable feature exploitable to interfere with individuals' choice-behavior in order to align it to pre-determined heteronomous goals. For collectivities (or groups): freedom from being profiled or constructed on the basis of features that are relevant with regard to heteronomously pre-determined goals and therefore from being profiled with the goal of discovering valuable patterns and driving choice-elements via the algorithmic a/b testing within the constructed group. For both individuals and collectivities (groups), freedom from algorithmic determinism is an ideal that grounds on treating individuals and groups as end-setters, and never as means to achieve pre-determined heteronomous goals or tasks, above all, when these overlaps with algorithmic maximization of click or economic-benefit of third-party interest at the expense of the identity of the individuals and the groups. This does not mean to exclude profiling, as well as the use of RS, algorithmic classifying and filtering. This means that they should be designed in a way that does not learn how to meet certain goals by using individuals and groups as means or tests to confirm the predictive potential of the probabilistic correlations and patterns inferred to meet certain pre-set goals. This means that algorithms should not mine users' data to construct profiles, attach labels to persons and groups, and so to categorize them depending on heteronomous pre-determined goals on the basis of which to bind users' choice-environment to them, for example, via determining their choice-options or triggering emotions by targeting techniques.

The ideal of freedom of algorithmic pre-determinism responds to a positive concept of moral freedom, as formulated in the first chapter, namely, our freedom to become genuine moral agents and therefore to develop genuine moral identity.

To operationalize the right of moral privacy we use the *conditiones sine qua non* elaborated above, as they can identify what has to be secured to prevent algorithmic pre-determinism in our informational societies, therefore they play as an evaluative standard to understand whether there is the risk of algorithmic pre-determinism. By detecting what has to be secured to guarantee the exercise of our moral freedom at a minimum threshold, these conditions allow also to develop the main axioms to decline and operationalize the respect of moral privacy and so the protection of moral identity.

To bring out these axioms I rely on the argumentation provided in chapter 2 (broadly speaking, the impact of ML algorithms on the heterogeneous availability of alternative options and moral autonomy as self-relational determination), as well as from the insights taken from the two privacy lenses above analyzed. Specifically, considering that 1. in that argumentation I have shown two scenarios opened by the algorithmic impact on users' choice-contexts that highlight two different kinds of algorithmic interferences (i.e., moral and immoral interference) and 2. considering that I have then claimed that the algorithmic predeterminism's undermining action on our moral freedom lie in moral and immoral interferences that algorithms can raise individuals' choice-behavior and that 3. if the algorithmic profiling is difficult nowadays to avoid given its pervasive application, these interferences are instead something that can be regulated and that if regulated they can mitigate the impact and consequences of algorithmic predeterminism, I have elaborated three moral privacy's axioms as a way to address to the moral and immoral interferences that algorithms can raise and in doing so endanger our moral freedom.

Moral Privacy 1st axiom declines the negative concept of moral freedom as freedom from immoral interference and consists in removing any algorithmic interference that labels individuals and groups according to wrong and illegitimate assumptions. This 1st axiom requires two aspects:

1) A technical aspect, i.e., the accuracy of probabilistic assumptions on the basis of which construct algorithmic profiles of individuals and groups. These assumptions (as well as patterns and valuable correlations) do not have to be derived by using individuals as tests to validate the predictive potential of inferred knowledge and above all by exploiting individuals' sensitive information inferable by their data to test the probabilistic patterns. These probabilistic assumptions and patterns can be tested on synthetic data[197] (that are cleaned from biases and that are not referred to specific individuals).
2) Epistemological aspect, i.e., knowledge of the assumptions behind the profiles and labels attached to individuals and groups. This means that the individual has the right to know her profile (the correlations behind the construction of that profile), as well as if she has been categorized in a certain group, she has the right to know what to be part of that group means (what are the common features on the basis of which the group is formed and what to be part of that group entails in terms of options I can or cannot have access to) so to easily contest it and ask to adjust it, or to contest it and ask for the elimination of that profile or grouping.

Immoral interferences (that are based on illegitimate or wrong assumptions) can constitute the most severe impact of algorithmic predeterminism as they are the main form of constraint of options both as information and real chances, as well as

[197] Synthetic data is artificial data that is created by using different algorithms that mirror the statistical properties of the original data but does not reveal any information regarding real people.

they are the main form of algorithmic injustice (see the debate in chapter 2). As previously outlined, in the hyper-algorithmic era a situation is preferable where my profile is morally legitimate, namely, is not based on wrong, biased and inexact assumptions, inasmuch as at least it can reflect our moral orientation, interests, values, and the options we have endorsed. Since algorithms are self-learning and we can exercise a control over our algorithmically constructed profiles, algorithms can tailor themselves on the basis of the choices we make and so the changes in our moral identity, as well as our feedbacks in reply to the disclosure of our profiles. Nevertheless, as previously pointed out, this preferable accuracy of our profiles can raise the problem of an excessive personalization of our choice options that in turn can undermine our availability of morally heterogeneous options (i.e., *first conditio sine qua non*).

For this reason, *moral privacy 2nd axiom* declines the negative concept of moral freedom as freedom from moral interference and consists of regulating any algorithmic interference that labels individuals and groups also when interferences are developed according to legitimate assumptions (i.e., assumptions that can track users' interests, goals, values, and so forth).

This 2nd axiom requires that even if moral interferences reflect individuals' orientation, they should always be regulable according to the users' choice. The user becomes who decides by default if algorithms can use that profile or not, even if it responds to her interests and goals. Indeed, in the hyper-profiled era in which we live, anonymity is not an option anymore and is not the solution to the excessive personalization carried out by algorithms, so we have the right (and maybe the duty) to determine our environment on the basis of who we are and above all who we would like to be.

Moral privacy 3rd axiom requires that algorithms are designed and trained to act like professionals (like physicians, lawyers, psychologists) who have the moral and legal duty of confidentiality to their users. This means that if some high-sensitive information (vulnerabilities, pathologies, weaknesses, trauma, and so on) can be inferred probabilistically or from health data (even if in some cases they are protected as a special category, e.g. in Europe by the GDPR), algorithms have the duty to inform the user about what discovered and do not sell, share, or disclose that data at any cost. This also means that algorithms should be trained on synthetic data to be able to understand and discern what inferences can be morally loaded as high-sensitive inferences and be trained to discern them on synthetic data.

Here one can say that what is sensitive for me is not for another person. Synthetic data can be constructed and trained on the basis of highly representative case studies on what can count as high-sensitive and morally loaded, but algorithms at the same time should be also designed in a way that can interact with the user, e.g., without asking if the information discovered is sensitive or not, they can ask instead if she wants to remove a specific inference or informational option – whatever is the reason – and they cannot consider that action of removal, for example, as something on which deriving individuals' choice-driving elements or some specific connoting features. Indeed, the algorithmic duty of confidentiality should extend from the high-sensitive information to the users' management of that information in response to the algorithmic questioning.

These three axioms are per definition general and regulative, as they have to play as criteria on which to rethink the design, the use, and deployment of ML algorithms and specifically of algorithmic profiling, filtering, and recommending in a way that prevents or mitigate the threat of algorithmic predeterminism at least at a minimum threshold. These three axioms indeed require to be operationalized in specific contexts and so they play as guidelines that should spur engineers and IT designers in thinking on how to technically translate them to protect the normative ethical value of our moral freedom, as highly fundamental at both the individual and at the collective level (see chapter 1). In this sense, moral privacy is the tool for the operationalization of the ethical-normative value of our moral freedom and hence for safeguarding individuals as autonomous and free to develop genuinely their moral reasons, values, projects, critically test and embrace them in a way that preserves the moral authorship of the individuals while keeping them open to the influences and the reasons, projects, and values developed by the others.

Therefore, these three axioms can safeguard the exercise of moral freedom at least at a minimum threshold, by safeguarding the *conditiones sine qua non* that underlie our freedom of choice and actions. Indeed, if implemented, these axioms define three-level of mitigation of the threat of algorithmic predeterminism and so of the protection of our moral freedom. The first level is the most important and it is delineated by the first axiom as the protection of individuals from the construction of profiles of them that can be based on illegitimate, wrong or inexact, and biased assumptions (moral freedom as freedom from immoral interferences).

The detection and elimination of immoral interferences and thus of profiles based on morally wrong assumption via the disclosure of profiles is fundamental to prevent discrimination and injustice phenomena and allows the subjects to exercise their power to know and intervene (and so to fight the epistemological and the power asymmetry between the profiler and the profiled) on what predetermine in an illegitimate way their choice-environment and so their alternative options both as information and also as real life-chances. Therefore, the first moral privacy axiom would prevent the algorithmic impact on the *first conditio sine qua non* – above all according to the second scenario delineated, i.e., when the profiles do not respect users as specific persons. The second moral privacy axiom safeguards users from the risk of ending to moral echo-chambers as a consequence of the correct tracking of their interests, values, and moral orientations (first scenario delineated), by giving to the users the power to refuse or ask for the regulation or the non-consideration of their profiles even if they are correctly constructed. Although the profiles indeed can correctly identify the users' orientation, the disclosure of the reasons behind the construction of a certain profile or the categorization of an individual into a certain group can make the individual aware of personal characteristics, dispositions, and beliefs of which before she was not aware of and that she does not really want to embrace or towards which she does not want to steer her behavior with regard the person she would like becoming. In this sense, the disclosure of moral dispositions as traced online and offline by users' interaction and research can lead the user to critically question whether they are aligned to what – projects, values, goals – they really want to embrace and the persons they want to become and so to become aware of whether the group of people or the environment they live is determining their behavior towards a direction they do not genuinely share. Via the second moral privacy axiom they have the power to become aware of that and have the possibility to change their profiles in a way that is more aligned to what they inspire to become. In this way, the second moral privacy axiom would prevent the risk of moral echo-chambers and bubbles as a consequence of the impact raised by algorithms (first scenario, i.e., moral interference) on the *first conditio* underlying moral freedom, i.e., the availability of morally heterogeneous options, by mitigating the risk of narrowing individuals' moral exposure and contamination by informing the user about how she is profiled and how she can do to change the informational options displayed to her. Furthermore, both the respect of the first axiom and the second one can protect the second *conditio sine qua non* of moral freedom, i.e., moral autonomy, inasmuch as the axioms recommend to not carry out the test of the predictive potential of ML algorithms by using individuals as means to achieve pre-determined goals and by exploiting their profiles and their categorization into groups to derive individuals' high-sensitive information, but to rely on the use of synthetic data as a promising field that can be further expanded as highly precious to preserve individuals' privacy. Though, the deep protection of moral autonomy as endorsement lies in the third moral privacy axioms that asks to design algorithms as confident that even when the profiling happens on individuals' data and not on synthetic data and this leads to infer highly-sensitive information, they have the duty to keep that information confidential at any cost. In doing so, they would highly prevent the exploitation of highly-sensitive characteristics and choice-driving elements that can lead to trigger individuals' emotional responses as the main causes of the phenomenon that I called the suspension of endorsement.

By operationalizing the moral privacy criteria above underlined, it sounds possible to preserve the two conditions of possibility detailed in chapter 1 for the exercise of moral freedom and so the protection of moral freedom as an ethical-normative value via the elimination or mitigation of moral and immoral interferences to our capacity to choose and act as authentic moral agents and therefore to our capacity to develop genuine moral identity.

Obviously, at the beginning, these axioms may decrease the economic potential derivable from users and group's exploitation as means to derive valuable patterns, but the long-term effect can be more beneficial and in line with both the users' goals and third-party interests. Indeed, designing algorithmic environments that can make people and groups to feel more recognized per se and protected from invasive, inexact, and unjust interferences (i.e., an environment that fosters individuals as moral agents) can boost and incentivize people's use of algorithmic ICTs in a way that can bring benefits to both technological companies and individuals, above all, in their power to choose and become who they want to be, and so by harnessing the technological potential to increase their heterogeneous exposure rather than the contrary, i.e., their capacity to choose and act according to the values they have formed, encountered, test, and embrace in a way aligned to how they want to form their ought to and their moral identity: an 'how' that only by exposing ourselves to morally heterogeneous relations, practices, ways to do things, we can understand, develop, and endorse as authors of our identity and lives.

To sum up, in this section, I have tried to show that there are practical reasons why we ought to operationalize moral privacy as a privacy tool to protect our moral freedom. The present section represents a first step in this direction and recognize that much work needs to be done to prevent the risk of algorithmic predeterminism in our informational societies. My goal indeed was that to begin a conversation about moral privacy and suggest reasons on why taking it seriously. Although unconsidered so far, moral privacy is crucial to safeguard what it takes to be free to choose as moral agents in our contemporary informational societies, and therefore, to be free to choose in the algorithmic era.

III.3 Champions of Moral Freedom

In the previous section, I have defined a further lens to protect our overall freedom in a way that includes our moral freedom: what I defined as moral privacy, as the protection of our freedom to become genuine moral agents and therefore to develop a genuine moral identity in the socio-relational dimension in which we live and act as moral agents between and with other moral agents. Moral privacy is a conceptual lens I develop drawing insights from the theory of informational privacy and the theory of intellectual privacy, as both precious but not sufficient to address the potential or actual threat of algorithmic pre-determinism we can experience in our informational societies, as more and more interconnected and algorithmically governed. Specifically, our right to moral privacy stands for our right to choose and act as moral agents in our informational societies (our right to moral freedom) and specifically is defined as our freedom from algorithmic predeterminism.

In this section, I define and call into action who are the main agents of moral freedom, those agents called to operationalize the protection of moral privacy: the algorithms' designer (as well as the company providers) and institutional decision-makers. I argued indeed why our moral freedom ought to be protected, what instead I have not defined is what are the legal, social, or technological structures that can build this protection.

The algorithmic designers should have the duty to design algorithmic ICTs according the ethical-normative value of moral freedom, that specifically means to operationalize the three axioms of moral privacy as elaborated in the previous section, while on the other side, the institutional agents are those who are called to supervise whether the algorithmic ICTs in use comply the moral privacy's criteria and therefore respect the ethical-normative value of moral freedom.

Let me synthetically sum up the moral privacy's axioms that ICTs designers and institutional decision makers have to meet in order to respect the ethical-normative value of moral freedom:

- *Moral Privacy 1st axiom* declines the negative concept of moral freedom as freedom from immoral interference and consists of removing any algorithmic interference that label individuals and groups according to wrong and illegitimate assumptions.
- *Moral privacy 2nd axiom* declines the negative concept of moral freedom as freedom from moral interference and consists of regulating any algorithmic interference that label individuals and groups according to legitimate assumptions (i.e., assumptions that reflect their moral orientation).
- *Moral privacy 3rd axiom* requires that algorithms are designed and trained to act with users as professionals (like physicians, lawyers, psychologists) who have the moral and legal duty of confidentiality to their users. This specifically means that even if some high-sensitive information (vulnerabilities, pathologies, weaknesses, trauma) can be inferred probabilistically or from health-related data

> or by proxy, algorithms have the duty to inform the user about what has been discovered and above all do not sell, share, or disclose that data at any cost. This would also require that algorithms would be trained on synthetic data – when it is possible – to be able to discern those high-sensitive inferences in general, and therefore not on the basis of users' history and interactions.

The *algorithmic ICTs designers* have a huge role. They have the right to think about a design of algorithmic ICTs that can prevent to predetermine users' choice behavior. Therefore, ICTs designers and architects have the duty to design them in a way that deeply considers the risk of algorithmic pre-determinism and operationalizes the axioms of moral privacy previously outlined in order to prevent it. This means that ML should be designed in order to inform the user about her profile and her categorization. In order to ensure the right to moral privacy, they have to fulfill the duty of designing algorithms that can disclose profiles of groups and people, as well as to justify the way in which that profiling has been carried out. This justification requires a low degree of explainability, therefore it cannot infringe the company's trade secret as well as it does not require the full transparency of the code and algorithmic functioning. At the same time, the justification provided, that means, the disclosure of the particular patterns or correlations behind the creation of a certain label or categorization, can lead users to understand which are the profiles according to which they have been labeled, on what basis these labels and profiles have been discovered and constructed, so to be able to understand on what basis to contest or refuse that profile or categorization, along with the key-understanding of whether that profile is legitimate or unjust. In the former case, individuals should be able to decide whether to accept it (by asking for a regulation or the possibility to intervene on it by just adjusting it), while in the latter they can exercise their right to immediately ask for the removal of that profile. Algorithmic ICTs designers have the duty to train ML algorithms that learn how to discover correlations and patterns on synthetic data (that are accurate, cleaned from biases, and not referred to real persons) so to prevent the risk of algorithmic pre-determinism, i.e., programming algorithmic tasks which learn how to achieve pre-determined goals at the expense of individuals' wellbeing. Algorithms have to be trained on synthetic data also to discern morally-loaded high-sensitive information: to do so they have to be trained according to a huge set of synthetic data that incorporates a heterogeneity of cultures, practices, and values. Last but not least, algorithmic ICTs designer have to develop design algorithms capable to interact as a confidant with the users. This means that algorithms should be designed in a way that may inform the user when something that has been inferred can be morally relevant or problematic to her, as well as to do not use the inferable information about why the user decide the removal of that informational option. Algorithms have to be designed by embedding the concept of confidant, where confidentiality extends to particularly high-sensitive options discovered by data (of which detection allows the user to ask the removal from consideration) or by proxy, as well as it should be extended to the action we perform to manage, or in response to, the particular high-sensitive option inferred. Nevertheless, ICTs' designers are not the sole to be entitled to work for the respect and protection of our moral freedom. Though their role is crucial in the design of algorithmic ICTs in a way that respects individuals' moral freedom, when it takes to tackle the impact of algorithms on moral freedom due to the complexity and pervasiveness of algorithmic application in our contemporary societies, we need to take into account a distributed scenario of duties and responsibilities amongst plural agents. Indeed, also what we can define as *institutional decision makers* play another important role in the protection of moral freedom via the moral privacy's lens.

For institutional decision-makers, the right to moral privacy ought to be a criterion of discernment in the initial stages – when they are called to adopt algorithms-based ICTs in crucial social sectors (such as justice and healthcare) and to discern the best deployment option in order to comply with the ethical-normative value of moral freedom – as well as later in the process, to exclude or prohibit an algorithmic ICTs when a person's request cannot be fulfilled and thus the duty of the algorithmic providers is not fulfilled. Moreover, institutional decision-makers have to supervise users' requests to contexts the application of algorithmic ICTs when they produce interferences that are based on morally wrong assumptions and make users' requests effective. To do so, institutional agents are called to create the societal sub-structures to assist users to prove their claims, support their requests, and carry on contestations towards algorithmic interference, above all, when those are against algorithms' providers and so particularly expensive in terms of personal and economic efforts. Both institutional agents and technological agents are called to take the issue of algorithmic predeterminism seriously. Indeed, the rise of individuals' distrust on algorithmic technology as a consequence of all the concerns recently underlined may turn out a huge loss in opportunity at the individual and societal levels. Although the present work has mainly focused on the critical side of algorithmic ICTs on moral freedom, it is worthy to underline how designing algorithms that can meet the ethical-normative criterion of our moral freedom by operationalizing moral privacy's axioms would result not only in the boosting of us as individuals, as autonomous agents, but it would also mean securing and boosting the moral fabric of our social sphere towards the possibility of more cooperation, open dialogue, and sharing of commitments on common moral goods. Both the moral dimension of the individual and of the social sphere can take huge benefit from the design and the deployment of algorithms in a way that respects the ethical value of moral freedom. In this last section, I have clarified that algorithmic designer and institutional decision-makers have the duty (and which kind of duty) to fulfill the right to moral privacy according to the criteria pointed out, respectively the former in the design of algorithms-based ICTs, while the latter in the decision-process of the phases of their adoption, regulation, or exclusion.

Although ICTs' designers and institutional agents are the main agents that have the duty to protect the moral dimension of the individuals and of our living together by algorithmic ICTs' design and adoption, they are not the sole ones. I have argued indeed how the rise of the algorithmic predeterminism and more specifically the exploitation of the predictive knowledge developed by algorithms on individuals and its use to rule evermore social domains in a way that very often is not aligned to persons' goals, values, and long-term ground projects but to third-party goals is a serious concern. This means that it requires more than the sole action of ICTs' designers and institutional decision-makers. It demands for the re-thought and the adoption of a broader design perspective of collective responsibility on how to secure and promote the ethical value of moral freedom in our contemporary informational societies, where we as individuals are not the sole agents anymore, but share our environment and tasks with artificial entities more and more capable to perform choices and activity autonomously while influencing deeply if not predetermining how we make our choices, actions, and develop our identity.

This work has tried to show the ethical value of our moral freedom not just as a philosophical concept but as our capacity to choose and act effectively in our informational societies, in the light of what we want to value and who we want to become from a moral standpoint. The work has tried to highlight what it takes at least at a minimum threshold to become genuine moral agents in our contemporary informational societies, and specifically, what it takes to become genuine moral agents when novel threats or forms of impediment can be raised by the exponential progress in the potential and use of algorithmic technology and threat to undermine this great ethical value; a fundamental ethical value both for us as individuals and specifically as moral agents and for the moral flourishing and the moral progress of our societies: an ethical value that is crucial for our possibility to decide which kind of moral agents we want to become, and – in this way – to what kind of societies we want to contribute from a moral standpoint, namely, which kind of moral progress we want to initiate or carry on as moral agents, by choosing and acting as genuine moral agents, agents that are free to embrace and pursue certain values rather than others, namely, agents that are free to choose and act in the algorithmic era.

CONCLUSION

The overall purpose of this dissertation is precisely to investigate one of the most crucial issues in our contemporary liberal-democratic societies in the light of the recent advancements in algorithmic technology: that issue of our freedom, and specifically, our moral freedom. The specific proposal underlying this work is that the algorithmic governance that is structuring by design in our mature informational societies is creating a new form of impediment to our moral freedom, what I called algorithmic predeterminism, by silently undermining those necessary conditions (or *conditiones sine qua non*) which secure our moral freedom at a minimum threshold and so that enable our freedom to choose and act as moral agents and thus become genuine moral agents.

To argue my proposal and achieve my goal, I articulated the dissertation in three main chapters.

The first chapter focused on the elaboration of an adequate account of moral freedom to recognize properly what it means to choose and act as moral agents, and therefore to evaluate the impact of algorithms-based ICTs on our freedom to choose and act as moral agents in our informational societies. In the first section of the chapter, I have shown the complexity and multidimensionality of the concept of freedom and the shortcomings characterizing the huge and heterogeneous debate on freedom when it deals to distinguish moral freedom from free will and from socio-political freedom. By drawing insights from theories developed in free will debate and socio-political one, I have defined the concept of moral freedom in both a positive and negative sense. Moral freedom in a positive sense is our freedom and power to become moral agents, and specifically genuine moral agents, that means, our freedom to develop genuine moral identity. I have also argued that freedom to develop genuine moral identity means freedom to develop genuine moral reasons, values, and ground-projects and so the freedom to choose and act in accordance with moral reasons, values, and ground-projects as options I can endorse as genuine motives for my choices and actions. I have then argued also that there is also a negative sense of moral freedom, that is, freedom from potential and actual, moral and immoral, interferences to develop a genuine moral identity, where by moral interference I meant an interference on us that even if it reflects my interests, values, and – broadly speaking – my moral orientation is nonetheless moral-freedom conditioning as it interferences anyway with the possibility to choose according my own reasons and values, and whereby immoral limitation I meant an interference on us that is morally illegitimate, wrong, or unjust as it does not consider our moral values and reasons (moral orientation) .

In the second section, I have brought out what conditions, not sufficient but at least necessary, need to be guaranteed to make the exercise of our freedom of choice and action as moral agents possible and so to evaluate whether algorithmic ICTs are affecting our moral freedom in both a positive and negative sense. I have highlighted, by drawing insights from theories developed in the free will debate and in the socio-political one, what are the *conditiones sine qua non* underlying our moral freedom as securing at a minimum threshold its exercise. The first condition for the exercise of our moral freedom is the availability of morally heterogeneous options, according to which an agent can develop a genuine moral identity if and only if the context of choice in which she chooses is characterized by morally heterogeneous options, i.e., options that embed plural moral values and reasons that are diversified, that in turn requires sufficient moral exposure of the subject to social relations, attachments, values (and so forth) morally diversified – insofar as they allow her to develop diverse, but (as critically tested and then endorsed via choices and actions over time) genuine, moral identity. The second condition is that of moral autonomy, according to which an agent can develop a genuine moral identity if and only if she can be the author of her moral identity by reflectively endorsing, amongst the morally heterogeneous options available, those values and reasons she embraces as motives for her choices and actions. Since the latter condition strictly requires the former, as we live and choose as moral agents among other social agents, I have shown how moral autonomy cannot be adequately thought as a full self-determination, as we cannot prescind from the role of social-dimension for the exercise of freedom of choice and action as moral agents; therefore, I have clarified that a sound definition of moral autonomy needs to recognize how the social and relational dimension inform the self-determination of the subjects: therefore, I have defined moral autonomy as relational self-determination.

In the third section, I have argued the value of our moral freedom for the flourishing of the moral dimension of both the individuals, insofar as moral freedom concerns the development of moral values and the dimension of ought to that oblige us to them, and of our living together, as moral freedom boosts moral openness and a culture of reason-giving underlying social dialogue, cooperations, and sharing of social commitments. Therefore, I have finally argued why we ought to protect it as ethical normative value from existing and novel forms of impediment, as that one posed by algorithms I aimed to show.

In the second chapter of the dissertation, I have questioned whether the algorithmic governance that is structuring into our reality in the light of its disruptive impact may hamper our moral freedom, and I have carried out an analysis on whether algorithmic ICTs can affect those necessary conditions underlying the exercise of our moral freedom as freedom of choice and actions as genuine moral agents.

In the first section, I have shown how algorithms governing the functioning of our ICTs are becoming architects of our choice-contexts, i.e., the contexts in which we make our choices, perform our actions, and develop our identities, insofar as they re-shape and structure the set of available options displayed to the users. I have argued how due to the digitalization of our societies, everything is datified, and therefore everything can be captured via information, and information, and specifically informational contents, embed in turn values, beliefs, attachments, thoughts (and so on): every information can be considered as a reason that can be endorsed as a motive for our choices and actions. This means that informational contents can be considered as options we can choose, and to the extent algorithms as gatekeepers manage information, algorithms are structuring our options and so reshaping our choice-contexts, by determining what is available or unavailable to us basing on our algorithmically constructed profiles, so generating algorithmic choice-architectures. I have then argued how algorithmic choice-architectures can affect the two *conditiones sine qua non* underlying the exercise of moral freedom.

Firstly, I have considered what I have defined the epistemological problem posed by algorithmic reshaping of our choice-contexts, i.e., the epistemological impact raised by algorithms on individuals by pre-determining the number and the qualitative kind of options available to the individuals on the basis of the inferred profiles on them: I called this impact as the algorithmic *hetero-definition of the availability of options* and I have underlined that the definition of our availability of options is hetero-connoted, i.e., depends on external forces and probabilistic assumptions on which we do not have the power to intervene, and that this hetero-definition is by default, insofar as algorithms by selecting the relevant information for us pre-determine the conditions of our choices and restrict the range of available options. Put it in other words: I shown how algorithms are the external forces which determine what option – which in turn embeds values, reasons, and therefore potential alternative courses of action – from being potentially considerable into my choice availability of options is effectively actualized as a part my availability of options, as part of my context of available options. This means that at the root of the algorithmic epistemological impact on the individuals there is an algorithmic (hetero-)definition of our availability of options that is based on the profiles of us as persons probabilistically inferred by algorithms. This is a first hetero-determination of algorithms on our freedom to choose and act as moral agents as it consists of pre-determining the available options to the subjects' choice on the basis of the profile that algorithms can discover and construct of them. Here I have developed two scenarios to show how this epistemological impact does not seem avoidable in the way in which algorithms are currently designed, and I have shown that it can epistemologically affect the *first conditio sine qua non* in different ways (i.e., the epistemological problem can be declined in different ways).

The first scenario is an algorithmic choice-architecture where the profiles of the individuals as inferred from probabilistic assumptions (on which classification and filtering algorithms will base the determination of the options that are available to them) reflect effectively subjects' values, reasons, beliefs, goals (and so forth), broadly speaking, reflect effectively subjects' moral orientation or their moral dispositions, and therefore algorithms tailor continuously the choice-contexts of the users on the basis of this self-learning prediction. This is made possible thanks to the large quantity of data available on us that can allow the construction of highly accurate or personalized profiles on us as persons (let us think about RSs and their capacity to capture the driving elements of our choices) and therefore to reshape our availability of options on the basis of this predictive knowledge on our moral orientation. In this scenario, profiling algorithms steer classifying and filtering algorithms to align the determination of the kind of informational options available to the subjects towards users' well-captured values, reasons, beliefs, or goals. In this sense, the subjects may result as fostered in their choices and actions by the algorithmic action, insofar as the algorithmic hetero-definition of options would result as aligned to their values, beliefs, intentions, broadly speaking, to their moral orientation (or if it is not formed yet to their moral dispositions). Therefore, following this reasoning, this hetero-determination would not pose a real limiting or hindering impact on the availability of alternative options, as they reflect users' probabilistically well-inferred moral orientation. To put it in different terms, the options algorithmically presented to the subjects would be the same the users would have chosen, if they would have the time and resources to perform a similar filtering operation. Although apparently this hetero-determination may increase the capacity of the subject to choose and act accordingly her values, goals and so forth, actually, I have argued how this algorithmic hetero-determination affects this *first conditio* underlying the exercise of moral freedom at its core, specifically, it undermines qualitatively our availability of alternative options, i.e., the alternativity factor, and specifically, the moral heterogeneity of the available options that is peculiar to this condition. I have shown how this alleged choice-enabling algorithmic functioning has been uncovered to undermine our exposure to different points of view, beliefs, values, and relations and creating filter bubbles or informational echo-chambers, i.e., environment characterized by like-minded people, hence with similar beliefs, orientations, and values, ideas of the good, and so on, therefore, categories or groups characterized within by morally similar or homogeneous options (that can produce the effect of a moral echo-chamber). Furthermore, I have clarified how algorithmic choice-architecture tends to narrow our exposure to diverse social interaction, information about people with other culture, values, and ways to do things, that is instead crucial to open the possibility of wondering whether the moral rules, values, and practices we are following are optimal, so to test and eventually change them – a process that is fundamental to embrace genuinely our moral values. In this way, algorithms by shaping our availability of options (and therefore who to get in touch with and what piece of information see) on the profiles probabilistically discovered similar just like ourselves lead to encounter those of exactly the same opinion or value sets as our own and this tends to make us more enclosed and radicalize our previous orientation, instead of critically test, challenge, eventually change, or deeply embrace it.

With this first scenario ad argument, I have maintained how the algorithmic choice-architecture is posing a risk to our moral freedom as freedom of choice and action as moral agents and specifically to develop a genuine moral identity, by undermining the first necessary but not sufficient condition underlying its exercise, i.e., the availability of morally heterogeneous options, that are pre-determined in a way that diminish the social exposure of the agents to diversified values and moral reasons. This lack of heterogeneity of relations, points of views, and orientations is indeed a lack of heterogeneous reasons and diversified ideas on what is good, and so it undermines the possibility to challenge and reasoning our moral orientation, and therefore the possibility of developing genuine reasons and values, and hence genuine moral identity, easily leading to develop self-enclosed reasons, values, and identity. I have specifically claimed that even when the options are shaped on the basis of values and reasons that reflect the moral orientation of the users, as captured via choice-driving elements, the moral echo-chamber or bubbles produced deeply undermine the moral heterogeneity of options that is crucial for the subject to act and choose as a moral agent, therefore, for the subject to become a genuine moral agent and develop a genuine moral identity. Indeed, reducing the heterogeneous expositions of the users' informational environment undermines the possibility for the subjects to develop their own idea of good, their moral values, and own moral ground projects in a genuine way, and therefore in a way that requires the encounter of heterogeneous values, beliefs, reasons (and so forth) crucial to act and choose as genuine moral agents. Indeed, in order to develop morally genuine identity, agents need to be able to critically form, expose, and test their moral orientation, and so those values, reasons (and so forth) they will embrace and then endorse as motives for their choices and actions, on which steering the development of their moral identity. I defined this impact of algorithmic choice-architecture on our availability of morally heterogeneous options as an epistemological impact to subjects' freedom to choose and act as moral agents, as it affects the way in which the subject develops her own idea of good, her moral values, moral reasons, namely, the formation of and critical reasoning on her *moral knowledge*, i.e., of those moral reasons and values she can endorse as motives for her choices and actions and on which she steers the development of her moral identity as a genuine moral identity. In this first scenario, I have argued that even when algorithms are able to capture our "moral orientation", they end to affect epistemologically the individual, specifically affect her moral knowledge, by narrowing individuals' exposure to diversified social-relations, points of views, and therefore, to qualitatively alternative moral reasons and values as a consequences of personalization techniques based on profiling. Here I have shown that one may reply that therefore the fact that very often ML algorithms work on de-individualized assumptions and correlations may constitute a counter-effect to personalization and therefore it may prevent the risk the availability of morally heterogeneous options.

This is specifically the second scenario that I have considered, the scenario in which the profiles algorithmically constructed of individuals do not take in consideration their interests, goals, and moral orientation. I have then argued how ignoring the individuality of a person via de-personalization is not the solution to the above problem of options' personalization that ends in moral echo-chambers. Rather, I have shown that when the predictive models driving profiling algorithms are imperfect and specifically ignore the specific user beyond the profile, the algorithmic reshaping impact on users' options can become even more problematic, above all when as – previously argued – the options pre-determined as available to people are not just informational *stricto sensu*, but are also considerable alternative possibilities such as real chances and social opportunities. We have indeed widely described how everything is interconnected via algorithms and the same profiling algorithms that rule our SNS can inform predictive recidivism algorithms, as well as algorithms that determine who can be denied a credit, a loan, and a housing, who can access to a job, to a subsidized rate of health insurance, or a particular health service. This means that, for example, a person can be subject to an adverse decision, such as being denied credit, due to de-individualized assumptions that are too general to be inexact or biased can see the algorithmic refusal simply in virtue of being similarly profiled to persons who are not credit-worthy – in a way that do not consider her as particular person.

I have argued then how algorithms which make decisions on subjects on the basis of profiles that do not consider them or reflect them as particular persons can affect even worse epistemologically individuals as well, by creating asymmetries both in knowledge and in power. Indeed, when the algorithmic probabilistic assumptions on the basis of which we are profiled and categorized into a certain group are inexact and de-individualize us as persons, not only we experience a limitation of available options by algorithms, inasmuch as they are hetero-determined and chosen in advance by algorithms on the basis of the profile of us probabilistically discovered or constructed, but the available options (and so what I can get access to or not) are based on an algorithmic knowledge of us as profiles that does not reflect us as agents and specific persons with a specific and particular history that cannot be generalized or taken for granted.

This results in an epistemological asymmetry between who I am as a moral agent, and therefore, my values, my attachments, moral ground projects, beliefs, and so forth, and how I have been profiled, known, and described by algorithms, and on the basis of this profile, re-influenced by the re-shaping of my informational environment or choice-context. Moreover, when we are categorized and grouped in arbitrary or in increasingly complex ways we are often unable to predict, understand, and therefore contest the algorithmic decisions made on us (e.g., this can be just about an information that is shown or not to me or instead my label as not creditworthy) and on which we are subject to. In this sense, the epistemological asymmetry is a real asymmetry in terms of power between the algorithmic profiler (and who is behind) and the person profiled.

In this sense, I have defined also this impact (in this second scenario) as an epistemological impact on the *first conditio sine qua non* underlying our moral freedom, insofar as it weakens the epistemological position of the "decisional" subjects, who is made unaware and passive towards her choice-context, as it cannot know and intervene on the options algorithmically pre-determined – options which in turn lead the agents to pre-determined alternative possibilities, opportunities, chances, and courses of actions.

There is a crucial difference between the two scenarios described and their consideration in the light of the impact of profiling algorithms on moral freedom that just its conceptualization in a negative sense has allowed us to underline.

The negative concept of moral freedom means that our moral freedom can be defined as freedom from potential and actual, moral and immoral, interferences to develop genuine moral identity, where by moral limitation I have meant an interference on us that even if reflects my moral orientation is nonetheless moral-freedom constraining as interferences with the possibility to choose according my own reasons and values, and whereby immoral limitation is meant an interference on us that is illegitimate or unjust as does not consider our moral values and reasons (moral orientation), i.e., there is no tracking of our interests, moral goals, and ground project and this results in a de-individualization of the subject that can lead to label and profile (and categorize) him on inexact, wrong, or biased assumptions. According to this definition, I have distinguished the algorithmic interference as delineated in the first scenario, where the algorithmic reshaping function of our options is based on profiles of us that can track our interests, goals, and moral orientation, as always freedom constraining (as argued above) but definable as moral, inasmuch as take into consideration us as person. Moral does not mean that is good per se, but that takes into account us as moral agents (i.e., our moral dimension), and that is not morally unjust or illegitimate, as not based on the wrong assumptions on them and so do not subject users to unfair or illegitimate outcomes. Nevertheless, this interference always remains a moral freedom-constraining, that needs a moral justification and can be regulated (and fixed) on the basis of what justification can be provided. The distinction between moral and immoral interferences does not change the fact that algorithms (in both the two scenarios) constrain our moral freedom to choose and act as moral agents, by binding us to pre-determined options. Nevertheless, the distinction can be helpful when we are called to distinguish what interferences need to be just regulated or fixes and those that instead must be eliminated as illegitimate or unjust. In the third section of the second chapter, I have shown how the impact of algorithmic choice-architectures can go beyond the epistemological level of persons (as profiled) and affects also their moral autonomy, as relational self-determination. I have argued how the relational dimension that is part of our possibility to be the author of genuine moral identity is undermined by the pre-determination of algorithms of the availability of options that very often reduce users' socio-relational exposure and thus the moral heterogeneity of values and reasons that are crucial for the agents to test the values embraced, therefore deeply limiting their autonomy and power to choose and act as genuine moral agent to a very narrowed – in quality and quantity – and so constrained availability of options algorithmically predetermined on the basis of patterns discovered and profiles constructed. Therefore, the predictive knowledge developed on the agent can become a way to bind her potential choice and action to a set of options algorithmically pre-determined. Furthermore, I have also showed how this algorithmic impact can also affect at the core of people's autonomy their ability to reflectively endorse those options they embrace as motives of their choices and actions. Reflective endorsement is our last call for moral freedom: in the reflective endorsement we exercise on the options available we can find the distinctive trait of our authorship over our choices, actions, and identity. In the reflective endorsement we find the way in which we can determine ourselves given a context of pre-determined options. Indeed, by exercising reflective endorsement, and so by endorsing certain options embedding certain moral reasons and values, those we embrace, by approving them as motives for our choices and actions, we develop our *ought to*, namely, the way in which we make those moral reasons and values not just the motives but the moral rules for our behavior, i.e., we make them normative for our conduct. This key-trait is indeed of crucial importance as by exercising our reflective endorsement we develop the way in which we respond to reality, conveyed by our mediated perceptions and emotions, by taking a moral stand and by doing so developing our moral posture: develop our moral identity as genuine persons. Indeed, by endorsing options as specific values and reasons as motives for our behavior we actualize our moral disposition towards a certain direction, rather than another, so exercising our freedom of become certain moral agents, rather than others, to choose and act with a certain moral posture, rather than another: in sum, to develop our moral identities in a genuine way as moral agents that are authors of their choices and actions. I have argued that even if who can pre-determine the 'availability' of our options can bind our choices (i.e., we cannot choose those options that result as unavailable) to certain options rather than others, and therefore can exercise a constrain our freedom of choice and action, this constrain is soft to the extent we as moral agents have always the power to decide to act against their informational options (as well as against our preferences and needs), or choose not to choose. This because our options do not necessarily determine us or constrain us to choose. This is because we can always 'yes or no' to a certain option (reason, value, event, desire, and so forth) and this approval lies exactly in the exercise of reflective endorsement. Therefore, I had to explore the algorithmic impact on our endorsement to understand whether they can effectively undermine the second condition of our moral freedom and so effectively jeopardize it, by making the risk of a novel threat to moral freedom real. I focused specifically on algorithmic RS, as one of the main techniques nowadays in use: indeed, I have shown that algorithmic recommendations are not nudges, but real pushes, that can create a hard constraint on our freedom of choice and action as moral agents. RSs create indeed a more pervasive action on individual: they do not just filter or re-order options available, but can use a few of them to specifically target individuals.

To understand the risk of algorithmic predeterminism, I have highlighted when the phenomenon I call 'endorsement suspension' may happen, and this is specifically when the information used to target individuals is highly sensitive (e.g., from information about individuals' physical or psychological status to personal vulnerabilities or weaknesses). Due to the RSs use in morally loaded contexts, the possibility for algorithms to capture highly personal and sensitive aspects of the individuals is very likely and this entails that is also very likely for RSs capture the choice-driving elements (also via cross-inferences between groups and within the same group) which connote users and this kind of information is highly valuable to push their behavior to a certain direction rather than another (to a choice from a purchase to a political vote, for example). In this sense, the options recommended (or better pushed) by RSs are extremely value-laden, as able to trigger the choice-driving elements caught. In turn, since RSs' use today ranges from contexts, such as health care, lifestyle, insurance, and the labor market, that are morally loaded, RSs' outputs – what can trigger a certain recommendation – produce consequences morally relevant for the individuals, where a choice rather than another can be life-changing. The options algorithmically recommended are pre-determined on the basis of the refined predictive knowledge on the users, and so they have a specific triggering potential: they can trigger physical or psychological weaknesses, as well as vulnerabilities, pathologies such as depression or anxiety, or evoking fears and trauma. This means that RSs have the power to raise emotional instinctive responses, and specifically, trigger primary emotions, those we have in common with very young child and animals (are instinctive and innate). This means that the options chosen by RSs to target user on the basis of the captured personal sensitive traits and choice-driving elements can work as *triggers*: such options can trigger users' emotional and instinctive behavior-response (fear, extreme joy, anger and so forth) in a way that can suspend users' exercise of reflective endorsement, by leading the option emotionally loaded to determine their choices and actions at users' place. I provided examples on the way in which the predictive knowledge of ML algorithms can lead to discover users' vulnerabilities, traumatic experiences, weaknesses (and so forth) by associating data capturable from diverse applications (such as GPS, SNS, healthy-apps…) and the users' history (and broadly on other users' interactions via collective-based filtering) and since algorithms are not capable to qualitatively distinguish the options recommend (the moral weight of a certain option embedded in an informational content), RSs can purposely or accidentally target users' vulnerabilities or choice-driving elements with information that can be sensitive or emotionally loaded so much to create not just a soft constraint on individuals' freedom of choice and action, as that created by pre-determining the options amongst which users can choose, but a hard constraint on it, by suspending the key-endorsement subjects can give to a certain option rather than to another by approving it as a motive for her choice and action. In these cases, the option recommended can become a real push that affects individuals' autonomy in-depth, by suspending reflective endorsement and transforming the informational *option* pushed from being a *motive* of people's choices and actions (an option they can approve as a motive for their choices by reflectively endorsing it) to be the main *cause* of users' choices and actions. These relentless RSs' pushes in triggering options (high-sensitive contents) can affect users' autonomy and their endorsement by transforming the main option pushed – algorithmically chosen for them – from being a motive (that users can endorse) to be the cause (strictly determining) of their choice and their behavior.

As a consequence, the RSs' recommendation of such options, instead of epistemologically informing agents' choices and action (informing them), ends to decide or choose at the users' place, in other words, to determine them. In this sense, algorithms are not just architects of the contexts of our choices, by informing our choices via the reshaping of our availability of options, but become the architects of our choices and actions themselves, by not just informing our choices, rather pre-determining us and them. The choices as determined can be life-changing, as in the case mentioned, but also when they are not so morally loaded, they can open certain courses of actions while declining others, in a way in which firstly do not express the authorship of the agents and above all do not respect their moral reasons and values, whose approval has been suspended – this means that I do not take a moral stand in response to a certain option/event or information, insofar as I have been determined, though my choice has always a consequence on how I form my identity and therefore is binding my future moral development. In this sense, I have argued that our moral autonomy would become not just influenceable (via the predetermination of the relational dimension), but also predeterminable (via the suspension of the endorsement), and so, also the second necessary condition underlying our moral freedom, might result as affectable by algorithmic choice-architecture. It follows that algorithmic choice-architecture cannot just affect (and interfere) on how persons develop their moral knowledge (epistemological level), but it might affect (and interfere) on how we form our ought to, our moral posture, that is, the corroboration of our moral dispositions towards our moral identity via the way in which we respond (or not respond) by taking a moral stand to our reality, by endorsing the values and moral reasons we embrace as motives of our choices and actions, which over time and in turn give form to our moral identity.

This algorithmic determination of our choice via options' targeting is based on a predictive knowledge (of my choice-driving or behavior-triggering elements) developed as probabilistically discovered by algorithms in order to meet pre-determined or pre-set goals. Following this reasoning, the algorithmic choice-architecture's potential action sounds capable to give rise to a novel threat to moral freedom: to an unprecedented form of predeterminism generated by the potential of ML algorithms to discover predictive knowledge on individuals – on the basis to the capacity to see and discover what is invisible or impossible to the human eye, as lies in the capacity to scale and compare huge amounts of data and correlations – whose controversial application can generate intentional or accidental, moral and unmoral, interferences or constraints on people's capacity to form their own ideas of the good, their own moral ground projects, as well as form and assess their own values and moral reasons, to the point of undermining their freedom to choose, act, and so become genuine moral agents.

This is exactly what I have defined as the rise of the threat of the algorithmic predeterminism. The analysis of the first impact of algorithms on the first condition necessary to our moral freedom has shown that as our availability of options can be reshaped and more pre-determined algorithmically, the potential risk of an algorithmic pre-determinism according to which our choices are pre-determined algorithmically does not seem so far. The analysis of the impact of algorithms on the second necessary condition on our moral freedom, our moral autonomy, have allowed me to claim that more than a risk the rise of algorithmic predeterminism driven by the huge predictive capacity of ML algorithms is becoming a reality.

The third chapter of the dissertation has been devoted to the understanding on how to secure our moral freedom as our freedom of choice and action as moral agents in our evermore algorithmically-governed societies.

In the first section, I have resumed the steps of the argumentation and I have deepened the threat posed by algorithmic pre-determinism to our moral freedom, by showing what is at the root of the algorithmic predeterminism and – by providing some examples – the extent of its impact. I have highlighted how predeterminism is a very old concept, and has assumed diverse declinations in the course of history (natural, biological, psychological, just to mention the main theories), although its underlying idea is that events (including human actions) have been already decided or already known, i.e., they are already pre-determined, and therefore there is no space for freedom of choices and actions, as the chain of events is pre-established and human action has no power to interfere and intervene on it. By algorithmic pre-determinism, I have argued the rise of an unprecedented form of predeterminism generated by the potential of ML algorithms to discover predictive knowledge on individuals, whose controversial application can generate intentional or accidental, moral and unmoral, interferences on people's capacity to form their own ideas of good and their own ground projects to the point of undermining their freedom to choose, act, and so become genuine moral agents.

This further definition of algorithmic predeterminism has been helpful to elaborate a novel conceptual lens to frame and then introduce this unexplored algorithmic threat inside the current ethical and legal debate on algorithms. Indeed, in the second section, I have drawn insights from theories in the privacy debate concerning the protection of freedom, and specifically, the theories of informational privacy and intellectual privacy, in order to elaborate and add a new specific conceptual lens, what I defined *moral privacy*, that I have argued it constitutes the third level of a three-layers privacy framework for a comprehensive protection of our freedom – that with moral privacy's lens specifically includes the protection of moral freedom. Particularly, I have underlined a third-level zone of legal protection of our moral freedom and I have elaborated three moral privacy's prescriptions that should spur and steer the development of novel techniques in order to operationalize by design the ethical-normative value of moral freedom. I have specifically argued how and why the right to informational privacy defines a zone of protection that is important for our moral freedom, as it constitutes a *first basic ground* for the protection of our overall freedom, namely, for protection of freedom in our deeply interconnected and more and more algorithmically ruled informational societies. Informational privacy detects the protection of a personal space to develop unique identity, a space that is free from interferences on 'who we are' and 'who we are becoming into', thus, for the protection of the informational identity of both individuals and collectivities/groups. Then, I have argued how and why the right to intellectual privacy is also valuable for moral freedom, as it should protect us from a re-definition of our choice-environment according to a personal identity that is private in thoughts, desires, needs (and so forth), an identity we do not are sure we want to publicly share and that does not overlap with our moral identity. Indeed, I have argued the protection of our moral identity both beyond the conception of informational and intellectual identity, as not all my information defines us as moral agents, but only that information – or the informational options – we have decided to endorse as motives for our choices and actions. The development of my moral identity as the formation of my ought to requires my endorsement amongst the available options (amongst which there are also desires, needs, short-terms goals, fantasies and so on: everything can connote my identity both as informational identity – so my general personal information – and as intellectual identity, i.e., the information that I do not have chosen or endorsed yet and that perhaps I will never endorse into my choices). So, the act of endorsement is what defines what information connotes my moral identity (at least actively). The protection of moral identity is anyway informational – is informational, as everything is conceivable by information in our informational societies, thus, it requires informational privacy as a first layer of protection. Though, informational privacy is not enough, it is a starting informational ground in which we need to discern what really pertains to our choosing and acting as moral agents and what does not, as above pointed out. At the same time, the protection of moral identity requires the protection of intellectual identity, as what we need on choosing on, as options on which we need to deliberate, but by which we do not want to be necessarily defined before our key approval or endorsement. Intellectual privacy is a part of our personal identity, but only what we decide to endorse defines us as moral agents and so our moral identity, by forming over time our ought to, our obligation to certain values and moral reasons, rather than to other options (e.g., desires, needs, and so forth). At the same time, the protection of our intellectual identity is not enough, as it does not secure at a minimum threshold our possibility to choose and act as genuine moral agents; intellectual privacy just secure the protection from interferences on the formation of new ideas, thoughts, the protection of our intellectual activities as a space of confidentiality we deserve on what we do not want to share publicly.

I have so argued how the protection of the development of our moral identity is the protection of the development of our ought to, namely, what corroborates our moral dispositions into our moral posture, into our moral identity, the development of the obligation we form to the values and moral reasons we decide to embrace, by endorsing them among a plurality of morally heterogeneous options: in other words, is the protection of our moral freedom, our freedom of choice and actions as moral agents via which we develop our moral identity. I have so introduced a further lens, *moral privacy*, as what defines a *zone of protection* for our moral freedom, hence for its underlying *conditiones sine qua non*: a. the moral heterogeneity of available options from a narrowing impact poseable by algorithms via information (as moral heterogeneity of our options can secure at a minimum threshold our possibility to choose and act alternatively as moral agent, therefore, to develop our own idea of the good, values, and reasons and so to test the values and moral reasons we embrace and decide to genuinely endorse), and b. for our moral autonomy, that is, our capacity to endorse as motives the information or option (values, moral reasons, and ground projects) into our choices and actions (as the endorsement can secure the authorship of our choices and actions from external constraints).

I have so claimed that moral privacy is therefore a zone of protection of our freedom of choice and action as genuine moral agents that underpins the ideal of freedom from algorithmic predeterminism, which in turn expresses in moral and immoral interferences on individuals' possibility to develop genuine moral identity. Moral privacy detects the zone of protection from interference via information that can alter our identity, but more specifically, from everything can interfere with the genuine development of our moral identity, and therefore, from anything can interfere with our possibility to exercise our reflective endorsement so much to suspend it and so determining at our place our choices and actions. As well as, it detects a zone of protection from interferences that can reshape our informational environment or choice-context by undermining (or narrowing) the moral exposure we need for the development of genuine moral identity by binding our choices to options to meet pre-determined heteronomous goals. I have elaborated the zone of protection of our moral freedom detected by the lens of moral privacy as declined in the respect of what I have developed as three main moral privacy's axioms that are prescriptions that if technically operationalized can safeguard the two necessary conditions underlying our moral freedom, and so secure its exercise at a minimum threshold, and specifically they protect our freedom of choice and action as moral agents by proposing how to address moral and immoral interferences raised by algorithms on our possibility to develop genuine moral identity and so to become genuine moral agents.

In the last section, I have defined the specific institutional and technological agents called into action in the application of the conceptual moral privacy axioms to protect our moral freedom in our contemporary algorithmic societies; the ones I defined the champions of moral freedom, as the agents entitled to safeguard and boost the value of our moral freedom in our contemporary societies. I have clarified that algorithms' designers and institutional decision-makers have the duty (and which kind of specific duties) to fulfill the right to moral privacy according to the criteria pointed out, respectively, the former mainly in the technical design of algorithms-based ICTs, while the latter in the evaluation process concerning the phases of algorithms-based ICTs' adoption, regulation, and supervision. I have also underlined that they are not the sole agents that are called to the protection of our moral freedom, as the issues of moral freedom and algorithmic predeterminism are complex and multidimensional, we need to further reason and design an approach of collective or shared responsibility capable to protect the multidimensionality of our freedom according to different levels that in turn should engage different actors.

Nevertheless, the technological and institutional agents are those actors that today have the crucial role and the duty to design, deploy, and supervise the use of algorithmic ICTs in a way that ought to be aligned with the ethical-normative value of our moral freedom, for example, via the operationalization of moral privacy's axioms, as they draw a path to protect individuals' moral freedom in the algorithmic era, and therefore, a path to protect our freedom to become genuine moral agents and develop genuine moral identity from existing and novel forms of social and technological interference: a path to preserve ourselves as agents that are free to choose and act according to the values and reasons we believe in, the moral society we want to design and build, and therefore, the moral progress we want to see and contribute to in our world.

BIBLIOGRAPHY

Adomavicius, G., Sankaranarayanan, R., Sen, S., & Tuzhilin, A. (2005). "Incorporating contextual information in recommender systems using a multidimensional approach". *ACM Transactions on Information Systems (TOIS)*, 23(1), 103–145.

Aggarwal, N. (2020). "The Norms of Algorithmic Credit Scoring". *Cambridge Law Journal* (*forthcoming*).

Albritton, A. (1985). "Freedom of Will and Freedom of Action". *Proceedings and Addresses of the American Philosophical Association*, 59 (2): 239-251.

Ananny, M. (2016). "Toward an ethics of algorithms convening, observation, probability, and timeliness". *Science, Technology*, *& Human Values*, 41(1): 93-117.

Ananny, M., & Crawford, K. (2016). "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability". *New Media & Society*, 17 (1): 973-989.

Anderson, M., & Anderson, S.L. (2007). "Machine ethics: Creating an ethical intelligent agent". *AI Magazine*, 28(4).

Anderson, E. (1999). What is the Point of Equality?. *Ethics*, *109* (2), 289–337.

Anderson, M., & Anderson, S.L. (2014). "Toward ethical intelligent autonomous healthcare agents: A case-supported principle-based behavior paradigm". *AISB 2014 - 50th Annual Convention of the AISB*.

Aneesh, A. (2006). *Virtual Migration*. Durham, NC (USA): Duke University Press.

Aneesh, A. (2008). “Global labor: algocratic modes of organization”. *Sociological Theory*, 27 (4): 347-370.

Angwin, J., Larson, J., Mattu, S. & Lauren, K. (2016, May 23). *Machine Bias*. Retrieved March 10, 2021, from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Applin, S.A., & Fischer, M.D. (2015). “New technologies and mixed-use convergence: How humans and algorithms are adapting to each other”. *2015 IEEE international symposium on technology and society (ISTAS)*, Dublin (Ireland): 1-6.

Arneson, R. (1991). “Autonomy and Preference Formation,” in Jules Coleman and Allen Buchanan (eds.), *In Harm’s Way: Essays in Honor of Joel Feinberg*, Cambridge: Cambridge University Press, pp. 42–73.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). “Privacy and human behavior in the age of information”. *Science*, 347, 509–514.

Bakardjieva, M., & Gaden, G. (2012). “Web 2.0 technologies of the self”. *Philosophy & Technology*, 25: 399-413.

Barnet, B.A. (2009). “Idiomedia: The rise of personalized, aggregated content”. *Continuum*, 23(1): 93-99.

Barocas, S. (2014). “Data mining and the discourse on discrimination”. *Proceedings of the Data Ethics Workshop, Conference on Knowledge Discovery*. Retrieved from: https:/semanticscholar.org/abbb/235cf3b1633784252b44fa.pdf

Barocas, S. (2014), Data mining and the discourse on discrimination. *Proceedings of the Data Ethics Workshop, Conference on Knowledge Discovery and Data Mining (KDD)*. Retrieved March 10, 2021.

Barocas, S., & Selbst, A.D. (2015). "Big data's disparate impact". *SSRN Scholarly Paper*, Rochester, NY (USA): Social Science Research Network.

Beer, D. (2017). "The Social Power of Algorithm". *Information, Communication & Society*, 20 (1): 1–13.

Benson, P. (2005). "Feminist Intuitions and the Normative Substance of Autonomy," in J.S. Taylor (ed.), pp. 124–42.

Benson, P. (1994). "Autonomy and Self-Worth," *Journal of Philosophy*, 91(12): 650–668.

Benjamin, R. (2019). *Race after Technology: Abolitionist Tools for the New Jim Code*. Medford, MA: Polity.

Berlin, I. (1969). *Two Concepts of Freedom*. Oxford (UK): Oxford University Press.

Berk, R., Heidari, H., Jabbari, S., Kearns, M. & Roth, A. (2018). Fairness in Criminal Justice Risk Assessments: The State of the Art. *Sociological Methods & Research*, 004912411878253. https://doi.org/10.1177/0049124118782533.

Binns, R. (2018). "Fairness in Machine Learning: Lessons from Political Philosophy". *Journal of Machine Learning Research*, 81: 1-11.

Blume, P. (2002). *Protection of Informational Privacy*, Djøf Forlag, Copenhagen.

Boyd, D., & Crawford, K. (2012). “Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon”. *Information, Communication & Society*, 15: 662-679.

Bostrom, N., & Ord, T. (2006). “The Reversal Test: Eliminating Status Quo Bias in Applied Ethics”. *Ethics*, 116 (4): 656-679.

Bozdag, E. (2013). “Bias in algorithmic filtering and personalization”. *Ethics and Information Technology*, 15(3): 209-227.

Bradshaw, S., & Howard, P. (2019). “The Global Disinformation Order: 2019 Global Inventory of Organized Social Media Manipulation”. *Project on Computational Propaganda.* Oxford (UK).

Brey, P., & Soraker, J.H. (2009). “Philosophy of Computing and Information Technology”. *Elsevier*.

Budd, B.S. Miller, & Manning, M.L., *et alia* (2020) “Digital technologies in the public-health response to COVID-19”. *Nature Medicine* 26: 1183-1192.

Buolamwini, J., & Gebru , T. (2018). Gender Shades: Intersectional Accuracy Disparities. *Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR*, *81*, 77-91. Retrieved 11 March, 2021.

Burr, C., Cristianini, N., & Ladyman, J. (2018). An Analysis of the Interaction Between Intelligent Software Agents and Human Users. *Minds and Machines*, *28*(4), 735–774.

Burrell, J. (2016). “How the machine ‘thinks:’ Understanding opacity in machine learning algorithms”. *Big Data & Society*, 3(1): 1-12.

Calo, M.R. (2014). “Digital Market Manipulation”. *The George Washington Law Review*, 82(4).

Calhoun, C.J., (2002). *Dictionary of the social sciences*. Oxford University Press, New York (NY, USA).

Carter, I. (1995). “The Independent Value of Freedom”. *Ethics*, 105 (4): 819-845.

Carter, I. (1999). *A Measure of Freedom*. Oxford (UK): OUP.

Carter, I. (2011). Respect and the Basis of Equality. *Ethics*, *121*(3), 538-571.

Chaslot, G. (2018, February 1). “How Algorithms Can Learn to Discredit the Media”. *Medium*.

Chiusi, F., Fischer, S., Kayser-Bril, N., & Spielkamp, M. (2020), “Automating Society Report 2020”, *AW AlgorithmWatch*, Berlin (Germany).

Citron, D.K., & Gray, D. (2013). “Addressing the harm of total surveillance: A reply to Professor Neil Richards”. *Harvard Law Review Forum*, 126 (262).

Citron, D.K., & Pasquale, F. (2014). “The scored society: due process for automated predictions”. *Washington Law Review*, 89 (1): 1-34.

Coll, S. (2013). Consumption as biopower: Governing bodies with loyalty cards. *Journal of Consumer Culture*, *13*(3), 201–220.

Committee of Ministers - European Union (13th February 2019). *Declaration on the Manipulative Capabilities of Algorithmic Processes*, Bruxelles (Belgium).

Corbett-Davies, S. & Goel, S. (2018). The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *ArXiv:1808.00023 [Cs]*, Retrieved March 11, 2021.

Crawford, K. (2013). "The hidden biases of Big Data". *Harvard Business Review.*

Danaher, J. (2018). "Moral Enhancement and Moral Freedom: A Critique of the Little Alex Problem", *Royal Institute of Philosophy Supplement*, 83 (10): 233-250.

Danks, D. & London, A.J. (2017). Algorithmic Bias in Autonomous Systems. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization*, 4691-4697.

Dastin, J. (2018, October 11). *Amazon scraps secret AI recruiting tool that showed bias against women.* Reuters. Retrieved March 7, 2021.

De Vries, K. (2010). "Identity, profiling algorithms and a world of ambient intelligence". *Ethics and Information Technology*, 12(1): 71-85.

Deville, J. (May 20, 2013). *Leaky Data: How Wonga Makes Lending Decisions*. Charisma: Consumer Market Studies. Retrieved March 11, 2021, from http://www.charisma-network.net/finance/leaky-data-how-wonga-makes-lending-decisions.

Diakopoulos, N. (2015). "Algorithmic accountability: Journalistic investigation of computational power structures". *Digital Journalism*, 3(3): 398-415.

De Caro, M. (2004). *Il libero arbitrio. Una Introduzione*, Roma-Bari (Italy): Laterza Editori.

Domingos, P. (2012). A few useful things to know about machine learning. Communications of the ACM, 55(10): 78–87

Domingos, P. (2015). *The master algorithm: how the quest for the ultimate learning machine will remake our* world. New York, NY: Basic Books.

Dworkin, G. (1998). *The theory and practice of autonomy*. Cambridge: CUP.

Dworkin, R. (2000). *Sovereign Virtue: The theory and Practice of Equality*. Cambridge, MA: Harvard University Press.

Ekstrom, L. (1993). "A Coherence Theory of Autonomy," *Philosophy and Phenomenological Research*, 53: 599-616.

Ekman, P. (2003). *Emotions Inside Out: 130 Years after Darwin's The Expression of the Emotions in Man and Animals.* New York: New York Academy of Sciences 2003.

Eppler, M. J., & Mengis, J. (2004). "The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines". *The Information Society*, 20(5), 325–344.

Eubanks, V. (2018). *Automating Inequality*. New York, NY (USA): St Martin's Publishing.

Edwards, J. 1754 [1957]. *Freedom of Will*, ed. Paul Ramsey, New Haven: Yale University Press.

Ferguson, A. (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law*. New York, NY (USA): NYU Press.

Floridi, L. (2005). "The Ontological Interpretation of Informational Privacy. Ethics and Information Technology". 7(185-200).

Floridi, L. (2006). “Four challenges for a theory of informational privacy”. *Ethics and Information Technology, 8*(3), 109-119.

Floridi, L. (2016). “On human dignity as a foundation for the right to privacy”. *Philosophy and Technology, 29*(4), 307-312.

Floridi, L. (2008). “The method of levels of abstraction”. *Minds and Machines*, 18(3): 303-329.

Floridi, L. (2010). *Information – a very short introduction*. Oxford: Oxford University Press.

Floridi, L. (2011). “The informational nature of personal identity”. *Minds and* Machines, 21(4): 549-566.

Floridi, L. (2012). “Big data and their epistemological challenge”. *Philosophy & Technology*, 25(4): 435-437.

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford (UK): OUP.

Floridi, L. (2013). *The ethics of information*. Oxford: Oxford University Press.

Floridi, L. & Sanders, J.W. (2004). “On the morality of artificial agents”. *Minds and Machines*, 14 (3).

Floridi, L. (2017). Group privacy: a defence and an interpretation. In Taylor, L., Floridi, L. and van der Sloot, B (eds.), *Group privacy: new challenges of data technologies*, Cham: Springer, pp. 83-100.

Floridi, L., & Taddeo, M. (2018). “How AI Can Be a Force for Good”. *Science*, 361 (6404): 751-753.

Floridi, L., & Cowls, J. (2019). “A Unified Framework of Five Principles for AI in Society”. *Harvard Data Science Review* 1(1).

Forst, R. (2014). Two Pictures of Justice. In *Justice, Democracy and the Right to Justification. Rainer Forst in Dialogue*, London: Bloomsbury.

Floridi, L. (2017). Group privacy: a defence and an interpretation. In Taylor, L., Floridi, L. and van der Sloot, B (eds.), *Group privacy: new challenges of data technologies*, Cham: Springer. 83-100.

Frankfurt, H. G. (1969), “Alternate Possibilities and Moral Responsibility”, *The Journal of Philosophy*, 66(23): 829–839. Reprinted in Fischer 1986, pp. 143–52; in Frankfurt 1988, pp. 1–10; and in Widerker and McKenna 2003: 17–25.

Frankfurt, H. G. (1983), “What We Are Morally Responsible For”, in L.S. Cauman et al. (eds.), *How Many Questions? Essays in Honor of Sidney Morgenbesser*, Indianapolis, IN: Hackett Publishing Company:321–335. Reprinted in Frankfurt 1988, pp. 95–103 and in Fischer and Ravizza 1993, pp. 286–295.

Frankfurt, H. (1971). “Freedom of the will and the concept of a person”. *Journal of Philosophy*, 68 (1971): 5-20.

Frankfurt, H. (1994) “Autonomy, Necessity, and Love,” in *Vernunftbegriffe in der Moderne: Stuttgarter Hegel-Kongress 1993*, eds. H.F. Fulda and R.P. Horstmann, Stuttgart: Klett-Cotta.

Friedman, B., & Nissenbaum, H. (1996). “Bias in computer systems”. *ACM Transactions on Information Systems (TOIS)*, 14(3): 330-347.

Frischmann, B., & Selinger, E. (2018). *Re-Engineering Humanity*. Cambridge (UK): Cambridge University Press.

Fule, P., & Roddick, J.F. (2004). “Detecting privacy and ethical sensitivity in data mining results”. *Proceedings of the 27th Australasian conference on computer science*, 26. Australian Computer Society, Dunedin (New Zealand):159-166.

Gabbert, M.R. (1927). “Moral Freedom”, *The Journal of Philosophy*, 24 (17): 464-472.

Gal, M. (2018). “Algorithmic Challenges to Autonomous Choice”. *Michigan Journal of Law and Technology*, 25: 59-104.

Garcia-Molina, H., Koutrika, G., & Parameswaran, A. (2011). Information seeking. Communications of the ACM, 54(11), 121.

Gräf, E. (2017). “When Automated Profiling Threatens Freedom: a Neo-Republican Account”. *European Data Protection Law* Journal, 4:1-11.

Green, B., & Yiling, C. (2019). “Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments”. *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* ’19*, Atlanta, GA: ACM Press: 90-99.

Gutwirth, S., and Hildebrandt, M. (2010). *Data Protection in a Profiled World.* Erasmus University Rotterdam; Springer Netherlands.

Harris, J. (2011). “Moral Enhancement and Freedom”. *Bioethics* 25: 102-111.

Hartmann, N. (1932). *Moral Freedom*. The MacMillan Company. New York (NY, USA).

Hauer, T. (2019). “Society Caught in a Labyrinth of Algorithms: Disputes, Promises, and Limitations of the New Order of Things”. *Society*, 56 (3): 222-230.

Hilbert, M. (2012). “Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making, Psychological Bulletin”, 138(2), 2021: 211–237.

Hildebrandt, M. (2008). “Defining profiling: A new type of knowledge?”. *Profiling the European Citizen* (edited by Hildebrandt M and Gutwirth S). The Netherlands: Springer: 17-45.

Hildebrandt, M., & Koops, B-J. (2010). “The challenges of ambient law and legal protection in the profiling era”. *The Modern Law Review*, 73(3): 428-460.

Hill, R.K. (2015). “What an algorithm is”. *Philosophy & Technology*, 29(1): 35-59.

Hilty, L. M. (2015). “Ethical issues in ubiquitous computing – three technology assessment studies revisited”. In: Kinder-Kurlanda, Katharina; Ehrwein Nihan, Céline. Ubiquitous Computing in the Workplace. Cham: Springer, 45-60.

Hinman, L.M. (2005). Esse est indicato in Google: Ethical and Political Issues in Search Engines. *International Review of Information Ethics*, *3*. Retrieved March 11, 2021.

Hinman, L.M. (2008). Searching Ethics: The Role of Search Engines in the Construction and Distribution of Knowledge. In: Spink A., Zimmer M. (eds.). *Web Search. Information Science and Knowledge Management*, Springer, *14*.

Hobbes, T. (1654 [1999]). “Of Liberty and Necessity” in *Hobbes and Bramhall on Liberty and Necessity* (edited by Chappell V), Cambridge (UK): Cambridge University Press: 15-42.

Holton. R. (2009). *Willing, Wanting, Waiting*, New York: Oxford University Press.

Hoven, J. V., & Rooksby, E. (2008). *Distributive justice and the value of information: A (broadly) Rawlsian approach*. England: Cambridge University Press.

Hoye, J.M., & Monaghan, J. (2018). “Surveillance, Freedom and the Republic”. *European Journal of Political Theory* 17 (3): 343-363.

Hu, Margaret (2017). Algorithmic Jim Crow. *Fordham Law Review*. Retrieved March 10, 2021.

Introna, L.D. (2016). “Algorithms, governance and governmentality: On governing academic writing”. *Science, Technology, & Human Values*, 41: 1-33.

Introna, L.D., & Nissenbaum, H. (2000). “Shaping the Web: Why the politics of search engines matters”. *The Information Society*, 16(3): 169-185.

Jobin, A., Ienca, M. & Vayena, E. (2019). Artificial intelligence: the global landscape of ethics guidelines. *Nat Mach Intell*, *1*, 389–399.

Kamishima, T., Akaho, S., Asoh, H., & Sakuma, J. (2012). “Considerations on fairness-aware data mining”. *IEEE 12th International Conference on Data Mining Workshops*, Brussels (Belgium): 378-385.

Kahneman, D. (2011). *Thinking, Fast and* Slow. Farrar, Straus & Giroux.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47, 263–91.

Kahneman D., & Tversky, A. (2000), eds., *Choices, Values and Frames*. Cambridge University Press.

Kant, I. (1998 [1781]). *Critique of Pure Reason* (trad. and ed. by Guyer, P. and Wood, A.W.). Cambridge (UK): Cambridge University Press.

Kant, I. (1956 [1788]). *Critique of Practical Reason* (trad. by White Beck, L.), New York, NY (USA): Macmillan Library of Liberal Arts.

Killmister, J. (2017). *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance*. London (UK): Routledge.

Kim, P.T. (2017). Data-Driven Discrimination at Work. *58 Wm. & Mary L. Rev*, *857* (3). Retrieved March 11, 2021.

Kitchin, R. (2016). "Thinking critically about and researching algorithms". *Information, Communication & Society* 20 (1): 14–29.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2017). Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*.

Korsgaard, C.M. (1996). *The Sources of Normativity*. Cambridge (UK): Cambridge University Press.

Korsgaard, C. M. (2014). "The Normative Constitution of Agency," in Manuel Vargas and Gideon Yaffe (eds.), *Rational and Social Agency: The Philosophy of Michael Bratman*, New York: Oxford University Press, pp. 190–214.

Kymlicka, W. (1989). *Liberalism, Community and Culture*, Oxford: Clarendon.

Labati, R., Donida, A.G., Muñoz, E., Piuri, V. (2016). "Biometric Recognition in Automated Border Control: A Survey". *ACM Computing Surveys*, 49 (2): 1-39.

Lacy, S. (14th November 2017). "Uber Executive Said the Company Would Spend 'A Million Dollars' to Shut Me Up", *Times*, 14th November 2017.

Laidlaw, E. B. (2008). Private Power, Public Interest: An Examination of Search Engine Accountability. *International Journal of Law and Information Technology*, 17*(1)*, 113–145.

Lee, M. K. (2018). Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management. *Big Data & Society*, *5* (1).

Lee, M.S.A., & Floridi, L. (2020). "Algorithmic Fairness in Mortgage Lending: From Absolute Conditions to Relational Trade-Offs". *SSRN Electronic Journal*.

Leese, M. (2014). "The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards". *The European Union*. *Security Dialogue*, 45(5): 494-511.

Lewis, D. (8 October 2019). "Social Credit Case Study: City Citizen Scores in Xiamen and Fuzhou". *Medium: Berkman Klein Center Collection*.

List, C., & Valentini, L. (2016). "Freedom as Independence". *Ethics* 126 (4): 1043-1074.

Lobosco, K. (2013, August 27). *Facebook friends could change your credit score*. CNN Business. Retrieved March 11, 2021.

McMahan, J. (2000), "Moral Intuition", in *Blackwell Guide to Ethical Theory*, H. LaFollette (ed.), Oxford: Blackwell, chap. 5.

Macnish, K. (2012). "Unblinking eyes: The ethics of automating surveillance". *Ethics and Information Technology*, 14(2): 151-167.

Malhotra, C., Kotwal, V., & Dalal, S. (2018). “ETHICAL FRAMEWORK FOR MACHINE LEARNING”. In *2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K)*: 1–8. Santa Fe: IEEE.

Matthias, A. (2004). “The responsibility gap: Ascribing responsibility for the actions of learning automata”. *Ethics and Information Technology*, 6(3): 175-183.

Michael, B (2005). “Planning Agency, Autonomous Agency,” in *Personal Autonomy*, ed. James Stacey Taylor, New York: Cambridge University Press.

Mitchell, T. (1997). *Machine Learning*, McGraw-Hill: Singapore 1997.

Milano, S., Taddeo, M., & Floridi, L. (2020). “Recommender Systems and Their Ethical Challenges”. *AI & SOCIETY*.

Mackenzie, C., & Stoljar, N. (eds.), (2000a). *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, New York: Oxford University Press.

Mackenzie, C., & Stoljar, N. (eds.) (2000b). “Introduction: Autonomy Refigured,” in Mackenzie and Stoljar (eds.), pp. 3-31.

Myrton Frye, A. (1931). “Moral Freedom and Power”. The Journal of Philosophy, 28 (10): 253-260.

Mittelstadt, B.D., & Floridi, L. (2016). “The ethics of big data: Current and foreseeable issues in biomedical contexts”. *Science and Engineering* Ethics, 22(2): 303-341.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L. (2016). “The Ethics of Algorithms: Mapping the Debate”. *Big Data & Society*, 3 (2): 1-21.

Möller, J., Trilling, D., Helberger, N., & van Es. B. (2018). “Do Not Blame It on the Algorithm: An Empirical Assessment of Multiple Recommender Systems and Their Impact on Content Diversity”. *Information, Communication & Society*, 21 (7): 959–77.

Mori, M. (2001). *Libertà, necessità, determinismo*, Bologna (Italy): Il Mulino.

Moor, J.H. (2006). “The nature, importance, and difficulty of machine ethics”. *Intelligent Systems IEEE*, 21(4).

Moor. J.H. (1997). Towards a Theory of Privacy in the Information Age. ACM SIGCAS Computers and Society, 27: 27– 32.

Morley, J., Machado, C., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2019). “The Debate on the Ethics of AI in Health Care: A Reconstruction and Critical Review”. *SSRN Electronic Journal*.

Nah, F.FH., & Siau, K (2020). “COVID-19 Pandemic – Role of Technology in Transforming Business to the New Normal”. *HCI International 2020. Late-Breaking Papers: Interaction, Knowledge and Social Media. HCII 2020. Lecture Notes in Computer Science* (edited by C. Stephanidis *et alia*), 12427.

Naik, G., & Bhide, S.S. (2014). “Will the future of knowledge work automation transform personalized medicine?”. *Applied & Translational Genomics*, Inaugural Issue, 3(3): 50-53.

Nakamura, L. (2013). *Cybertypes: Race, Ethnicity, and Identity on the Internet*. New York, NY (USA): Routledge.

Negroponte, N. (1995). *Being Digital*. New York, NY (USA): Vintage Books.

Newell, S., & Marabelli, M. (2015). "Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'". *The Journal of Strategic Information Systems*, 24(1): 3-14.

Neyland, D. (2016). "Bearing accountable witness to the ethical algorithmic system". *Science, Technology & Human Values*, 41(1): 50-76.

Nichols, P. (201). 'Wide reflective equilibrium as a method of justification in bioethics,' *Theoretical medicine and Bioethics*, 33(5): 325-341.

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*, Stanford, CA (USA): Stanford University Press.

Noble, S. (2018). *Algorithms of Oppression*. New York: NYU Press.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations". *Science*, 366 (6464): 447-453.

Ochigame, R. (2019, December 20). *The Invention of "Ethical AI"*, 2019. Retrieved March 10, 2021.

Orseau, L., & Armstrong, S. (2016). "Safely interruptible agents". *UAI'16: Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*: 557-566.

O'Neil, C. (2016). *Weapons of Math Destruction*. London (UK): Penguin.

Oshana, M., (1998). "Personal Autonomy and Society," *Journal of Social Philosophy*, 29(1): 81–102.

Oshana, M. (2006). *Personal Autonomy in Society*, Hampshire, UK: Ashgate.

Oshana, M. (2005). “Autonomy and Self Identity,” in Christman and Anderson (eds.), pp. 77–100.

Paraschakis, D. (2018). “Algorithmic and Ethical Aspects of Recommender Systems in E-Commerce”. *Studies in Computer Science* (4), Malmö (Sweden).

Pariser, E. (2011), *The Filter Bubble: What the Internet is Hiding from You*. London (UK): Viking.

Parsell, M. (2008). “Pernicious Virtual Communities: Identity, Polarisation and the Web 2.0”. *Ethics and Information Technology*, 10 (1).

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA (USA): Harvard University Press.

Pasquale, F. (2016). “Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power”. *Social Science Research Network*, Social Science Research Network Rochester, NY.

Patterson, S. (2013). *Dark Pools: The Rise of AI Trading Machines and the Looming Threat to Wall Street.* Random House.

Perra, N., & Rocha, L. (2019). “Modelling Opinion Dynamics in the Age of Algorithmic Personalization”. *Scientific Reports* 9 (1) 7261.

Persson, I., & Savulescu, J. (2016). “Moral Bioenhancement, Freedom and Reason”, *Bioethics* (9): 263-268.

Pettit, P. (2001). *Republicanism: A Theory of Freedom and Government*. Oxford: OUP.

Pettit, P. (2011). “The Instability of Freedom as Non-Interference: The Case of Isaiah Berlin”. *Ethics*, 121 (4): 693-716.

Pettit, P. (2001). *A Theory of Freedom: From the Psychology to the Politics of Agency*, Cambridge (UK), Polity Press.

Pettit, P. (2014). *Just Freedom: A Moral Compass for a Complex World*. New York, NY (USA): WW Norton and Co.

Portmess, L., & Tower, S. (2014). “Data barns, ambient intelligence and cloud computing: The tacit epistemology and linguistic representation of Big Data”. *Ethics and Information Technology*, 17(1): 1-9.

Rawls, J. (1971). *A Theory of Justice*, Cambridge, MA: Harvard University Press.

Raz, J. (1986). *The Morality of Freedom*. Oxford (UK): OUP.

Raymond, A. (2014). “The dilemma of private justice systems: Big Data sources, the cloud and predictive analytics”. *Northwestern Journal of International Law & Business*.

Richards, N.M. (2013). “The Dangers of Surveillance”. *Harvard Law Review*, 126(7): 1934-1965.

Richards, N.M. (2008) Intellectual Privacy. Texas Law Review, Vol. 87, p.387, 2008, Washington U. School of Law Working Paper No. 08-08-03.

Richardson, R., Schultz, J., & Crawford., K. (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice, *N.Y.U. L. Review*, *94* (192), Retrieved March 10, 2021.

Robbins, S. (2019). A Misdirected Principle with a Catch: Explicability for AI. *Minds and Machines*, *29* (4), 495–514.

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2019). "The Chinese Approach to Artificial Intelligence: An Analysis of Policy and Regulation". *SSRN Electronic Journal*.

Rosenberg, M. (2018). "Bolton Was Early Beneficiary of Cambridge Analytica's Facebook Data". *The New York Times*. March 23 2018.

Rouvroy, A. (29th January 2015). "Algorithmic governmentality: a passion for the real and the exhaustion of the virtual". *All watched over by algorithms*. Berlin.

Royakkers, L., Timmer, J., Kool, L. & van Est, R. (2018). Societal and Ethical Issues of Digitization. *Ethics and Information Technology*, *20*(2), 127–142.

Sandel, M.J. (1982). *Liberalism and the Limits of Justice*, Cambridge: Cambridge University Press, 2nd ed., 1999.

Scanlon, T.M. (2002), "Rawls on Justification", in The Cambridge *Companion to Rawls*, S. Freeman (ed.), Cambridge: Cambridge University Press, pp. 139–167.

Schermer, B.W. (2013). "Risks of profiling and the limits of data protection law". In: Custers, B, Calders, T, Schermer, B, et al. (eds) *Discrimination and Privacy in the Information Society*. Berlin: Springer, pp. 137–152.

Schermer, B.W. (2011). "The limits of privacy in automated profiling and data mining". *Computer Law & Security Review*, 27(1): 45-52.

Scheibehenne, B., Greifeneder, R., & Todd, P.M. (2010). "Can there ever be too many options? A meta-analytic review of choice overload". *Journal of Consumer Research*, 37: 409-425.

Schreurs, W., Hildebrandt, M., Kindt, E., et al. (2008) “Cogitas, ergo sum. The role of data protection law and non-discrimination law in group profiling in the private secto”r. In: Hildebrandt, M, Gutwirth, S (eds) *Profiling the European Citizen: Cross-Disciplinary Perspectives*. Dordrecht: Springer, pp. 241–270.

Schroeter, F. (2004), “Reflective Equilibrium and Anti-theory”, *Noûs*, 38(1): 110-134.

Schwartz, B. (2004). *The paradox of choice: Why less is more.* New York, NY (USA): Harper Collins.

Siegel, E. (2013). *Predictive Analytics: the Power to Predict who will Click, Buy, Lie or Die*. John Wiley and Sons.

Seaver, N. (2018). Captivating algorithms: Recommender systems as traps. *Journal of Material Culture*, 135918351882036.

Seng Ah Lee, M. & Floridi, L. (2020). Algorithmic Fairness in Mortgage Lending: From Absolute Conditions to Relational Trade-Offs. *Minds & Machines*.

Shah, H. (2018). Algorithmic Accountability. *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences*, *376* (2128): 20170362.

Shapiro, S. (2020). Algorithmic Television in the Age of Large-scale Customization. *Television & New Media*, *21*(6).

Shardanand, U., & Maes, P. (1995). “Social information filtering: algorithms for automating word of mouth”. In I. R. Katz, R. Mack, L. Marks, M. B Rosson & J.

Solove, D. J. (2013). “Privacy self management and the consent dilemma”. *Harvard Law Review*, 126, 1880–1893.

Nielsen (Eds.), Proceedings of the SIGCHI conference on human factors in computing systems (CHI ’95) (pp. 210–217). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co.

Shin, D. & Park, Y. G. (2019). “Role of Fairness, Accountability, and Transparency in Algorithmic Affordance”. *Computers in Human Behavior*, 98 (September): 277–84.

Silverstone, R., (2007). *Media and Morality: On the Rise of Mediapolis*, Polity Press, Cambridge (MA).

Simon, J. (2015). “Distributed epistemic responsibility in a hyper-connected era”. *The Onlife Manifesto* (edited by Floridi L). Springer International Publishing: 145-159.

Simonite, T. (2020, October 7). *Meet the Secret Algorithm That's Keeping Students Out of College*. Wired. Retrieved March 11, 2021.

Stark, M., & Fins, J.J. (2013). “Engineering medical decisions”. *Cambridge Quarterly of Healthcare Ethics*, 22(4): 373-381.

Skinner, Q. (2012). *Liberty before Liberalism*. Cambridge (UK): Cambridge University Press.

Skinner, Q. (2008). *Hobbes and Republican Liberty*. Cambridge (UK): Cambridge University Press.

Skinner, Q. (2008). *The Genealogy of Liberty*. Public Lecture, UC Berkley.

Sunstein, C.R. (2001). *Republic.com*. Princeton, NJ (USA): Princeton University Press.

Sunstein, C.R. (2007). *Republic.com 2.0*. Princeton, NJ (USA): Princeton University Press.

Sunstein, C. (2008). "Democracy and the Internet". In J. van den Hoven & J. Weckert (Eds.), *Information Technology and Moral Philosophy*. Cambridge University Press, 93–110.

Sunstein, C.R. (2016). *The ethics of influence*. Cambridge (UK): Cambridge University Press.

Sunstein, C.R. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ (USA): Princeton University Press.

Sweeney, L. (2013). "Discrimination in online ad delivery". *Queue*, 11(3) 10:10-29.

Reid, R. 1788 [1969]. *Essays on the Active Powers of the Human Mind*, ed. Baruch Brody, Cambridge, MA: MIT Press.

Taddeo, M. (2010). "Modelling trust in artificial agents, a first step toward the analysis of e-trust". *Minds and Machines*, 20(2): 243-257.

Taddeo, M., & Floridi, L. (2015). "The debate on the moral responsibilities of online service providers". *Science and Engineering Ethics*: 1-29.

Taylor, C. (1991). *The Ethics of Genuineity*, Cambridge, MA: Harvard University Press.

Taylor, L., Floridi, L., & van der Sloot, B. (2017). *Group Privacy: New Challenges of Data Technologies*, 1st ed. New York, NY (USA): Springer.

Tene, O., & Polonetsky, J. (2013). *Big Data for all: Privacy and user control in the age of analytics*, Nw. J. Tech. & Intell. Prop.

Thaler, R., & Sunstein, C. (2009). *Nudge: Improving decisions about health, wealth and happiness*. London (UK): Penguin.

Thomas, P. (2017). *Self-determination: The Ethics of Action*, volume 1, Oxford: Oxford University Press.

Torous, J. *et alia*. (2020). "Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow", *JMIR Mental Health*, 7(3).

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2020). *The Ethics of Algorithms: Key Problems and Solutions*.

Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Journal on Telecommunications and High Technology Law*, *13*(203). Retrieved March 11, 2021.

Turner Lee, N. (2018). Detecting Racial Bias in Algorithms and Machine Learning. *Journal of Information, Communication and Ethics in Society*, *16* (3), 252–60.

Turilli, M. (2007). "Ethical protocols design". *Ethics and Information Technology*, 9(1): 49-62.

Turilli, M., & Floridi, L. (2009). "The ethics of information transparency". *Ethics and Information Technology*, 11(2): 105-112.

Turow, J. (2011). *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. New Haven, CT (USA): Yale University Press.

Tutt, A. (2016). *An FDA for algorithms. SSRN*, Rochester, NY: Social Science Research.

Upbin, B. (2011). “Facebook ushers in era of new social gestures” – Forbes.

Vaidhyanathan, S. (2018). *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. Oxford (UK): Oxford University Press.

Valentini, L. (2019) Respect for persons and the moral force of socially constructed norms. *Noûs*. 2019. 1–24.

Van den Hoven, J., & Rooksby, E. (2008). “Distributive justice and the value of information: A (broadly) Rawlsian approach”. *Information Technology and Moral Philosophy* (Edited by Van den Hoven J and Weckert J). Cambridge (UK): Cambridge University Press: 376-396.

Van den Hoven, J. (2010). “The use of normative theories in computer ethics”. *The Cambridge Handbook of Information and Computer Ethics* (Edited by Floridi L), Cambridge (UK): Cambridge University Press.

Van Otterlo, M. (2013). “A machine learning view on profiling”. *Privacy, Due Process and the Computational Turn-Philosophers of Law Meet Philosophers of Technology* (edited by Hildebrandt M and de Vries K). Abingdon (UK). Routledge: 41-64.

Van Wel, L., & Royakkers, L. (2004). “Ethical issues in web data mining”. *Ethics and Information Technology*, 6(2): 129-140.

Vasilevsky, N.A., Brush, M.H., Paddock, H., Ponting, L., Tripathy, S., Larocca, G.M., Haendel, M. (2013). “On the reproducibility of science: Unique identification of research resources in the biomedical literature”. *PeerJ*, 1.

Veale, M., & Binns. R. (2017). Fairer Machine Learning in the Real World: Mitigating Discrimination without Collecting Sensitive Data. *Big Data & Society*, *4* (2).

Vedder, A. (1999). "KDD: The Challenge to Individualism." *Ethics and Information Technology* 1 (4): 275–81.

Warren, S. and Brandeis, L.D. (1890) "The Right to Privacy". *Harvard Law Review*, 193(4).

Westlund, A. (2014). "Autonomy and Self-Care," in Veltman and Piper (eds.), pp. 181–98.

Yeung, K. (2017). "'Hyper-nudge': Big data as a mode of regulation by design". *Information, Communication and Society*, 20 (1): 118-136.

Yeung, K. (2018). "Algorithmic Regulation: A Critical Interrogation". *Regulation and Governance*, 12 (3): 505-523.

Yu, M., & Du, G. (2019). "Why Are Chinese Courts Turning to AI?". *The Diplomat*.

Zarsky, T. (2013). "Transparent predictions". *University of Illinois Law Review*, 2013 (4).

Zarsky, T. (2016). "The trouble with algorithmic decisions an analytic road map to examine efficiency and fairness in automated and opaque decision making". *Science, Technology & Human Values*, 41(1): 118-132.

Zhou, N., Zhang, C.T., Lv, H.Y., Hao, C.X., Li, T.J., Zhu, J.J., Zhu, H., Jiang, M., Liu, K.W., Hou, H.L., Liu, D., Li, H.Q., Zhang, G.Q., Tian, Z.B., Zhang, X.C. (2019). "Concordance Study Between IBM Watson for Oncology and Clinical Practice for Patients with Cancer in China". *The Oncologist*, 24 (6): 812-19.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York, NY (USA): Public Affairs.

Zuiderveen Borgesius, F.J., Trilling, D., Möller, J., Bodó, B., de Vreese, C.H., Helberger, N. (2016). “Should We Worry About Filter Bubbles?”. *Internet Policy Review*, *5*(1).