Alma Mater Studiorum – Università di Bologna
in partnership with LAST-JD Consoritum
Università degli studi di Torino
Universitat Autonoma de Barcelona
Mykolas Romeris University
Tilburg University

DOTTORATO DI RICERCA IN

Erasmus Mundus Joint International Doctoral Degree in
Law, Science and Technology

Ciclo XXXII – A.A. 2016/2017

Settore Concorsuale di afferenza: 12H3
Settore Scientifico disciplinare: IUS20

TITOLO TESI

PERSONALIZATION IN SOCIAL MEDIA:
CHALLENGES AND OPPORTUNITIES FOR DEMOCRATIC SOCIETIES

Presentata da: Urbano Reviglio

Coordinatore Dottorato   Supervisore

Prof. Monica Palmirani   Prof. Giovanni Ziccardi

Esame finale anno 2020

Alma Mater Studiorum – Università di Bologna
in partnership with LAST-JD Consoritum
Università degli studi di Torino
Universitat Autonoma de Barcelona
Mykolas Romeris University
Tilburg University

Ph.D.Programme In

Erasmus Mundus Joint International Doctoral Degree in
Law, Science and Technology

Cycle XXXII

Settore Concorsuale di afferenza: 12H3
Settore Scientifico disciplinare: IUS20

Title of the Thesis

PERSONALIZATION IN SOCIAL MEDIA:
CHALLENGES AND OPPORTUNITIES FOR DEMOCRATIC
SOCIETIES

Submitted by: Urbano Reviglio

The Ph.D. Programme Coordinator   Supervisor
Prof. Monica Palmirani

                                                   Prof. Giovanni Ziccardi

Year 2020

ABSTRACT

Personalization algorithms perform a fundamental role of knowledge management in order to restrain information overload, reduce complexity and satisfy individuals. Personalization of media content in mainstream social media, however, can be used for micro-target political messages, and can also create filter bubbles and strengthen echo chambers that restrain the exposure to diverse, challenging and serendipitous information. These represent fundamental issues for media law and ethics both seeking to preserve autonomy of choice and media pluralism in democratic societies. As a result, informational empowerment may be reduced and group polarization, audience fragmentation, conspiratorial thinking and other democratically negative consequences could arise. Even though research about the detrimental effects of personalization is more often inconsistent, there is no doubt that in the long run the algorithmic capacity to steer our lives in increasingly sophisticated ways will dramatically expand. Key questions need to be further discussed; for instance, to what extent can profiling account for the complexity of individual identity? To what extent are users, media and algorithms responsible in such practices? What are the main values and trade-offs that inform designers in such a fundamental societal algorithmic arbitrage? How is social media's personalization directly or indirectly regulated in the European Union? Is there the need for further regulation to tackle its challenges?

The thesis firstly presents a critical overview of information societies, analyzing social media content personalization practices, dynamics and unintended consequences. Secondly, it explores the role of serendipity as a design and ethical principle for social media. Serendipity is also a metric for assessing personalization quality. With serendipity being both limited and cultivated in the digital environment, the research reveals a theoretical trade-off between relevance (what a user is supposed to want) and serendipity (what a user may want). As such, it represents not only a cutting-edge technical challenge but it also underlies a powerful metaphorical and educational value. Thirdly, the European legal landscape with regard to personalization is analyzed from a regulatory, governance and ethical perspective. It is thus advocated co-regulatory approach to tackle the outlined risks of social media personalization. Finally, it is introduced the concept of 'algorithmic sovereignty' as a valuable abstraction to begin to frame technical, legal and political preconditions and standards to preserve users' autonomy, and to minimize the risks arising in the context of personalization.

Keywords: Social Media, Personalization, Internet Regulation, Algorithm Governance, Digital Ethics, Serendipity

ACKNOWLEDGEMENTS

Contents

0. Thesis Outline

This thesis is based on seven articles published during the Ph.D. period (5 individual articles and 2 collaborations in which I was the correspondent author as well as the main contributor). The research was focused on increasing concerns resulting from mainstream social media's personalization, that is, the design and algorithmic systems that provide affordances for managing information filtering and that automatically adapt information to users' (supposed) interests. In particular, information asymmetries between service providers and users, biases in algorithms, so-called 'informational bubbles' and the threats to autonomy, privacy and self-determination are analyzed. As such, the thesis explores these risks drawing from the phenomenon of online personalization, with a special focus on news personalization in social media, with the aim to understand its effects on identity, societies, to eventually analyze its ethics, regulation and governance. In particular, the following research questions are addressed:

> **RQ1**: *What are the challenges of social media's personalization for democratic societies?*

> **RQ2**: *How social media's personalization is regulated? More specifically, is the European legislation able to prevent the risks arising from personalization?*

> **RQ3**: *Should social media's personalization be further regulated? If yes, how?*

Taking into account that the interdisciplinarity of the proposed research spans from sociology to ethics, from governance studies to European regulation, the research methodology benefitted both from direct sources of documentation and from data collection, processing, interpretation and verification. The method of triangulation was therefore the preferred one. In order to carry out the analysis, the methodological approach was based on the established methods of interpretive research and analysis of Science and Technology Studies, critically reviewing literatures from an interdisciplinary perspective. The thesis is developed as follows;

Chapter 1 introduces a critical overview of information societies. In the last decade, an oligopolistic social media market emerged, as well as a new 'economy': the attention economy. The goal of mainstream social media platforms is chiefly to engage users and maximize data collection. Data is the core element of information societies. Users are thus surveilled and data is analyzed for the purposes of generating profit. Subsequently, data is refined in order to offer users increasingly sophisticated services. Yet, while big corporations 'win and take-all', dramatically enhancing their capacity to analyze such data and provide useful and persuasive services, human beings are overloaded and unable to effectively manage this new form of life and to properly navigate the complexities and risks that digitally-mediated information consumption and social relations actually imply. In a personalized and weaponized infotainment landscape, many individuals often experience an illusion of knowledge and socialization that, among several potentially negative consequences, tend to lead to narcissism, isolation and political polarization. A major consequence is the pollution of public debate, in which truth becomes increasingly subjective, thus contested; an era in which misinformation and disinformation strive to capture users' attention, either for profit or to persuade and control them. This chapter introduces the main characteristics and limitations of current mainstream social media as a public sphere.

Chapter 2 unveils and discusses the phenomenon of online personalization, with a special focus on news recommender systems in social media. Online personalization performs a fundamental social role of knowledge management in order to restrain information overload, reduce complexity and, ultimately, satisfy individuals. Among several challenges, online personalization – particularly in social media – can create phenomena such as 'filter bubbles' and strengthen 'echo chambers' which tend to restrain the exposure to diverse, challenging and serendipitous information. These represent fundamental issues for media law and ethics seeking to preserve autonomy of choice and pluralism in democratic societies. The risk is the reduction of individual agency and the polarization and fragmentation of society at large. Even though research on the detrimental effects of online personalization is often unconvincing, there is no doubt that in the long term the algorithmic capacity to steer our lives in increasingly sophisticated ways will dramatically expand. Therefore, key questions are addressed; for example, to

what extent profiling accounts for the complexity of individual identity? And what are the main values and trade-offs that inform designers in the fundamental societal arbitrage of information filtering and design choices?

Chapter 3 presents and discusses the phenomenon of serendipity, its metaphorical, political and design value in digital environments, especially in social media. Often undervalued because of its unpredictable nature, operational difficulty and wide or sometimes vague definition, serendipity as a design and ethical principle is attracting ever more interest, especially 'artificial' and 'digital' serendipity. Being influenced by environmental and human factors, the experience of serendipity encompasses fundamental phases of production, distribution and consumption of information. On the one hand, design for serendipity can increase media diversity, information quality and, more generally, discoverability. On the other hand, serendipity is a capability that has a strong political value. It helps individuals to internalize and adopt strategies that increase the chances of experiencing it. As such, the pursuit for serendipity can actually help to burst filter bubbles and weaken echo chambers in social media. The chapter firstly analyzes serendipity and its study. Then, it discusses the role and the ethics of digital serendipity and its design from an interdisciplinary perspective. The conclusion is that serendipity can be conceived as a metaphor, a technical challenge, an ethical design principle, and an educational goal. Limitations are also discussed.

Chapter 4 analyzes the phenomenon of personalization from a regulatory (4.1), ethical (4.2) and governance (4.3) perspective. The analysis takes into account the interdisciplinary nature of personalization and it is focused on the European legal context. To begin with, modes of regulation and an overview of social media regulation are broadly introduced. Then, an analysis of the European Legislation is provided. After having discussed a human rights perspective, there are briefly analyzed the E-Commerce Directive, E-privacy Directive, the General Data Protection Regulation and, more generally, media law. In the second Section, an ethical approach to personalization is presented, introducing the role of ethics in European legislation and Artificial Intelligence and other relevant issues such as 'group privacy' and the ethics of nudging. In the third Section the governance of personalization as a synthesis of regulation and ethics is

explored. Additionally, it is questioned the role of Public Service Media (PSM) in cultivating a healthier digital public sphere.

Chapter 5 draws conclusions and provides some general recommendations. By doing so, it is initially framed and discussed the notion of "algorithmic sovereignty" as a fundamental concept to further develop in order to ultimately provide, defend and cultivate people's autonomy to govern one's own digitally personalized life.

1. An Introduction to Information Societies

"Every technology is used before it is completely understood"

— Leon Wieseltier, Critic and writer

We live in Information Societies, where the information flowed by Information Communication Technologies (ICTs) have profoundly changed many aspects of our lives. This shift is now pervasive and it is still changing many aspects of social life. In essence, it is blurring the distinction between reality and virtual; public and private; truth and opinion; and between human, machine and nature. We now live *onlife*, in the sense that the majority of people are always hyper-connected and rely on Internet as a second skin for almost everything. We have thus moved inside the "infosphere", that is the whole informational environment constituted by all informational entities, their properties, interactions, processes, and mutual relations (Floridi, 2014). Our information society is now better seen as a neo-manufacturing society in which data and information are replacing raw materials and energy; not just communication and transactions but the creation, design, and management of data and information.

In information societies, networks constitute the new social morphology. The ICTs revolution, in fact, brought societies from information scarcity to information abundance, and from the primacy of entities to the primacy of interactions. Communication networks are indeed the fundamental networks of power-making in society, giving rise to a 'Networked Information Economy' based on digital information (Castells, 2011), whose disruptive features are the zero-marginal cost of reproduction and the non-rivalry feature of data (Benkler, 2006). At the same time, media convergence, namely the merging of the old media over a single digital network - the Internet - is also profoundly changing our social relations and the production and consumption of information (Jenkins, 2007).

Internet creates a number of incomprehensible therabyte of data, a number expected to double every year, this has exponential growth potential. Everything and everybody will be increasingly connected to the Internet of (every)Things (so-called IoT or IoE). We have only recently entered into the era of *big data*, in which the use of predictive analytics or other certain advanced methods to extract value from large amounts of data is becoming increasingly pervasive and effective. Several risks and opportunities arise from such a revolution; for example, private firms, government and academic research often

uncritically accept this socio-technical phenomenon which carries particular epistemological and cultural views on scientific praxis and the nature of reality. Conversely, critics note that big data contributes to the creation of an Orwellian society, in which privacy and civil rights are disregarded to favour corporate interest and state control. While many advantages are clear, contrastingly the challenges are unpredictable and, eventually, extremely complex to govern.

Data is not only changing individuals, but is also demonstrative of collective processes of knowledge and decision-making. Pentland (2015) anticipates the emergence of a truly Comtian "social physics", the possibility to finally create a data-driven mathematical model of social behaviours from big data. It appears indeed imminent a profound change at the levels of epistemology and ethics. This trend led some scholars like Chris Andersen (2008) to even argue "the end of theory". In his famous article, he claimed that science can nowadays advance even without coherent models, unified theories, or really any mechanistic explanation at all. Yet, as Thomas Kuhn (1962) famously argued, any data is always "theory-laden", and any data correlation is useful only under certain background assumptions, which ultimately come from theory. Without it, correlations can be as misleading as they are informative (so-called *apophenia*). The assumption that big data alone will always result in more predictive power is *hubris*: it might have profound consequences on information societies. It may even climax in what historian Noah Yuval Harari (2016) provocatively defines as "dataism", a worldview that perceives "the entire universe as a flow of data, see organisms as little more than biochemical algorithms and believe that humanity's cosmic vocation is to create an all-encompassing data-processing system and then merge into it (p.134)." 'Dataists' would believe in the invisible hand of the dataflow. Thus, given enough biometric data and computing power, this all-encompassing system could understand humans much better than humans understand themselves to the point that they will uncritically and willingly delegate their autonomy and any other shared authority to algorithms. This temptation is frequently becoming a reality that many do not fully recognize, specifically within its ethical and socio-political consequences.

In this emerging and fast-paced landscape, the challenges and the opportunities of the governance of data and the Internet are unique and extremely complex. No other network in the past has exhibited the Internet features, a fact that poses unprecedented

methodological challenges for scholars. The search for concepts, tools, and categories in order to make sense of the governance of the Internet is very much open-ended. Moreover, it is becoming an area with profound socio-economical and geopolitical conflicts. Its increasing privatization also poses nascent issues. Commercial entities, in fact, increasingly undermine the Internet's pioneer spirit and, thus, its traditional democratic – and to some extent libertarian – values. They often use closed systems and integrate proprietary technologies that constitute new modes of economic domination that are imposed on the users without a method of opting out entirely. More often, these tech giants are quasi-monopolies in the markets in which they operate. Similarly, nation States are becoming mutually dependent by the services they increasingly offer to society. These trends radically challenge societal functions. More than ever Internet needs to be safeguarded critically and reflexively. The negotiation of competing values on these conflicts, in fact, will have a significant bearing on global innovation policy, national security, freedom of expression and societal cohesion.

Social media, in particular, are fundamental gatekeepers of information societies. They are increasingly more essential for individuals to discover information, form opinions and participate in public life. Though in social media and search engines individuals receive unprecedented access to information and people of widely varied backgrounds, this increasing availability of information does not guarantee the quality of information and communication. In particular, the exposure to diverse and challenging information, viewpoints and worldviews is fundamental for promoting critical thinking and informed decision-making, and even preventing or correcting inaccurate beliefs or dangerous radicalizations. Several concurring phenomena constrain the supposed power of the Internet to connect people and ideas, depending on the main actors involved: individuals, information intermediaries (e.g. platforms or newspapers) and technology itself (e.g. algorithms or design).

Naturally, there are human proclivities that limit the quality of debate in the public arena. In general, selective exposure to information, confirmation bias and homophily negatively affect the processes of information seeking and knowledge acquisition. Then, the increasingly complexity of interconnected societies, along with the rapid development of science and technology, make these processes even more difficult. Secondly, Internet itself as an artifact shapes how we acquire information: the so-called 'information

overload' and the risks that it leads to illusionary knowledge. Thirdly, the role of information intermediaries is prominent. They design the platforms that mediate our social relations and the algorithms that filter information and provide us a personalized Internet experience. These can produce two intertwined phenomena: at the individual level *filter bubbles* while at the collective level *echo chambers*. Yet, whether these ultimately lead to collective political action or simply more audience fragmentation and political polarization is a question that has been contested by social and political theorists since the Internet has developed into a topic for academic enquiry. Lastly, the current social media landscape is weaponized by states and private actors willing to micro-target users in order to persuade them and, in some cases, for manipulative purposes. As a consequence, trust and truth are also undermined.

This chapter will briefly analyze the Internet as a 'public sphere', in the Habermasian sense. After having highlighted the ambiguities of the Internet, an introduction to social media will be done, firstly from a more general perspective and, consequently, from the perspective of all the actors involved. It will be analyzed through the role of information intermediaries and governments in their sustainment and benefit of a surveillance-complex industry. Followed by, a user-perspective on the limitations of acquiring knowledge and socialization in the information era. Finally, it will analyze how the Internet is increasingly fragmented and weaponized by several actors, and how this may lead to what has been referred to as the 'post-truth era of politics'.

1.1 Internet is a Public Sphere?

Most of the debate on the public sphere is based on the normative prerogatives of Habermas' theory of public sphere, which is the "realm of our social life in which something approaching public opinion can be formed" (Habermas et al., p.49, 1974). Castells (2011) presented the idea that the public sphere has moved from the physical world to the network. Indeed, Internet can be viewed as a new type of networked public sphere that allows all citizens to change their relationship to the public sphere. They no longer need to be consumers and passive spectators, they can become creators and primary subjects. In this sense the Internet is a democratizing force (Benkler 2006). Papacharissi (2008) also described the emergence of a "virtual sphere 2.0", in which citizen-consumers participate and express "dissent with a public agenda […] by

expressing political opinion on blogs, viewing or posting content on YouTube, or posting a comment in an online discussion group" (p. 244). Yet, the Internet is a network of different communication spaces rather than one monolithic public sphere. Certainly, mainstream social media platforms like Facebook, Twitter and Youtube represent a major space of the western public sphere.

According to Habermas, two conditions are necessary to structure a public space: freedom of expression, and discussion as an integrating force. The architecture of the Internet was said to articulate these two conditions. Currently, however, it deviates from these ideals, and its development into a public sphere without the limitations inherent in this model has been proven to be largely false hope. As Splichal (2007) argues, the public sphere has actually never materialized because of "unequal access to communication channels, uneven distribution of communicative competence, and the reduction of public debates to a legitimisation of dominant opinions created by either the 'business type' or the 'government type' of power elites" (p. 242). For these reasons, Fuchs (2015) argues against idealistic interpretations of Habermas and advocates for a cultural-materialist understanding of the public sphere concept that is grounded in political economy. Habermas' theory should then best be understood as a method of immanent critique that critically scrutinizes limits of the media and culture grounded in power relations and political economy or, better, the political economy of knowledge.

Yet, it is also true that over the last decades Internet has – to some extent – democratized knowledge. This occurred chiefly in three ways (Lynch, 2015):

1. *Availability*. First, and most conspicuously, the Internet, like the printing press before it, has made bodies of knowledge more widely available. It has greatly expanded this process whilst also changing both the amount of different kinds of information available and also, the speed at which that information can be accessed. A prominent example is piracy for books and academic papers with websites such as *Scihub* (for academic articles) and *Libgen* (for books).

2. *Inclusivity*. The Internet has also democratizing knowledge by making its production more inclusive. One common example is *open source* software like Mozilla's Firefox web browser. When security vulnerabilities or bugs arise, a diverse and widespread community of volunteers work on fixes and plug-ins. Open source software is software by the people, for the people: it is crowd and

collective intelligence in action purely for the common good.

3. *Transparency*. A third way that the Internet has democratized knowledge is by making what is known more transparent - particularly with regard to information held by governments. The most obvious and controversial example of this is *WikiLeaks*, a nonprofit organization that publishes news leaks and classified governmental information online. The Internet can be used to shine a light on all sorts of activities, arguably empowering citizens.

On the one hand, it is often argued that the Internet, by promoting equal access to diverging preferences and opinions in society, actually increases information diversity. Many scholars characterize the online media landscape as the "age of plenty", with an almost infinite choice and unparalleled pluralization of voices that have access to the public sphere (Karppinen, 2014). Some argue that social media will disrupt the traditional elite control of media and amplify the political voice of non-elites (Castells, 2011). Some argue that digital tools such as social media will inevitably lead to the pluralization of the public sphere giving way to non-mainstream political actors to influence the political agenda. It is well known that traditional media have a bias in selecting what to report and in choosing perspective on a particular topic. Selection bias, organizational factors, advertisers, and government influences can all affect which items will become news. Evidence of bias may range from the topic choice of the *New York Times* to the choice of think-tanks that the media refer to. Yet, even in social media it is a small number of users that generally determine political communication gatekeepers that exert strong and selective influence on the information.

On the other hand, there are skeptical voices that argue that the Internet has not fundamentally changed the concentrated structure typical of mass media, but reflects the previously recognized inequalities (Karpinnen, 2008). It is also argued that it has brought about new forms of exclusion and hierarchy. While it has increased political participation to some degree, it has also empowered a small set of elites that strongly shape how political material is presented and accessed. Others have pointed out the danger of fragmentation of information caused by personalization algorithms (Pariser, 2011; Sunstein, 2017). They argue that the filters we choose on the internet, or the filters that are imposed upon us, are weakening and will weaken the democratic process. On the one hand, they allow citizens to be exposed mostly to what they like and join into groups that

share their own views and values, and cut themselves off from any information that might challenge their beliefs. Group deliberation among like-minded people can create polarization; individuals may lead each other to falsehoods, simply because of the limited argument pool and the role of social influences. Increased polarization makes it more difficult for society to find common ground on important issues. On the other hand, the ability of online intermediaries such as recommender systems and social networks to customize their items to the taste of individuals, together with users' preference to reading opinions which reinforce their own viewpoints, raises the phenomenon referred to as filter bubbles, which might limit the discovery of novel, serendipitous information and, eventually, interests.

The changes in how knowledge is produced and distributed affect the politics of knowledge because they directly influence the 'economy of knowledge'. The economy of knowledge (Lynch 2015) refers to the structure of relations that divides epistemic labor and governs its exchange. The globalization of the economy of knowledge, however, has some of the same effects as the globalization of the economy in general. The unfettered global economy is not only increasing economic inequality, it is also encouraging epistemic inequality. The value of epistemic equality is that all persons have a basic claim to the same epistemic resources. An epistemic resource is a structure or institution that provides information and, at least, the basis for knowledge. Thus, epistemic inequality is the result of an unfair distribution of structural epistemic resources. Thus, even if knowledge is more "democratized" now it means little in conditions of increasing epistemic equality. Social media has the potential to be a public sphere where inequality is reduced, however, this is limited by the steering media of political power and money so that corporations own and control and states monitor users' behaviors on social media.

1.2 An Introduction to Social Media

Social network sites (SNS) are the epitome of the *web 2.0*, and of the internet in general. Boyd and Ellison (2007) define SNS as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users – friends, contacts, followers, fans – with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site. The

terminology is also unclear, with some referring to social networking services as social media.

Instead, 'social media' became a buzzword in the mid-2000s. Most of the people understand SNS, blogs, wikis, user-generated content sharing sites and microblogs as social media. This, however, depends on what we define as "social" about social media (Fuchs, 2015). Sociality can in fact correspond to the social theory concepts of *social facts* (how human thought is shaped by society), *social relations* (humans exchange symbols by communicating in social relations), *co-operation* (humans work together and thereby create use-values) and *community* (how humans form and maintain communities). Described as information processes, it can also be expressed as a threefold interconnected process of (a) cognition, (b) communication and (c) cooperation. Today's social media afford all of these processes to a different extent. One of the main consequences is that traditional media is declining in their *gatekeeping* role in determining the agenda and selecting which issues and viewpoints reach their audiences. Most citizens have moved from traditional mediums such as newspapers and television to using social media. Political discussions on these platforms are becoming an increasingly relevant source of political information, often used as a source of quotes for media outlets, they are nowadays still the main gatekeepers.

Yet, social media are not just communication companies that uniquely afford socialization processes, but also – and mostly – large advertising agencies. It was already Marx's (1842) concern that the primary freedom of the press lies in not being a trade. There are indeed several problems of how capitalist media limits the public sphere:

- *Market concentration*. On the web, this is evident on the light of the "*network effect*", that is the service becomes more valuable when more people use it, in a self-reinforcing spiral effect. Today the social media landscape is mostly constituted by a bunch of platforms: Facebook, Instagram, Twitter, LinkedIn and Snapchat. The remaining platforms represent a rather small minority.

- *Infotainment*. Ads-financed media tend to focus on entertainment more than news, documentaries and educational content because this content is better suited for attracting advertisers. The influence of economic logic on the media would result in '*tabloidisation*': "Reporting facts as human-interest stories, mixing information with entertainment, arranging material episodically, and breaking down complex

relationships into smaller fragments – all of this comes together to form a syndrome that works to depoliticize public communication" (Habermas, 1996, p. 377).

- *Power inequalities*. There are power differentials in commercial media that disadvantage individuals and groups that lack financial weight, political influence and reputation, and disempower their voices and visibility. Private media ownership gives owners the possibility to influence media content (i.e. Facebook's *Instant Articles* in which the more you spend the larger the audience you reach). To some extent they become dependent on market and commodity logic, and prone to exclude voices that question these logics. Finally, there is an educational and economic gap that can privilege educated and wealthy individuals in the consumption of demanding and costly culture.

The contemporary social media world is thus shaped by three antagonisms (Fuchs, 2015): a) the political antagonism between users' privacy and the surveillance-industrial complex as well as citizens' desire for accountability of the powerful and the secrecy of power, b) users' data and social media corporations' profit interests, and c) the civil society antagonism between the creation of public spheres and their corporate and state colonization. These dimensions will be now briefly explored in order to outline their main characteristics and, especially, their main misconceptions. Followed by, a broader analysis of the limitations of socialization and knowledge acquisition processes in the digital era will be undertaken.

1.2.1 The Surveillance-Complex Industry
"*These programs were never about terrorism: they're about economic spying, social control, and diplomatic manipulation. They're about power.*"
— Edward Snowden, Former CIA analyst and whistleblower

Surveillance has become a key dimension of the contemporary world. Since 9/11 there has been a massive intensification and extension of surveillance that is based on the ideology that monitoring technologies, big data analysis, and predictive algorithms can prevent terrorism. It has been justified with a typical modern dichotomy, between security and freedom: we need both, but we cannot have one without sacrificing the other at least in part. The more we have of one, the less we have of the other (Bauman and Lyon, 2013).

Thus, instead of an "iron cage" of Weberian memory, around each of us is built relentlessly a more powerful "digital cage". This, however, will never thwart terrorism because terrorists are strategic enough not to announce their intentions on the Internet. On the contrary, evidence has shown that social media surveillance does not solely target terrorists, but has also been directed at protestors and civil society activists, while at the same time inducing the "*chilling effects*", which describes situations in which rights are threatened by the possible negative results of exercising these same rights, namely a deterrent effect of self-restriction (Büchi et al., 2019). This occurs not only in State surveillance but also in commercial surveillance. In other words, a significant negative externality of commercial algorithmic profiling is the potential threat to autonomy.

Edward Snowden's revelations about the existence of the *Prism* system have shed new light on the extension and intensity of state institutions' Internet and social media surveillance. Today, in light of the Datagate scandal and the dissemination of new systems of pervasive surveillance – Video-surveillance, Facial Recognition, Predictive Models, Emotion Detectors, Social Credit Systems, brain readers applied to workers, Lethal Autonomous Weapons etc. – we can say to live increasingly in a 'society of control' (Deleuze, 1992). In several countries – particularly in China – it emerged a massive "surveillance-industry complex" that risks to escape democratic control and accountability and threatens the free and open character of our societies (Hayes, 2012).

The concept of the military-industrial complex stresses the existence of collaborations between private corporations and the state's institutions of internal and external defence in the security realm. Sociologist Charles Wright Mills argued there is a power elite that connects economic, political and military power: "there is no longer, on the one hand, an economy, and, on the other hand, a political order containing a military establishment unimportant to politics and to money-making. There is a political economy linked, in a thousand ways, with military institutions and decisions. […] there is an ever-increasing interlocking of economic, military, and political structures" (Mills, p.4, 1956). In fact, according to the Snowden's leaked documents, the NSA in the PRISM programme obtained direct access to user data from seven mainstream online/ICT companies: Aol, Apple, Facebook, Google, Microsoft, Paltalk, Skype, Yahoo! (Fuchs, 2015).

Despite the hopes that the *Datagate* scandal could have triggered a serious debate on human and digital rights, the effects have been very limited. *Dataveillance* is an invisible

and complex practice so that most of the users argue that a lack of concern proves they have "nothing to hide", a fallacious argument resulting from a narrow conception of privacy (Solove, 2007). Snowden brightly summarized this incongruous widespread attitude: "arguing that you don't care about the right to privacy because you have nothing to hide is no different than saying you don't care about free speech because you have nothing to say" (Ha, 2014). Instead, the attempts to rule the Internet in order to constrain it have steadily increased since the Datagate scandal. Social media is still facing an economic and political antagonism between "users' interest in data protection and corporate tax accountability on the one side and corporations' interest in user data's transparency/commodification and corporate secrecy on the other side." (Fuchs, 2015, p.83). Exactly the opposite of the *cyberpunk* philosophy who animated many Internet pioneers – privacy for the weak, transparency for the powerful – and this is evident in the configuration of social media data-driven business model.

The currently capitalistic logic of accumulation produces "hyperscale assemblages of objective and subjective data about individuals and their habitats for the purposes of knowing, controlling, and modifying behavior to produce new varieties of commodification, monetization, and control" (Zuboff, p.85, 2015). This new global architecture of data capture and analysis indeed produces rewards and punishments aimed at modifying and commoditizing behaviour for profit. Many of the practices involved in what Zuboff famously defined as *surveillance capitalism* are challenging social norms associated with privacy and, thus, are contested as violations of rights and laws. Consequentially, big tech corporations have learned to "obscure their operations, choosing to invade undefended individual and social territory until opposition is encountered, at which point they can use their substantial resources to defend at low cost what had already been taken" (Ibid). In this way, surveillance assets are accumulated and attract significant surveillance capital while producing their own surprising new politics and social relations.

Why has such 'surveillance capitalism' emerged? There are a variety of reasons. Firstly, it was constructed at high velocity and designed to be undetectable. Structural asymmetries of knowledge and rights, in fact, made it impossible for people to learn about these practices. There was no historic precedent, so there were few defensive barriers for protection. Leading tech companies have been over-estimated, in a way respected and

treated as 'emissaries of the future'. On the other hand, individuals quickly came to depend upon the new information and communication tools as necessary resources, requirements – at times even preconditions – for social participation. Finally, as Zuboff (2015) critically argues: "the rapid build-up of institutionalized facts (…) produced an overwhelming sense of inevitability (p.85)." As Hildebrandt (2015) also put it: 'The temptation to accept that things are the way they are because technology is the way it is, has a strong hold on public imagination […] once a technology has consolidated it acquires a tenacity that is not easily disrupted.' (p. 174).

This totalizing system of control stands on users' attention and, eventually, money. And to capture them, increasingly sophisticated design systems are put in place without a true social awareness and, thus, defence and oversight. It is a pervasive and perverse system that might be detrimental for individual's well-being, social cohesion, and, ultimately, democracy.

1.2.2 The Attention Economy

"The average person checks their phone 150 times a day. Why do we do this? Are we making 150 conscious choices?"

— Tristan Harris, Former Google employee and founder of *Time Well Spent*

An overwhelming amount of information stimuli compete for our cognitive resources, giving rise to the 'attention economy'. In 1971 Herbert Simon articulated the concept of "attention economy" ante-litteram: "[T]he wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention". (Simon, 1971, p.40). The establishment of value in the attention economy is a dual register of what one pays attention to and what one chooses to ignore. Even if mainly used by economists, the notion of attention economy has also become an important object of critical analysis in recent years.

Big Tech companies conceives the attention of each user as a "sign of intention" of his or her individual interests and desires, and it is then used to identify patterns of consumer behaviours, preferences, tendencies, etc. Thus, it uses it a mechanism for the harvesting of large amounts of data (i.e. big data). The aim of the attention economy as a specific power apparatus is not to repress a given set of differences nor to normalise them under

a given norm but, indeed, to dispose them in a certain way so as to achieve the most effective economic result.

The attention of the users is, therefore, exploited, as it was – to a more limited extent – also in the pre-digital era. *Click-baiting* is, above all, based on titular sensationalism. For example, unfinished article titles create a "curiosity gap" that triggers the user to click. Today, however, information intermediaries' techniques for increasing profit through users' engagement also manifests itself in the development of more complex, often unconscious, addictive rituals. Algorithms, in fact, raise interesting questions with regard to psychological hedonism, that is the view that all human action is ultimately motivated by the desire for pleasure and the avoidance of pain (Gal, 2017). What if we create algorithms that could predict our choices and simply grant them to us? Emerging persuasive techniques can indeed create addictive behaviours that function acting directly on the system of the neurotransmitter dopamine (Turel et al., 2014). This may explain, why the majority (59%) of the URLs mentioned on Twitter are not clicked at all (Gabielkov et al., 2016). Users crave to share in order to receive notifications but, in the end, they rarely have enough attention to read each other's.

Internet users, of course, are not simply hedonistic consumers, but also producers or, more accurately, unpaid workers who generate economic value. Information societies can be seen as a global *click-farm*. In this context, the notion of 'digital labour' has been coined (Scholz, 2013). In order to access any social service platform, users have to give up the social data they generated. It is the business of the meta-data that allows the existence of "free" services. Thus, every time users surf the internet they leave their digital fingerprint, as if we were at the scene of a crime. Indeed, profiling is an activity that comes from criminology. In a sense, Internet users are like criminals to get to know in order to predict our desires and satisfy a compulsive thirst of consumerism (Ippolita, 2011). This is the chimera of the data-driven society, in which the human role is widely irrelevant and in which sociality is turned into economic value. Thus, users hold in a subordinate position, firstly being raw data producers without any reasonable possibility to opt-out rather than agree or leave their social networks and, secondly, by being fed with addictive habit loops, alienated by attractive multifunctional smartphones, more often with low quality information, even if they keep in their hand one of most powerful tool ever created by mankind (Ippolita, 2016).

It is also true that traditional commercial broadcasting treats audiences as a commodity. On mainstream social media, however, this is qualitatively different in a number of respects. For example, measuring audiences in traditional media has been based on studies with small samples of audience members, whilst on corporate social media it is constant, total and algorithmic; audience commodification on social media is based on the constant real-time surveillance of users; user measurement uses highly accurate predictive algorithms (i.e. collaborative filtering, if you like A, you may also like B because 100 000 people who like A also like B); user prices are often set based on 'algorithmic auctions' (*pay per view*, *pay per click*).

Turning user data into a private good controlled by social media companies is prominently legitimised for the sake of efficiency and innovation – in order to 'provide better services' – even with the help of privacy policies – in order to 'protect users' interests' – and finally with the "sharing is caring" narrative. Sharing, in effect, is not primarily a sharing of data with friends and the public, but sharing with social media platforms, who are the largest data processors and data profiteers in the world. This explains not just the recent rise of the term "big data", but also their interest in hiding their commercial interests ideologically behind the ideas of sharing and openness. Certainly, there is no easy-solution to make this system more 'sustainable' and 'socially acceptable'. Innovative principles and legal frameworks can certainly strengthen privacy. Still, most issues lie in the economic model in the first place. One of the most appealing and viable alternative models for social media – *subscription fee* – would not even represent a solution as it would likely increase inequalities further. Who is able to pay has the right to privacy, the rest have to give up their digital rights on a platform where even the quality of socialization is highly disputed.

1.2.3 The Limits of Digitally-Mediated Socialization

"We expect more from technology and less from each other."
— Sherry Turkle, Sociologist, Psychologist, Technologist

Nowadays, people join social media mostly to socialize with friends they already know offline (Marwick and Boyd, 2014). Hence, instead of connecting across vast spaces, for many users Internet is a mere intermediary of daily social relations. As well-known in cyber-psychology, however, Internet is changing our ability to experiment and recognize

emotions. Since the mediated interaction substitutes our body with the medium, to some extent this impedes the activation of the mechanisms of body simulation that make people understand each other's emotions (*mirror neurons,* namely those who unfold empathy) (Turkle, 2017). Thus, emotions are 'disembodied'. This is not a problem in itself; it is a natural form of evasion as literature and movies always represented. The problems indeed arise when most of people's emotions are disembodied because it brings, especially to the so-called "digital natives", a widespread "emotional illiteracy" (Galimberti, 2010). This phenomenon may explain the ailments of recent generations as their limited empathy which hampers their ability to understand themselves. An extreme and debatable example of this trend may be that of *Hikikomori,* which emerged in Japan but extended in the rest of the world to different degrees. It regards young individuals socially atomized and living mostly at home – or even only in their own bedrooms – while living and socializing mostly or only online.

Based on the literature available at this time, it seems that social media affects individual's well-being dependent on how it is used (Verduyn et al., 2017). On the one hand, social media has the potential to increase our subjective well-being by allowing us to increase our social capital and feeling of connectedness due to active usage. On the other hand, they can also be a significant cause of distress, especially when they elicit social comparisons and envy due to passive usage. Apparently, this seems to depend on how active you are on the platform: the more the better. More research, of course, is needed. In particular, to answer other relevant questions like: How do "active users" actually use platforms? What really drives them? Or: why do "passive users" not fully participate? How to design more inclusive platforms?

In mainstream social media, friends are often counted as followers and everything is gamified and accountable to be translated in the language of performance and efficiency, so that "everything that is not accountable cease to exist" (Han, 2015, p.51). As such, most of the users are dramatically persuaded to compete to seek more *likes*, whatever it takes. Ippolita (2016) defines this widespread proclivity as "emotional pornography", stressing the compulsive need of the users to share their own intimacy, namely valuable data for both advertising and mass surveillance. Bauman and Lyon (2013) explain this trend exemplifying a new strategic change, what they call "post-panoptic society": "on the one hand, the old ruse panoptic (you'll never know when I look at your body, and in

this way your mind will never stop feeling observed) is implemented gradually but consistently and seemingly unstoppable, on a scale almost universal. On the other hand, now that the old nightmare of panoptic "never alone" has given way to the apparent hope of "never to be alone" (or abandoned, ignored, rejected and excluded), the joy of being noticed takes the upper hand on the fear of being revealed" (p.39). In other words, the fear of loneliness overcomes the hope to preserve freedom.

Currently, mainstream social media does not manifest as a public space but falls within in the frame of the private and the exposure of the self. As a consequence, the public sphere is hampered and the growing narcissism can lead, parallel to *informational bubbles*, to a de-politicization of society (Byung-Chul Han, 2013). A paradigm of this phenomenon is the so-called *slacktivism*, which refers to the act of showing support for a political or social cause with the main purpose of boosting the egos of participants in the movement, with very limited involvement required nor concrete effect. On the other hand, those who conform to such a system becomes popular, and their content goes viral. To some extent, the resulting environment impedes many users to express themselves, sometimes becoming *lurkers*, mere watchers of what's happening, and often making them fall into a *spiral of silence* in which they do not publicly express their (likely critical) opinions (Noel-Neumann, 1984).

1.2.4 The Limits of Internet Knowledge

"We live in a world where there is more and more information and less and less meaning."
— Jean Baudrillard, Philosopher (p.79, 1994)

The digital revolution is not only deeply changing how individuals think and socialize, but even more how society and knowledge are structured and constructed. While some deterministically argue that the consequent multitasking and disintermediation of information are worsening our capacity of problem-solving and critical analysis (Carr, 2011), providing a "google-knowing" which represents a form of *receptive knowledge* far from a true understanding of reality (Lynch, 2015), other scholars stress the primary role of digital literacy in tackling such risks (Rheingold, 2012).

People generally believe that the more information they have the better their decisions will be. Yet, more knowledge can reduce the accuracy of prediction of uncertain outcomes and simultaneously can increase confidence in their prediction (???). In the

pyramid of knowledge there is data at its base, with information above, then knowledge and, finally, wisdom at the top (see Figure 1). It is not just about accumulating data and information: it's more about turning data and information into knowledge and, eventually, wisdom. The fast-paced media environment short-circuits the training processes of individuals and collectives, upsetting and putting at serious risk cognitive faculties, social cohesion and relationships (Stiegler, 2014).[1]

Figure 1. The Pyramid of Knowledge.

According to Lynch (2015) information technology even if they expand our ability to know in one way, is actually impeding our ability to know in other, more complex ways. In particular, ways that require to 1) take responsibility for our own beliefs and 2) work creatively to grasp and reason how information fits together. Because the Internet has promoted such passive way of knowing, it makes real objective knowledge harder to reach. With objective knowledge can be intended a correct belief that is grounded or justified, and which can therefore guide our action. It is then possible to frame three main

---

[1] The encyclical Laudato Si by Pope Francis (2015) shares such criticism with convincing simplicity: 'when media and the digital world become omnipresent, their influence can stop people from learning how to live wisely, to think deeply and to love generously. In this context, the great sages of the past run the risk of going unheard amid the noise and distractions of an information overload. Efforts need to be made to help these media become sources of new cultural progress for humanity and not a threat to our deepest riches. True wisdom, as the fruit of self-examination, dialogue and generous encounter between persons, is not acquired by a mere accumulation of data which eventually leads to overload and confusion, a sort of mental pollution. Real relationships with others, with all the challenges they entail, now tend to be replaced by a type of Internet communication which enables us to choose or eliminate relationships at whim, thus giving rise to a new type of contrived emotion which has more to do with devices and displays than with other people and with nature. Today's media do enable us to communicate and to share our knowledge and affections. Yet at times they also shield us from direct contact with the pain, the fears and the joys of others and the complexity of their personal experiences. For this reason, we should be concerned that, alongside the exciting possibilities offered by these media, a deep and melancholic dissatisfaction with interpersonal relations, or a harmful sense of isolation, can also arise' (47).

forms of knowledge:

1. *Receptive knowledge*. It occurs when we absorb information from expert textbooks or reputable Internet resources. Toward digitally acquired information we often adopt an attitude of default trust because, overloaded by information, it is what we need. As we strive to filter the good (true) information from the bad (false)—and do so quickly, mechanically and reliably – we need to be receptive. Receptive thought, however, is also non-reflective and it works under the surface of conscious attention, allowing us to more easily select and choose sources that validate our existing opinions (*confirmation bias*).

2. *Responsible knowledge*. The second is the sort of knowing we value whenever possessing reasons or experience matters. You can give some evidence on a matter, you know many facts, but still lack comprehension.

3. *Understanding*. And the third is different still—it is the sort of knowing we expect of our most creative experts—even if those experts are more intuitive than discursive in their abilities. Understanding is what we have when we know not only the "what" but the "why." Understanding involves knowing not just the facts, but also the how or why something is the case. You understand when you see not just the isolated aspects, but how those aspects hang together. In turn, understanding is a form of knowing that involves grasping relationships—the network, or parts and whole.

Thus, there is a very sharp difference between knowing facts and knowing how. Knowing-which, then, is another fundamental element of understanding. Understanding increases the ability to ask the right questions. Experts (those who understand a given subject best) are often able to increase their understanding even further because they have the ability to know which question they should ask in the face of new information. "Downloading knowledge" other people have acquired via experience isn't the same as having that experience yourself, isn't the same as personal trial and error and creative adaptation in the face of circumstance, so it does not give mastery, and not even the understanding that comes with it. Even worse, somehow paradoxically, it is often the case that the ones who really have a certain expertise likely feel the so-called "*impostor syndrome*", which means that even if they are experts they are full of doubts about their knowledge, while those

who really don't understand often have an "*overconfidence effect*", which means that one is so confident about its own knowledge. Similarly, as the philosopher Bertrand Russell famously and concisely argued in a never confirmed quotation: "the whole problem with the world is that fools and fanatics are always so certain of themselves, and wise people so full of doubts."

As a result, conspiracy theories find a natural medium for their diffusion in social media and, also, for their validation through search engines. Certainly, the growth of knowledge fostered by information societies with the unprecedented acceleration of scientific and technological progress has exposed society to an increasing level of complexity to explain reality and its phenomena. Narratives grounded on conspiracy theories tend to reduce the complexity of reality and are able to contain the uncertainty they generate (Bessi et al., 2015). Literature addresses the study of social dynamics on socio-technical systems from social contagion up to social reinforcement. The extent of the phenomenon is dramatic. Human imagination, paranoia and self-deception in order to reduce complexity, indeed, are extremely powerful and influencing a significant amount of population. Consider that 4% of U.S. population, 12 millions of people endorsed the notion that 'shape-shifting reptilian people control our world by taking on human form and gaining power' (Oksman, 2016). Other more widespread beliefs nourished by suspicious, ill-informed thinking raise greater concerns. Notably, the no-vax movement that moved from theoretical accusations to practical consequences, namely no-vax-supporting politicians. Inevitably conspiracy theory results, in the dissemination of distrust for media, politics and science. While the formers may even deserve to be highly contested in order to be reformed, the latest should remain democratic societies' safety net.

To conclude, the internet is – to some extent – causing a collective deception: users consume a lot of information but rarely internalize it and act upon it. Google-knowing, while a basis for understanding, is not itself the same as understanding because it is not a creative act. Also, social media, as a novel source of information, are designed in a way that makes us comfortable with our pre-existing beliefs while making it more difficult to confront dissonant information. Reaching understanding is something that one must do for oneself as involves an element of individual cognitive achievement, without delegating to the medium, and it requires an open and constructive confrontation with the otherness, preferably offline. A true understanding needs a reflexive element and, most

importantly, reliable sources of information. Apparently, internet and social media do not openly sustain these preconditions.

1.2.5 The Risks of Misinformation

The World Economic Forum (WEF) listed massive digital misinformation as one of the main risks for modern society (Howell, 2013). Internet and in particular social media are increasingly weaponised, especially for propaganda reasons. Certainly, it is not a new phenomenon. Yet, the manipulation of our perception of the world is taking place on previously unimaginable scales of time, space and intentionality (EU commission, 2018). There is indeed plentiful of information wars among several actors — state or non-state political actors, for-profit actors, media, citizens, individually or groups — to the detriment of citizens and societies at large. The risk of harm includes threats to democratic processes, including electoral integrity, and to democratic values that shape public policies in a variety of sectors, such as health, science, finance and more. Bot, "fake-news" and micro-targeting are the primary weapons. This has become evident in the last years, especially after Brexit and Trump's election, with the potential role played by Russia and other actors. Fake-news, intended as "fabricated information that mimics news media content in form but not in organizational process or intent" (Lazer et al., 2018), became a buzzword - not to be confused with overlapping concepts like: *misinformation* – "which is information that is false, but not created with the intention of causing harm" –*disinformation* – "which is information that is false and deliberately created to harm a person, social group, organization or country" – and *malinformation* – "which is information that is based on reality, used to inflict harm on a person, organization or country" (Wardle and Derakhshan, p.20, 2017).

 Social Media in particular, and the Internet more generally, faces accusations of deteriorating civil debate to the point that facts and truth are now fertile ground for dispute and subjectivity, while trust with experts and authorities have decreased. This is particularly relevant on political issues. Politics no longer functions through "rational discourse", if ever. Traditionally, politicians bend the truth, in the sense that if they are found lying they would provide justifications. This tradition is deteriorating. More recently, the distrust with media led many people to accept politicians (prominently the president of United States Donald Trump) to refuse to justify their actions and words. On

the one hand, what is occurring is that voters are treated as malleable – aiming at the visceral and emotional level – rather than convinced with reason. On the other hand, the optimism about the coexistence of democracy and expertise may be well displaced. If the "post-truth era" starts to implode current knowledge structures, then it is not democratization anymore, rather it is the path to authoritarianism.

Epistemic competition is as much about which truth can be relevant and about which claims can be considered true or false, and these choices have important consequences. Civil society not only needs a common currency to exchange money but they also need a common currency to exchange reasons. This could be what philosopher David Hume called a "common point of view"— reasons that can move some universal principle of the human frame, and touch a string, to which all mankind have an accord and symphony (in Lynch, 2015). The point, as Lynch (Ibid) argues, is not that we should all agree but that if we cannot agree on what counts as evidence, on our epistemic principles, then we are not playing by the same rules. When you can't agree on your principles of evidence and rationality, you can't agree on the facts and, in turn, you can hardly agree on what to do in the face of the facts. This just increases tribalization, and so on and on in a recurring loop. That is the major risk of a post-truth era.

The paradigm of the extent of information war is that it has grown to such proportions that the idea of "fact-checking" is treated with suspicion. The main reason is that our politics are governed by passion. Reasons are indeed far less influential than intuition and emotion. Trying to draw reasons to convince our cultural opponents usually fail. At the same time, the Internet makes it considerably easier to manipulate people's expectations, partly due to its relative anonymity. Partly because the expectation-setting context is increasingly difficult to track. The Internet is a bloodied, chaotic battleground for the truth wars. And as we can't extricate ourselves from our perceptions, how can we decide which of those perceptions reflect things as they really are and which are the products of our own minds?

Social Media could undermine the importance of objectivity. But what we mean by objective truths? Truths are objective when what makes them true isn't just up to us, when they aren't constructed. Yet, they can always be in part constructed. Rather, one *can be objective*. According to Lynch (2015), a person is objective, or has an objective attitude, to the extent to which he or she is sensitive to reason. Being sensitive to reason involves

an awareness of individual limitation, being alert to the fact that beliefs may not stem from reason but prejudice, a viewpoint alone. Objectivity requires open-mindedness. Being objective, or sensitive to reasons, however, is no guarantee of certainty. A simpler explanation is that the apparent 'post-truth' can be explained as the result of the manipulations of trust brought by echo chambers, where certain communities have a vastly divergent set of trusted authorities. Members of an echo chamber are not irrational but misinformed about where to place their trust (Nguyen, 2018). What is, then, the role of personalization in all this?

2. Online Personalization and Its Challenges

Algorithms play an increasingly important role in the media; smart tools that assist journalists in producing their stories, fully automated production of news stories (so-called robot journalism), audience analytics that inform editorial board decisions and, of course, recommender systems and personalization systems. They are more than a technology. They can actually affect the news ecosystem on at least three levels: news production and distribution, individual users, the broader media ecology and society more generally. They in fact need to be intended as 'a socio-economic construct, that is, as technologies that are embedded in organisations with their own goals, values and fundamental freedoms, and that mediate and impact interactions with the human/economic/social environment in which they are functioning' (CoE, 2020, p. 3). While still in their early days, these tools can have profound implications for the overall quality and diversity of the information offer provided by the news media. They also promise several opportunities; for example by featuring societally relevant stories from the long tail, identify the most relevant stories for a user, taking into account the context of news use, cater to different reading habits related to our different social roles, enabling new forms of investigative data-driven journalism, enhancing diversity on various levels, and fostering a deeper understanding between societal groups.

In this chapter, we focus specifically on the role of online personalization in social media, mainly driven by recommender systems but, also, by design choices that can affect the algorithmic outputs as well as users' behaviors. As such, we try to shed lights on what is personalization, how can we define it, how it works, what are its limitations, challenges and opportunities, how users think of and know about it, how platforms are currently using it, and, ultimately, how future personalization might look like. By doing this, we implicitly apply the analytical toolbox of Science and Technology Studies to analyze online personalization as a complex socio-technical construct, reviewing the most up-to-date literature from an interdisciplinary perspective, and highlighting challenges and opportunities for designers, platforms, users, regulators and, more generally, democratic societies.

2.1 What is Online Personalization?

Nowadays, personalization has become ubiquitous and it underlies every social media engagement, many apps and platforms. Yet, personalization remains an ambiguous and under-examined concept, lacking of consensus on its essential characteristics. With personalization it can be broadly intended "data-driven profiling of consumers or citizens to deliver a customised or personalized service, advert or legal response. Personal data is interpreted in conjunction with big data, using sophisticated algorithms, to create a picture of who someone is based on who they were - their past preferences, activities, networks and behaviours - in order to make a future-oriented prediction of what they might like (i.e. which film), what might persuade them (i.e. which ad), how they might act (i.e. commit a crime or succeed in a job) or what they might deserve (i.e. promotion or public housing)." (Kohl et al., 2019). More concretely, in everyday Internet usage most people usually benefit from such personalization when Netflix suggests them a movie, Youtube a video, Facebook a stream of information, Spotify new music and Amazon related products. Still, what is intended with personalization need to be further discussed.

Basically, how does personalization work? Normally, it relies on *data* used to build *profiles* with which one can infer certain predictions through more or less sophisticated *algorithms*. As humans naturally categorize, generalize and classify the world around them to reduce complexity, algorithms can indeed be programmed to automatically process information in similar ways. Profiling practices, thus, create, discover or construct knowledge from large sets of data from a variety of sources that then are used to make or inform decisions. In the book *Profiling the European Citizen*, Hildebrandt and Gutwirth (2008) gave a concise definition of profiling: "Profiling is a matter of pattern recognition, which is comparable to categorization, generalization and stereotyping (p.365)." Of course, profiling can occur in a range of contexts and for a variety of purposes. In the case of social media personalization, profiling makes or informs decisions (presumed preferences) that personalize a user's media environment (e.g. content selection and ordering). With large media providers no longer serving a gatekeeping function, the consumption of information turned on the choices of individual users and algorithms, mediated by profiling.

Clearly, these complex predictive processes of decision-guidance and decision-making raise serious theoretical issues, offering limitless benign opportunities as well as

dystopian realities. Before to analyze these, in order to better grasp the outcomes and potential undesirable and unintended consequences of personalization, it is worth provide an overview on the major notions, processes and theories involved in these techniques. Some general questions are addressed, such as: what are the technical steps for personalization? How data is collected and how inferences drawn? What kind of data and decisions are inferred? What is (or what could be) the individual role in such process? Ultimately, what might look personalization in the future?

2.1.1 Personalization in Social Media

Another useful definition of personalization is provided by Thurman et al. (2013); personalization is defined as "a form of user-to-system interactivity that uses a set of technological features to adapt the content, delivery, and arrangement of a communication to individual users' explicitly registered and/or implicitly determined preferences" (p.2). This definition highlights a fundamental issue: personalization can be explicit or implicit, that is, it can depend on user's requests and/or user's behavioural data (which is more often created unknowingly and unwittingly). Personalization can thus be based on the individual autonomy of choice or on the algorithmic/platform delegation to infer one's personal preferences. In more detail, personalization can be:

- *Explicit*: the user customizes the information source himself. The user can register his interests or other personal information before the personalization starts. The user can also rate topics of interest. Explicit user information collection will allow the user to know that the personalization is taking place and he can tailor it to his needs. Yet, one problem with explicit feedback is that it places an additional burden on the user. Because of this, or because of privacy concerns, the user may not choose to participate. It is also known that users may not accurately report their own interests or demographic data and that the user's interests may change over time.

- *Implicit*: the system determines what the user is interested in through various factors, including web usage mining (i.e., previous interaction with the system such as click throughs, browsing history, previous queries, time spent reading information about a product), IP address, cookies, etc. Implicit user information collection, on the other hand, does not require any additional intervention by the user during the process of constructing profiles. It also automatically updates as

the user interacts with the system. One drawback of implicit feedback techniques is that they can typically only capture positive feedback. When a user clicks on an item or views a page, it seems reasonable to assume that this indicates some user interest in the item. It is not clear, however, when a user fails to examine some data item whether it is an indication of disinterest.

Both implicit and explicit personalization increased dramatically in the last years, though many websites have acted to make passive forms of personalization the fastest growing forms (Thurman, 2011). This is concerning and need further discussion (as will highlight in Chapter 4).
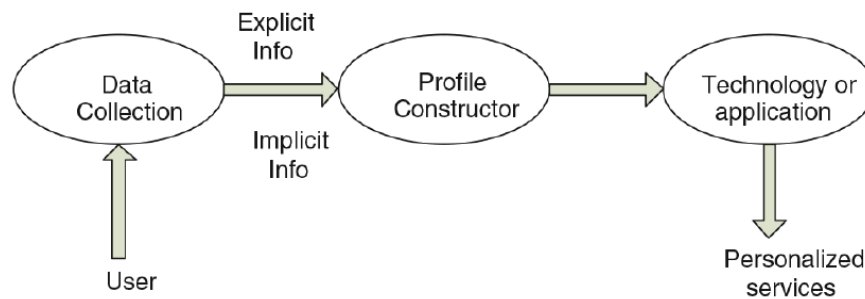
Figure 2. Basic User profile construction for personalization (from Bozdag and van de Poel, 2013).

Function	Type	Examples
Prioritization	General and search engines, meta search engines, semantic search engines, questions & answers services	Google, Bing, Baidu; image search; social media; Quora; Ask.com
Classification	Reputation systems, news scoring, credit scoring, social scoring	Ebay, Uber, Airbnb; Reddit, Digg; CreditKarma; Klout
Association	Predicting developments and trends	ScoreAhit, Music Xray, Google Flu Trends
Filtering	Spam filters, child protection filters, recommender systems, news aggregators	Norton; Net Nanny; Spotify, Netflix; Facebook Newsfeed

Figure 3. A General Taxonomy of the Main Algorithms Employed in Online Recommendations (from Lepri et al., 2017).

2.1.2 Recommender Systems

Among the diverse algorithmic functions, in this research we are more interested in information filtering and ranking. Specifically, in recommender systems (RSs). With RSs is intended computer-based data-driven software tools and techniques that provide suggestions for items to be of use to a user (Ricci et al., 2015). These systems represent the most important engines of personalization. They have emerged in the early '90s and in 2006 they attracted much attention with the famous Netflix prize to improve movies' recommendations with hybrid RSs. Then, the widespread of social media and smartphones providing a lot of contextual information such as time, place, the emotion of people and groups opened up a new avenue of recommendation known as contextual RS.

In essence, RSs try to analyze how a user values certain products or services and then predict what the user will be interested in next. "Item" is the general term used to denote what the system recommends to users. RSs are primarily directed towards individuals who lack sufficient personal experience or competence to evaluate the potentially overwhelming number of alternative items that a platform, for example, may offer. To carry out this task every RS employs some information retrieval techniques such as Machine learning and different algorithms like logistic regression, decision tree, association rule learning, cluster, Bayesian networks and support vector machine, etc. Peoples from various disciplines such as data mining, information retrieval, knowledge discovery, artificial intelligence, approximation theory, forecasting theory, information security and privacy, and business and marketing have contributed extensively with diverse research approaches.

Personalization and recommender systems are often used interchangeably. Actually, a recommendation is a form of personalization, but personalization is not a form of recommendation. For example, YouTube might suggest related videos based on previous viewing habits, this is a recommendation based on what other YouTube users also watched. A restaurant, however, might suggest a table by the window based on a previous booking you have made. This is personalization, as it is based on the specific habits of the individual and not a broad algorithm. The more you know about a person, not just their viewing habits, the better. In other words, a recommendation is often built around items (e.g. collaborative-filtering), whereas personalization is built around individuals

(i.e. individual profiling). There is of course much overlap, and the more informed and well designed a recommendation engine becomes, more on that in a moment, the closer to personalization such methods become. Also, personalization is depending on explicit inputs and design choices and, therefore, it is much more adjustable.

This distinction is important also for the subject choice of this research and need to be clarified. In fact, one could have chosen RSs as the paradigm of personalization and potentially a more specific object of analysis to undertake. Instead, personalization is a more complex and comprehensive phenomenon of interest; it is indeed more specific to the individuals, it is composed of design choices that affect and influence recommender systems and, eventually, it sustains a marketing rhetoric that is more attractive than RSs themselves. Personalization is thus a broader subject that ultimately shape the whole personal information environment online.

Going back to RSs, recommendation mechanism typically does not use an explicit query but rather 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 x or y). Then the recommendation mechanism presents to the user one or more descriptions of objects (e.g., books, people, movies) that may be of interest.

Generally, RSs can be divided into three main types: content-based (also called "semantic filtering"), collaborative (also called "social filtering"), and hybrid (most of the RSs) (Ricci et al., 2015):

- *Content-based filtering*. If recommendations are done solely by analyzing the associations between the user's past choices and the descriptions of new objects.
- *Collaborative filtering*. It automates the process of "word-of-mouth" recommendations: items are recommended to a user based upon values assigned by other people with similar taste. The system determines which users have similar taste via standard formulas for computing statistical correlations.
- *Hybrid*. The most common form of RSs today are hybrid recommenders, which combine features from both content-based and collaborative systems or other elements such as demographics, communities or editorial selections.

What we are interested in this paper-based thesis, however, is, in particular, news[2] RSs – that are essentially used by media organizations, news aggregators and social media platforms – and, more generally, social media's personalization. They are in fact increasingly relevant. According to recent findings (Newman, 2016), almost 70 percent of online news users surveyed across 37 different markets worldwide identified distributed forms of discovery as their main way of accessing and finding news online, with search and social media being by far the most influential factors, followed by aggregators, email and mobile alerts.

With the advent of the Internet, traditional models of information and news production, distribution and consumption have indeed radically changed. In order to provide relevant information to users, the epochal transition from information scarcity to information abundance brought the need for a balance between a pull and a push approach or, in other words, between user requests and algorithmic delegation (Thurman and Schifferes, 2012). This legitimized mainstream social media's gatekeeping role. The gatekeeping process is studied extensively by multiple disciplines, including media studies, sociology and management in order to address traditional media bias, namely how certain events are being treated more *newsworthy* than others and how institutions or influential individuals determine which information passes to the receivers (i.e. what are the values or moral perspective from which select the news). In this context, some major changes occurred: a) the increasing role of the audience in which users can determine what is newsworthy through social networks, b) the changing role of the journalist, from a gatekeeper to a gatewatcher, and c) the delegation to increasingly sophisticated algorithms (i.e. news RSs).

As such, scholars argued that by using online services, the audience can exert a greater control over news selection and can focus on issues that they consider more relevant, which in turn empowers audiences and erodes the degree of editorial influence over the public's issue agenda. Others even argued that the gatekeeping role performed by the traditional media becomes irrelevant and gates are disappearing. Importantly, in general user interests become more important than content quality or social significance (DeVito,

---

[2] While in this thesis the word 'information' is used more generically, news is considered as a specific kind of information with respect to its unique function in society, namely "information that should enable citizens to know what is regarded as important, contested, or an issue of public interest that should be deliberated upon (p.5)" (Bernstein et al., 2020).

2017). This kind of rhetoric makes data-driven personalization very appealing in our hyper-individualized societies. As Kohl et al. (2019) notes, terminology offers some insights on the acceptability of the phenomenon. In the market place where predictive profiling is welcomed by users and offered by businesses as added value, it is referred to as 'personalization', 'customisation', 'optimisation' or 'smart' technology.

To provide a paradigmatic example of a RS, Facebook uses a collaborative filtering called *Edgerank*, which adds a weight to produced user stories (i.e. links, images, comments) and relationships between people. 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 contact list. All stories produced by user X can be completely hidden in user Y's newsfeed, without the knowledge of both users. In fact, social media platforms are not just communication companies that uniquely shape social networks and their dynamics processes, but also large advertising agencies driven by profit (Fuchs, 2015). The underlying automated 'engagement maximization' of mainstream platforms brings several concerns that we will highlight throughout the thesis to which they add also other challenges.

Some of the most relevant technical and methodological challenges that arise from RSs – especially relevant to the context of social media's personalization – are summarized as follows:

- *Data collection*. Above all, data collection, validity and reliability of data are some of the most pressing issues. For example, there is the need to acknowledge the gap between behavioral data and actual user experience (Loecherbach and Trilling, 2019).
- *Time*. The performance of most RSs evolves over time and an increasing familiarity of the users with the system also changes how they interact with it (Ricci et al., 2015). Especially in news RSs, it is also very important as well as challenging to focus on timeliness (e.g. while an old book might still be interesting, the news of yesterday might be not).
- *Design*. For example, objectively acknowledged practical challenge is the small screen size of mobile phones, which nowadays constitute most of the accesses to Internet platforms. This becomes ever more relevant when designers want to

empower users with tools to use during navigation.

- *Interdisciplinary Debate*. There is a fragmentation of the debate, partly because of the newness of the technology, partly because of the proprietary algorithms and privacy issues involved (Milano et al., 2019).

- *User's RSs Control*. Balancing implicit and explicit personalization is fundamental both for RSs quality and, also, for users' autonomy and, eventually, dignity.

- *Trust*. A significant issue is the so-called *black-box problem* which occurs when the system is opaque towards the end-user, causing decreased levels of confidence. The potential of the recommendation is thus diminished leading to unsuccessful attempts of recommendation. By providing the transparent reasons to the user, such as ratings on an item provided by similar users, the users are motivated to place their trust and confidence in the recommendation as well as the recommended items.

- *Generalization of Results*. Fundamentally, the assessment of mainstream private social media RSs is restricted for outside researchers. Given that each user is served a personalized selection of recommendations, it is difficult for any individual observer to make generalizable conclusions about the performance of the system as a whole. As we only know our own news feeds we can only guess what others are seeing (Leersen, 2020).

- *Performance Assessment*. Most studies often remain in an evaluation setting where RSs are compared without user interaction. To better understand the risks of RSs, it might be also meaningful to confront participants with 'failed' algorithms that lock them in filter bubbles or ignore their wishes (Loecherbach and Trilling, 2019).

- *Metrics*. It is paramount the development of metrics to evaluate the quality of RSs (e.g. accuracy, diversity, serendipity etc.).

- *Values*. In general, it is fundamental to understand how to identify, operationalize, inscribe and assess values into the system (Helberger, 2019).

- *Diversity*. Similarly, identifying, measuring and optimising for diverse recommendations is acknowledged as relevant. For example, the *over-specialization problem*, as the risk to provide too many narrow and similar items

or the *long-tail problem*, as the risk that if an item is initially not well-rated or not rated at all then over the time it will perish from the recommendation catalog, despite its potential quality.

To conclude, it should be highlighted that RSs are highly context-dependent, that is, they depend on often changing parameters: dataset (what the space of options is), values, (what counts as a good recommendation) and assessment (how RS's performance can be evaluated) (Milano et al., 2019). Furthermore, RSs serve multiple goals and a purely user-centered approach does not allow all such goals to enter into the design so that the perspectives and utilities of multiple stakeholders (e.g. companies, investors, news outlet and, broadly, society) must be taken into account. For example, a multistakeholder approach could be applied based on a utility-oriented approach so as to integrate concerns from all actors involved, not only users, nor platforms, (Abdollahpouri et al., 2017). Such approach is also useful to conceptualize explicitly the impact that RSs have on different levels.

2.1.3 Future Personalization

To date almost all of the RSs have been designed for sellers, producers, and service providers; i.e., they are designed to attract potential customers. Future RSs will not only be limited to business, but they will have a much greater impact on our daily life. If we imagine the future of personalization, we can quite confidently argue that algorithms with powerful design (e.g. smart assistants) will guide us and make ever more sophisticated and reliable decisions for us and about us. Imaginative efforts of this emerging worlds have been done from literature to academic books and papers. It is often told that algorithms will wake up us softly with our most preferred music, make the perfect breakfast, the one that our physician recommended us, choose the street we take, the people we meet, the books and news we read, similarly to nowadays, but even suggesting the most important choices of life – what to study, whether to accept a job offer or whether to marry. These kind of *cognitive outsourcing* are provided by AI-driven algorithmic decision-making and decision-guidance. They promise us to enhance our lives preserving our time and energy and nudging us towards healthier behaviors and better decisions. The ethical questions of this potential development will be discussed in Section 4.2.6. Here only general expectation towards the development of personalization is done.

The ideal RS should be like someone who knows us better than we know ourselves. They should sense our need and will suggest instinctively, even if we do not express explicitly. Few other fields like the Internet of Things (IoT), Internet of Everything (IoE), Artificial Intelligence (AI), Cognitive Computing, Affective Computing (cognitive science and psychology), etc., will play a significant role in future RSs. Clearly, there is a fair amount of unpredictability in communication technology development, preventing precise predictions regarding what future implementations of personalization will look like. Yet, it is expected the rising of Ambient Intelligence related to the Internet of Things. This construct offers a vision in which automatic smart online and offline environments and devices interact with each other, taking an unprecedented number of decisions for us and about us to cater to our inferred preferences, representing a new paradigm in the construction of knowledge (Hildebrandt and Koops, 2010).

The major differentiating point of future RSs will be the intelligent use of ubiquitous data. Data will be captured, assessed and analyzed literally from anywhere and for anything. Though tackling the ever-increasing data will be a great challenge for the future RS designers because the current algorithms may not be straightforwardly scalable to cope up the unforeseen amount of data. The major challenge in designing the RSs will be to restrict them to be irritatingly interfering in the personalized domain. On the other hand, existing RSs are generally generic in nature. They don't really reflect each user's individualized taste. So, the future RS designers have to set a balance between the two, i.e., it will neither be too generic nor too intrusive.

2.2 Personalization and Its Challenges

Personalization occurs largely beyond the control of users as it is based on implicit personalization—behavioural data collected from subconscious activity—rather than on deliberate and expressed preferences. Even though there is a fair amount of unpredictability in communication technology development, it seems that implicit personalization is likely to become a default choice in future personalization (Thurman and Schifferes 2012). It is indeed imagined the idea of Ambient Intelligence related to the Internet of Things. This construct offers a vision in which automatic smart online and offline environments and devices interact with each other, taking an unprecedented number of decisions for us and about us in order to cater to our inferred preferences. This

may represent a new paradigm in the construction of knowledge (Hildebrandt and Koops, 2010). This often uninformed delegation can lead to several undesirable unintended consequences. For example, it increases political selective exposure—people's ability to see information that conforms to their pre-existing ideas and priorities— as it makes information avoidance less psychologically costly (Dylko et al., 2018).

Personalization algorithms are novel and complex entities, and they require novel solutions and public discussions around their development and regulation. There are, in fact, three main intertwined 'paradoxes' that highlight the complexities to fairly govern personalization. Firstly, the 'personalization paradox', that is, a trade-off between privacy and personalization accuracy. Secondly, and strictly related, the 'privacy paradox', that is, the infamous users' inconsistent will to protect their own privacy. And thirdly, the 'paradox of choice', that is, the more choices users have, the more easily they rely on simple-to-use personalization tools.[3] In other words, in order for personalization systems to provide a 'better service', users are surveilled and datafied to extract ever more value. Even if they disagree – as they more often do – they do not proactively react. Even in cases where the users are provided with more agency, they are unlikely to take advantage. Relatedly, there are arise other specific-related concerns, such as reductionism – in particular various related biases – information asymmetry, privacy, diversity and autonomy concerns.

2.2.1 Reductionism

Profiling technologies that allow personalization create a kind of knowledge that is inherently probabilistic. Moreover, data used are derived only from what is actually observed and are a small subset of possibly observed behaviors. Then, this narrow subset of recorded behaviors must be converted into a digital (database) representation. In this process, Information may be lost. Thus, three key characteristics of personalization arise:

1. Personalization is highly behaviorist its assumptions. Behaviorism seeks to eliminate theoretical causal mechanisms of human behavior (beliefs, intentions, goals, etc.) and focus instead on what can be observed measured and recorded. As such, behaviors most amenable to measurement tend to be recorded.

---

[3] A paradigmatic example is Youtube's recommendations which already drive more than 70% of the time spent in the video sharing platform and 90% of the 'related content' is indeed personalized.
See https: //www.cnet.com/news/youtube-ces-2018-neal-mohan and https://youtube.tracking.exposed/data.

2. Personalization focuses on predicting a very narrow set of possible behaviors, often limited by the context of the application. This can make it seem more powerful and accurate than it really is, especially when predictive performance is evaluated.

3. Personalization uses data not only from the individual user, but also from other users. This is clear in the case of RSs using social network data (e.g. collaborative filtering). This questions the extent to which personalization are actually personalized.

As a matter of fact, profiling technologies that fuel personalization cannot produce or detect a sense of self (Hildebrandt, 2009). The problem is that they can influence the individual's sense of self, especially in the long-run daily social media information consumption. The underlined – somehow inevitable – reductionism is concerning because it can undermine the sense and development of the self and may ultimately serve only the interests of service providers.

On the one hand, profiling algorithms are naturally stuck in the past, and this can steer individuals to conform to the status quo of their past actions chosen by past selves. Individuals might even start to want what is recommended to them without even recognizing it, in a sort of *self-fulfilling prophecy*. This could perpetuate existing inequalities and other pathologies, while threatening the foundational microeconomics principle of *preference formation*. Individuals, in fact, have different "orders" of preferences: "first-order preference" is expressed in how we behave in the moment that a stimulus or temptation affects our consciousness. In contrast, "second-order preference" is the choice we make for ourselves upon further reflection, generally separated from the immediate temptation. Think of the snake offering the forbidden fruit in the Garden of Eden (first-order preference: eat fruit), and the initial resistance (second-order preference: don't eat it). Online behavior seems to enable much more fluid expression of first-order preferences. In this way, personalization algorithms can shrink the individual "aspirational self" (or second-order preferences). In the longer term, this process might undermine individuals' capacity to develop and maintain an integral *sense of self*, that is preconditional to autonomy.

On the other hand, under cover of "personalizing" information, services and products we are witnessing a colonization of public space by a hypertrophied private sphere. Social

structures and social relationships, which are historically contingent, find themselves reduced to an algorithm and/or a computer application. They imply the reproduction and naturalization of diverse power relations and forms of social inequality. Such algorithmic governance is indifferent to individuals, insofar as it simply focuses on and controls our "statistical doubles", in other words combinations of correlations, produced automatically and using big data, themselves constituted or collected "by default". In general, engineers' approach can take two directions: build more advanced techniques, or simplify the user interface to convert the recorded human activity in a defined set of possibilities, reducing complexity. Modern advancements in technology use both (Harris, 2019). In many cases, in fact, personalization relies on the comparability or even sameness (by some simplified categories) of the user with others (e.g. collaborative filtering). Thus, personalzsation paradoxically denies individual uniqueness through a "smart" homogenization that negotiates the diversity of humankind (Yeung, 2018). Through constructing, manipulating, and reinforcing such homogenizing categories, data-driven personalization thus works on the premise of "divide et impera"—that is, segment individuals into markets and sell as much as possible personalized ads.
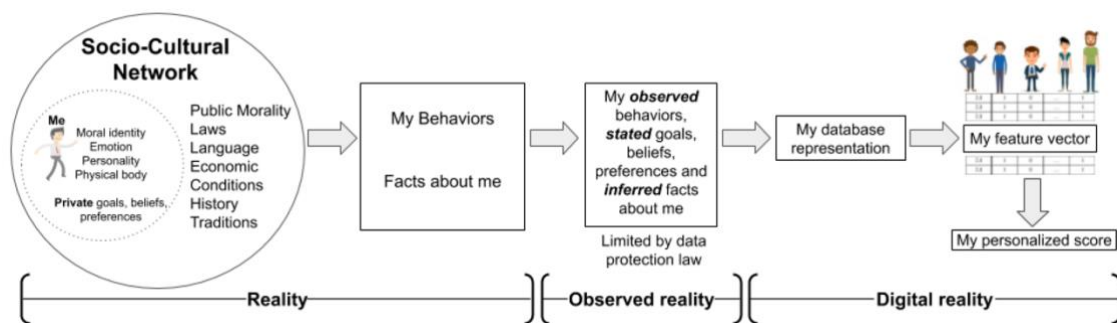
Figure 4. From humanistic concepts of the person to personalized scores: ML operationalization requires narrowing a person embedded in social and cultural space to a feature vector embedded in feature space (from Greene and Shmueli, 2019).

Eventually, here it lies difficult philosophical questions such as: which information should constitute me as a person? And, ultimately, what is a person? These questions are problematic as "person" is often justified through a particular worldview (Greene and Shmueli, 2019); It could mean "human animal," "moral agent," "rational, self-conscious subject" etc.. Yet, Western thought has historically focused on two general aspects of the person, one of which is revealed through the etymology of the word itself. "Person" comes

from the Latin *persona*, which derives from the Ancient Greek word for a type of mask worn by dramatic actors. For the Ancient Greeks, the idea of a person was inherently connected to context and role. One's persona was a specific kind of self-identity that was public, socially defined, and varied depending on context. Over time, the outward-facing Greek conception was complemented by a Christian emphasis on self-reflection and awareness, resulting in a focus on the human capacity for rational introspection. As a result, Western thought has generally viewed the person as consisting in two co-existing inner and outer domains.

The inner aspects of individuals, however, are inaccessible through observational methods. Individual, subjective input from the same person is therefore crucial in measuring the level of personalization. Hildebrandt (2017) also argues that the 'multiplicity of the self' is not computable and that such incomputable nature of the self should be somehow protected. In fact, not only no personal profile can ever entirely identify an individual, but there are always many—and sometimes radically different—ways of computing the same person.

Understanding the person in fact requires description at multiple levels. Personality psychologists define personality as a an organized set of relatively enduring psychological traits that guide a person's interactions with her "intrapsychic, physical, and social environments." These traits can be expressed as the "Big 5" dimensions of personality: Openness to New Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (known as OCEAN), and have been shown to have varying levels of predictive power. Consider that just dozens of Facebook 'likes' can reveal highly accurate correlations on personality (Youyu et al., 2015).

Yet, identity/self/personality are complex and intertwined concepts that cannot be easily reduced. Personality and social psychologists highlight the power of the cultural context in defining identity. For identity theorists in sociology and psychology, the self is fluid and occupies multiple social roles (identity theory) or group identities (social identity theory) that coexist and vary over time. The dynamism of the self-concept is reflected in its numerous sub-components or self-representations: the past, present, and future self; the ideal, "ought," actual, possible, and undesired self. These views ultimately highlight the processual, dynamic, and emergent nature of the self that, in mainstream social media, is substantially reduced to a cold, reduced, opaque commercial profiling.

2.2.2 Privacy

One of the primary challenges for RSs is, consequently, privacy. Algorithmic biases, information asymmetries and reductionism, all raise the question of how to better protect privacy. Furthermore, there is an obvious trade-off between privacy and accuracy of recommendations, thus making privacy fundamentally at stake in personalization practices. Yet, privacy is not a clear-cut concept and, therefore, we need to problematize it providing an overview of its meaning as well as a brief history of its debate.

Privacy is a very broad concept encompassing "freedom of thought, control over one's body, solitude in one's home, control over information about oneself, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations." (Solove 2002, p.1088). The lack of a consensus on the substance of privacy notwithstanding, there have been considerable efforts to conceptualize the notion. Drawing from the privacy literature, Bygrave (2001) posits that privacy as a concept could be classified into four types:

- Privacy in terms of 'non-interference' advanced by Samuel Warren and Louis Brandeis in their 1890 *Harvard Law Review* article titled 'The Right to be Let Alone'.

- Privacy in terms of 'information control,' popularized by Alan Westin's definition of privacy (1967) as 'the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others'.

- Privacy in terms of 'limited accessibility' popularized by Ruth Gavison which presupposes a condition consisting of three elements: 'secrecy' ('the extent to which we are known to others'); 'solitude' ('the extent to which others have physical access to us'); and 'anonymity' ('the extent to which we are the subject of others' attention').

- Privacy in terms of those aspects of persons' lives that are 'intimate' or 'sensitive', defined by Julie Inness 'the state of possessing control over a realm of intimate decisions, which includes decisions about intimate access, intimate information, and intimate actions'.

More generally, Fuster (2014) observes that privacy in the sense of serving the realization

of individuals' own lives has furthermore been coupled with the notion of human dignity; the basic assumption being that it is inherent to the human condition to develop freely without any form of external dictation (from the community or the State).

Despite the variety of research and attempts to categorize it, privacy has been principally regarded in terms of control and autonomy over (personal) information. Over the years, the paradigm has evolved from the control of the communication of information to others to an emphasis on the individual's control over their information. The control-based definition of privacy has remained largely unchanged since then. Yet, it has been criticized for several reasons, among them, for being "counterintuitive" and for relying on the false assumption that information is something that can be controlled, but also because privacy as control rarely aligns with the reality of people's practice with their personal information and it ignores information asymmetries and 'bounded rationality' (Nissenbaum, 2011; Acquisti et al., 2015). Also, it is argued that privacy can promote an individualistic approach that can go against the common good and that can even be used to carry out illegal or antisocial activities (e.g. financial privacy).

Despite inevitable criticism to the above privacy conceptions, there emerges a significant theoretical tension: the classical Western conception of privacy based on the idea of individual autonomy as opposed to *relational autonomy*, based instead on a contextualized perspective of autonomy. The concept of individual autonomy has been questioned by many critics from many other areas other than privacy studies, including psychiatry, moral philosophy, medicine, etc.. Instead, the term relational privacy (Bannerman 2019) is used to broadly categorize all conceptions of privacy that rest on relational autonomy and hence adopt a contextualized (or networked) perspective. Feminist theory, in particular, rejects the classic Kantian concept of autonomy as being atomistic, advocating relational autonomy as an alternative. In different forms, that perspective is also embraced by Japanese culture, Confucianism, Buddhism, Indian culture and the African Ubuntu philosophy (Reviglio and Alunge, 2020).

The application of big data and machine learning further challenges the existence of a clear information boundary that can be used to build a clear definition of personal information. For example, it is unclear if information inferred by machine learning algorithms about a person falls within the categorization of personal information. In theory, if all data has a potential to impact people, then all data could be considered

personal data (Purtova 2018). In addition, algorithms can often conflict with existing human understanding of personal information, and it may not even be clear how they can be interpreted. Therefore, the focus on individual control over data already seems problematic and too narrow and may eventually need to be supplemented by an interpretation of privacy which takes account of broader data uses, interests and practices.

Two aspects of the understanding of relational autonomy from this perspective are especially relevant. First, relational autonomy recognizes the social embeddedness of the person; the person, instead of being deprived of any social relations, is influenced by all sorts of social contingencies. Second, relational autonomy conceives of autonomy as a competency that is contextualized and, most importantly, that it can be developed.

A relational autonomy perspective supports a networked conception of privacy. Network privacy conceives the context and norm as not only co-constructed by participants but they also evolve over time (Nissenbaum 2011). Thus, it considers the possibility that individuals may interpret contexts differently, that contexts may be destabilized or collapse, or that other people may have control over the context in ways that are beyond the purview of the individual (e.g. surveillance, information leakage, or data-mining). Therefore, networked privacy requires individuals to not only refer to some external rules/regulations but also to construct on their own a clear expectation of privacy.

Relatedly, there is a more concrete development of the concept in what has been defined as 'group privacy' (Taylor et al., 2016; Mittelstadt, 2017). Algorithms can create ad-hoc and temporary groups to which none of the existing interpretations of privacy can be applied; for example, a group can be "female students in the city of Nairobi who received a university grant", or 'private sector workers in Accra below the age of 40 who own a car'. Being identified as a member of this group could drive a variety of automated decision-making with harmful or even beneficial effects for individual members, such as a preferential rate for health insurance. The existence of such groups is essentially informational, and these informational traits are only meaningful to algorithms on the group level when a multitude of individuals' information are considered together.

The above considerations suggest that (a right to) group privacy is a materialization of the concepts of relational as well as networked privacy. The view that the individual self is considered constrained from the start of its own existence by social factors (relational privacy) and that people are not always in control of information regarding them

(networked privacy) necessitates the consideration of top-to-bottom norms which emphasize the dignity of humanity (as a community or group) over the dignity of individuals. The challenge is indeed how to translate such principles into practice, for example due to the widespread focus on Artificial Intelligence ethics tools development on the need to protect the individual over the collective (Morley et al., 2019).

Milano et al. (2019) stress how privacy risks occurs in at least four stages in personalization systems: 1) when data are collected or shared without the user's explicit consent. 2) once data sets are stored, there is the further risk that they may be leaked to external agents, or become subject to de-anonymization attempts. 3) clearly, privacy concerns can also arise at the stage of inferences that the system can (enable one to) draw from the data. Thus, sensitive information can be inferred without the user knowledge of the nature of such inferences. And finally, 4) at the stage of collaborative filtering the system can construct a model of the user based on the data it has gathered on other users' interactions, that is, the issue of group privacy (that it will be discussed in 4.2.5) (Taylor et al., 2016).

Current and general solutions to these challenges fall into three broad categories (Friedman et al., 2015):

- *Privacy-enhancing architectures* that aim to mitigate privacy risks by storing user data in separate and decentralised databases, to minimise the risk of leaks.
- *Algorithmic solutions* focus on using encryption to minimise the risk that user data could be exploited by external agents for unwarranted purposes.
- *Policy approaches*, including GDPR legislation, introduce explicit guidelines and sanctions to regulate data collection, use, and storage.

In general, the issue of privacy would deserve much more space for reasoning. What is needed to know in this context, however, is that most of the platforms are opaque and much of the privacy users may benefit is not actually enacted. These issues will be variously discussed throughout the thesis.

2.2.3 Information Asymmetry

Nowadays a bunch of platforms dominate the social media landscape. Basically, these are Facebook, Youtube and Twitter. It is well-known in the literature about competition in social networks that an equilibrium can sustain only a small number of such

intermediaries and a concentrated market structure is thus expected. Economies of scale in the production of news lead to monopolies (Lovink and Rasch, 2013; Helberger, 2014). The current media landscape indeed reflects these dynamics and brings to a couple of social networks an unimaginable power. This is because of *network effects* which occur when the value of a platform to any user increases exponentially with the number of already present users. Indeed, the history of alternative media is a history of enormous challenges, as the consequence of a political economy that limits the possibilities for civil society because hearing alternative voices is a matter of money and political resources that afford visibility (Fuchs and Marisol, 2015).

Today social media policies, technical design choices and business models actually serve as a form of *privatized governance* directly enacting rights and regulating the flow of information online and, in doing so, promote or constrain civil liberties (DeNardis and Hackl, 2015). In particular, Facebook, at the time of writing, has 2.6 billion users and is the most pervasive and powerful intermediary of the Internet. Its unique relevance depends also on the fact that regarding personalization, search engines like Google, are, to some extent, less problematic because they tend to deliver more one-size-fits-all services, while the diversity in social media communication is significantly lower than that of search and also inter-personal communication (Nikolov et al., 2015).

The system of news updating in Facebook – the NewsFeed – has a growing central role in the global information flow. On average, it selects 200 posts on 2.000, therefore hiding 90% of social media information – circa 1 hour per day per user (Reviglio and Agosti, 2020). It is based on EdgeRank, a complex algorithm which constantly changes its outputs. By doing so, some values are embedded in the process but these, unfortunately, are secret and assessable only indirectly. The understanding of algorithms in the academic world is indeed weak due to two major factors: the impermanence of Internet technologies and the black-boxed nature of most influential algorithms. The first means that by nature the Internet is transient, rapidly changing at a rate that usually outpaces the research process. Secondly, the black-boxed nature of algorithms occurs not only to protect trade secrets but also prevent malicious hacking and gaming of the system (Pasquale, 2015). Of course, these features make not only the research but also the policy focus on algorithms very challenging.

In sum, mainstream Social Media, notably Facebook, are supranational entities,

divorced from the ratification of a social contract (not a mere 'terms and conditions' contract). Even though they are becoming our being-in-the-world, their interests conflict with the opportunities that social networks might entail. Their policy and profit-driven model, indeed, negatively affect the externalities of personalization. Facebook, for example, offers several ways to personalize one's own newsfeed. Yet, it does not provide clear and effective tools to opt-out, though in theory they could radically empower users through education and tools.

2.2.4 Various Biases

In traditional media, newspaper editors select some of the messages produced by journalist to make news. Algorithms used in web services (such as ranking algorithm in a search engine, or news feed algorithm in a social network) make similar decisions. The design of these algorithms is affected by choices made by designers, i.e., which factors to include in the algorithm, and how to weigh them. The most known biases are the following:

- *Popularity bias.* To serve majority interests, information intermediaries often include popularity metric in their ranking algorithm. A search algorithm, for instance, can give more weight to information coming from popular websites, to support majority interests and values. As a result, users may have troubles finding the less popular and smaller sites.

- *Third-party influence/maniplulation.* Because the information filtering are automated, they might be manipulated by activities from third parties. This could occur in several ways, from clickbait and Search Engine Optimization (SEO) techniques to 'bots' that game social media's metrics in order to further the spread of potentially problematic content.

- *Product/service prioritization.* In the last decade, the EU received complaints that claimed how their traffic drop after Google began promoting its own services above conventional search results. Studies showed that Google and Bing search engines both reference their own content in its first results position when no other engine does (Bozdag, 2013). Facebook was also criticized for favoring the products of its partners.

- *Novelty bias.* In Google search engine, the number of years a domain name is

registered has an impact on search ranking; domain names that exist for a period of time are preferred over newly registered ones. In Facebook, the longer a status update has been out there, the less weight it carries. A news item is prioritized over an old item. This might, for instance, lead companies to post updates when their audience is most likely to be online and using Facebook.

- *Other kind of biases.* The algorithm can also prioritize certain types of information over others. For instance, the engagement maximization model of mainstream SNS tends to prioritize ephemeral content to durable one, short videos (snippets) to long ones and snapshots (casual images) to written text as they better lock-in users in their "walled-garden" (Derakhshan 2016).

What are, then, all these 'biases'? Are they really biases or, instead, they are companies' values and interests inscribed into algorithms and design? Indeed, with algorithmic bias, especially in the context of AI, it is often referred to clearly unfair and discriminatory outputs. In this context, however, they are more of a series of editorial choices to proactively sustain commercial interests rather than unpredictable or even socially problematic outcomes.

2.3.5 Diversity

The effect of personalization on the diversity of information and news diets (or 'media diversity') has mostly been discussed negatively, assuming that such systems limit the breadth of viewpoints and topics. The major risk is the creation of "informational bubbles": filter bubbles (Pariser, 2011) and echo chambers (Sunstein, 2017), two sides of the same token. The first is a kind of cultural and ideological bubble in which an individual continues to see, listen and read what reinforces its opinions and interests. The latter is a group situation where established information, ideas, and beliefs are uncritically spread and amplified, while dissenting views are ignored. The crucial difference is that the former may not depend on user's autonomy and awareness (Bodzag, 2011) – therefore is mainly caused by technological affordances – while the latter pre-existed digital age and thus it is primarily social. There is limited evidence of the existence of these phenomena. Also, they are poorly defined and, in fact, are used more as generalizing (thus limiting) metaphors (Bruns, 2019). Similarly, the concept of media diversity is not really clear. More recently, a systematic literature review of this concept has shed more light

(Loecherbach et al., 2020).

From an individual perspective, personalization might reduce opportunities to self-determine and could negatively affect information finding by reducing the exposure to alternative points of view in the 'marketplace of ideas' (Pariser, 2011; Sunstein, 2017) and, more generally, to serendipitous encounters (Reviglio, 2019a). The consequences may be various: from the limitation of personal creativity to a reduction in the ability to build productive social capital (in particular *weak ties*, see Granovetter, 1977). More generally, the main consequence to provide a 'too familiar world' is that our online life would eventually shift from an intersubjective to a subjective one (Keymolen ,2016).

From a collective perspective, personalization can weaken media pluralism, making people more politically polarized and vulnerable to censorship and propaganda or, better, to self-censorship and self-propaganda (Sunstein, 2017a). This, in turn, would help to spread misinformation (Vicario et al., 2016) and erode interpersonal trust (Keymolen, 2016). As such, the audience becomes increasingly fragmented and people – especially the less skilled and literate – become more politically polarized and vulnerable to censorship and propaganda (see Figure 5).

Critics, however, argue that these are moral panics, and that personalization might instead foster the cultivation of "expert citizens" with stronger group identities (Harambam et al., 2018). Another prominent risk, however, remains growing the 'digital divide' and 'cultural divide' (Fabris, 2018) or 'epistemic inequality' (Lynch, 2016). Certain privileged group of users, that have higher (digital) literacy, are able to fruitfully manage personalization. Instead, a larger group of users would risk to be exposed only to a minimum, qualitatively inferior, range of information (see, for example, Figure 6). Also, the wealthier the social networks, the more the benefits of personalization, and vice versa.

In practice, research on diversity, as well as more generally personalization and its challenges, is often contradictory, ambiguous and, thus, unreliable (Zuiderveen Borgesius et al., 2016; Bodo et al., 2017; Tucker et al., 2018). The consequent risks of social media usage (Verduyn et al., 2017) and filter bubbles (Zuiderveen Borgesius et al., 2016), in fact, are very hard to prove and, eventually, to mitigate. Most research is ultimately inconclusive because it is generally survey-based, or is correlational or based on a small or unsatisfactory sample. In light of the fast-changing media landscape, many studies become rapidly outdated. This contributes to the current crisis with regard to the study of

algorithms (Bodo et al., 2017). They are "black-boxed" (Pasquale, 2015) which means their functioning is opaque and their interpretability[4] may not even be clear to engineers (Albanie et al., 2017).

While insights on the main causes and risks of personalization have been currently understood (Tucker et al., 2018), we still lack evidence with regard to the extent of their consequences. There is not, in other words, anticipatory knowledge to assess personalization systems as a technology. Also, information filtering processes take place not only at the technological level (e.g., affordances) but also on the individual (e.g., selective exposure) and the social (e.g., sharing practices). And given the vast heterogeneity of users, causes and effects of personalization vary widely. Furthermore, concerns are growing because of the rise of persuasive technologies and the ability of Big-Data to nudge individuals and effectively deceive them (Yeung, 2018, Zarsky, 2019).
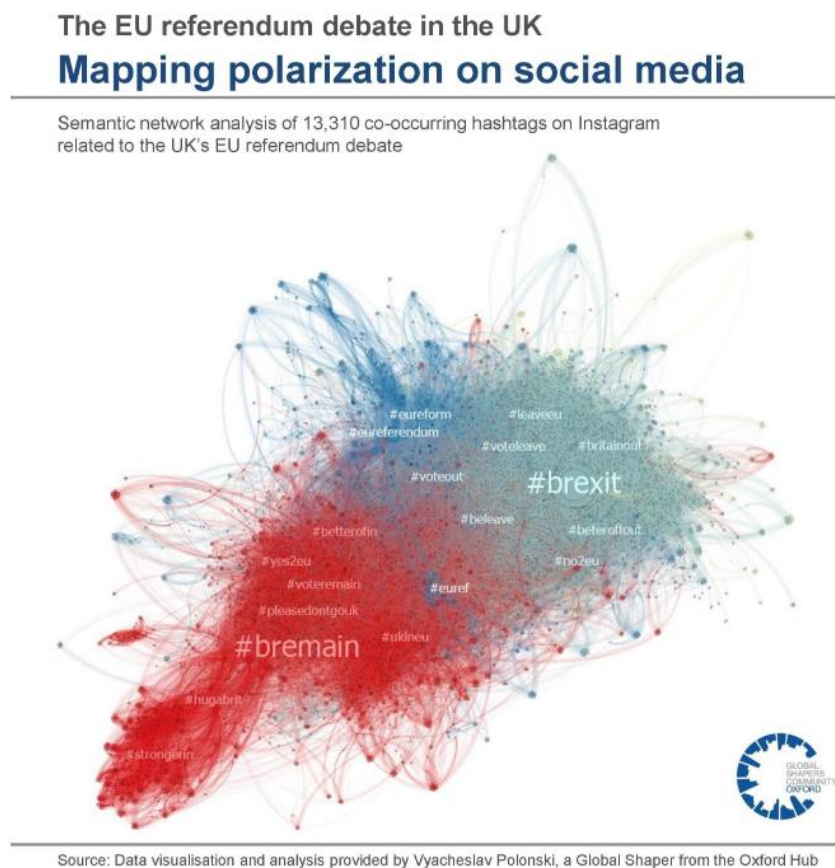
Figure 5. Semantic network analysis of the hashtags utilized on Instagram before 2016 UK referendum.

---

[4] Interpretability is the degree to which a human can understand the cause of a decision or consistently predict the model's result.
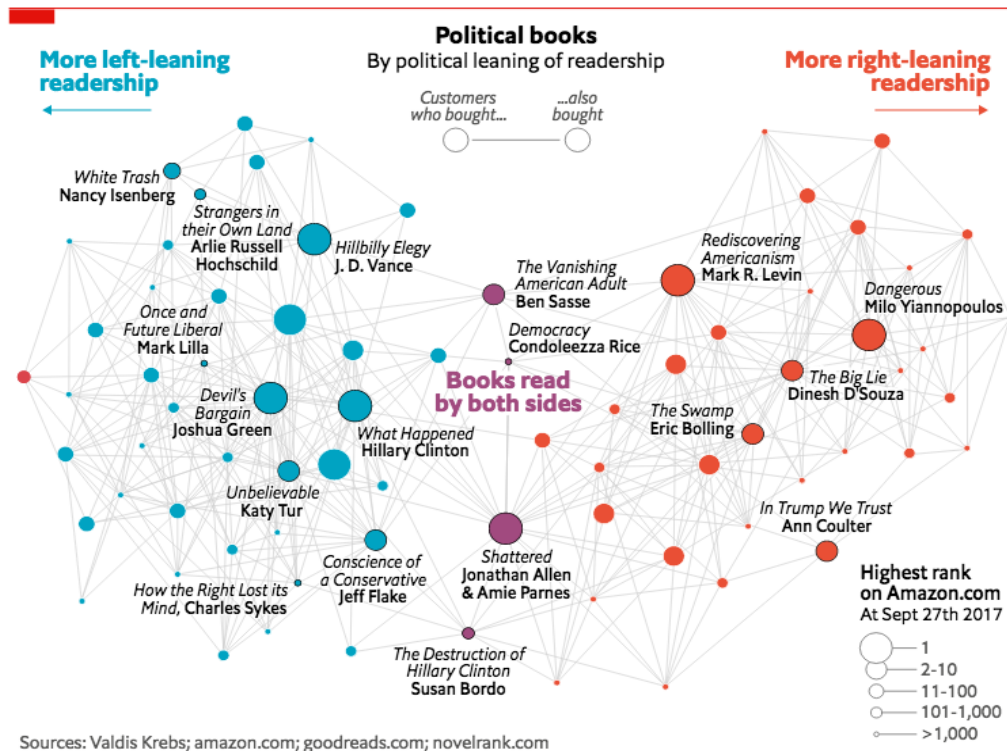
Figure 6. Political books relations for Amazon's recommender system (from The Economist, 2017)

Thus, key issues remain unsolved: to what extent personalization is detrimental, to what extent are information intermediaries complicit and responsible, whether they should be the target of a policy focus and, eventually, what kind of interventions might be pursued. As a reaction to the new media environment, media ethicists and policy-makers began to discuss how to maintain an exposure to diverse information. From a theoretical perspective, in fact, all the models of democracy consider the consequences of filter bubbles problematic for particular different reasons (see Figure 7) (Bozdag and Van den Hoven, 2015; Helberger, 2019).

More generally, media pluralism is achieved when users autonomously enjoy a diverse media diet. The exposure to various sources and content, however, seems insufficient to ensure actual experience of media diversity. Even if media diversity online is generally more than in traditional media, such exposure does not always end up in an actual *experience of diversity* (Hoffmann et al., 2015). Cognitive and affective factors that drive Internet users must also be considered. This requires to employ a user-centric perspective and extend beyond the assumption that supply diversity equals experience of diversity,

and that diversity of sources equals diversity of content. In fact, in the very first place the question is: what is actually meant with media diversity? There are indeed several interpretations.

Model of democracy	Norms	Criticism of the filter bubble
Liberal	Awareness of available preferences Self-determination Autonomy Adaptive preferences Free media Respect human dignity	User is unaware of the availability of options User is restrained and individual liberty is curtailed The media is not free, it serves the interests of certain parties (e.g. advertisers) Powers are not separated (advertiser and the information provider are the same)
Deliberative	Discover facts, perspectives and disagreements Determine common interests Construct identity by self-discovery Refine arguments and provide better epistemic justifications Consensus Respect towards each other's opinions A collective spirit Free and equal participants Rationality	Epistemic quality of information suffers Civic discourse is undermined No need to have better epistemic justifications Respect for other opinions is decreased Legitimacy is more difficult to achieve. There is a loss of a sense of an informational commons Communication suffers as gaining mutual understanding and sense-making is undermined
Republican and contestatory	Freedom from domination by oppressors Contest matters effectively Be aware of the oppressors	Diminishes one's ability to contest Diminishes one's awareness of the oppressors and their potentially manipulative interventions
Agonistic/inclusive political communication	Conflict rather than consensus Passions rather than rationality Struggle rather than agreement Inclusion: Measures must be taken to explicitly include the representation of social groups, relatively small minorities, or socially or economically disadvantaged ones Measures must be taken so that antagonism is transformed into agonism	The adversary becomes the enemy The minorities are excluded from the democratic process, their voices are lost

Figure 7. Models of Democracy and Design Criteria (from Bozdag and van den Hoven, 2015)

Media diversity is a rich and complex value that can be conceptualized in many different ways, and its interpretation differs significantly per discipline (Loecherbach et al., 2020; Bernstein et al., 2020). It is not as simple as the maximization of diversity is the most optimal solution. In media studies, diversity might be translated as "minority voices having equal access in the media" or "the degree which the media relates to the society in such a way to reflect the distribution of opinion as it appears in the population". Similarly, in communication studies, diversity is also an important concept. The freedom of media, a multiplicity of opinions and the good of society are inextricably intertwined. A theory of media diversity states that we establish and preserve conditions that provide many alternative voices, regardless of intrinsic merit or truth, with the condition that they

emerge from those whom society is supposed to benefit its individual members and constituent groups. Whereas in Computer Science literature, diversity is rather fuzzier. It can be defined as "variety in the products offered by the system", "helping user find items he cannot easily find himself" or "identifying a list of items that are dissimilar with each other, but nonetheless relevant to the user's interests". While media studies are analyzing this ethical value in detail, almost all scholars of search engine diversity seem to be limiting their understanding of "bias" and "diversity" to popularity bias (Bozdag, 2015).

While many scholars from different disciplines agree that media diversity is an important value that we should include in the design of institutions, policies and online services, this value is often reduced to a single definition, such as "source diversity", or "hearing the opinion of the other side". There is a need for a more detailed normative conceptualization of the value media diversity (Loecherbach et al., 2020; Bernstein et al., 2020). Only then, perhaps, can we translate this complex value into design requirements of information intermediaries and move towards a solution. To do this, all the following dimensions of diversity must be taken into account: source diversity and content diversity, exposure diversity and diversity experience (namely, what are you exposed to and what you actually consume) (Hoffman et al. 2015), individual diversity and aggregate diversity, viewpoint diversity (Bozdag, 2015), and the inclusion of minorities and 'long-tail' Information. Standardized measures of these core "sub-dimensions" of diversity are highly needed (Loecherbach et al., 2020). Moreover, there are in fact never-ending tensions in diverse design proposals, for example 'diversity' and 'trustworthiness' on the one hand, and 'non-discrimination' and 'neutrality' on the other. While the former requires RSs to prioritize certain content, the latter could arguably prohibit the drawing of such distinctions. Also, democracy theories have other conflicting values (Helberger, 2019). Thus, the issue of diversity cannot be 'solved' objectively or definitively, rather, throughout more interdisciplinary experimentations and collaborations it may be possible to approximate a more diverse media consumption.

2.2.5.1 The Case of Facebook

According to De Vito et al. (2015), NewsFeed's values in descending order of influence seem to be: friend relationships, explicitly expressed user interests, prior user engagement, implicitly expressed user preferences, post age, platform priorities, page

relationships, negatively expressed preferences, and content quality. Friend relationships, in particular, moderate how the other values will be expressed. Friends on social media, however, tend to be ideologically clustered and this potentially "places the lens through which the News Feed algorithm filters all other values firmly within your personal bubble to begin with" (Ibid, p.15). Indeed, homophily is the primary driver of content diffusion, especially misinformation and conspiracy theories, and one frequent result is the formation of homogeneous, polarized clusters (Del Vicario et al., 2016).

Facebook filters in fact combine to decrease exposure to ideologically challenging news from a random baseline by more than 25% for conservative users, and close to 50% for liberal users (Ibid). Another study by Facebook (Bakhsy et al., 2015), shows that liberals tend to be connected to fewer friends who share information from the other side of the political spectrum, compared to their conservative counterparts: 35% of the hard content shared by conservatives' friends are cross-cutting, compared to 24% for liberals. As a consequence, it seems that social media are not a space apt to belief change.

Facebook performed poorly concerning the overall level of deliberative quality and performed worse than the other platforms (Esau et al., 2017). Deliberative democracy is indeed one of the most influential theoretical concepts in the ongoing debate about the relationship between democracy and Internet technology. This was measured through certain criteria (rationality, reciprocity, respect, and constructiveness). For example, regarding users' comments platforms differ in moderation, asynchronicity, availability of information, and level of focus in topic definition.

Facebook is currently designed in such a fashion that promotes our natural homophilic tendencies of preferring to be with likeminded people. Indeed, users are usually more likely to share articles with which they agree, and the content they share tend to be confined only within their echo chamber (Quattrociocchi et al., 2016). In general, the more a polarized user is active, the more the user has friends with similar profiles. Therefore, users belonging to different communities tend not to interact. Then, if any users begin even to filter out uncongenial news, then its attitudes may harden. In fact, the filter bubble thesis is supported particularly in the context of online radicalization (Sunstein, 2007).

Nevertheless, not only algorithms but also design choices can deeply affect users' behaviours. Facebook, indeed, provides some personalization options (see Figure 8).

Instead, personalization opt-outs are not taken explicitly into consideration. Actually, the only evident way to look for a more generalized or de-personalized experience in Facebook is to select the option "Most recent" stories in the Newsfeed. However, this is still a personalized option, nor to say that the setting spontaneously reset itself to "Top Stories" after a certain amount of time. Furthermore, Facebook's relation with content suppliers poses other challenges for access of a diverse choice. For instance, *Instant Articles,* the news feature of Facebook in which users are able to open an article directly on the app newsfeed. This feature gives Facebook unprecedented power and control over the news market, since already 66% of users get news on Facebook (Gottfried and Shearer, 2016). However, many have warned that this system is favouring big publishers. As a matter of fact, exposure is partly related to how much money the publisher put into advertisement, and this gives large media group a comparative advantage, threatening the openness and the variety of online news found on Facebook.
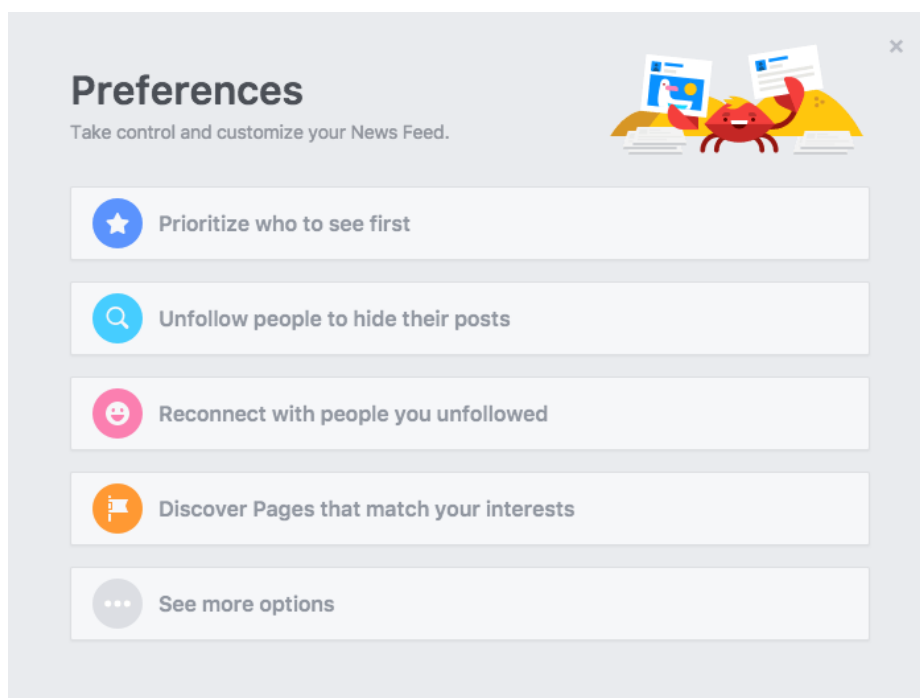
Figure 8. Newsfeed management (screenshot, May 2017)

2.2.5.2 Bursting Filter bubbles and Echo Chambers?

A related fundamental question to tackle the potential decrease in diversity is whether it is possible to 'dispel' or 'weaken' filter bubbles and echo chambers, and how that would be possible. In fact, some political radical groups have claimed that they use exposure to

opposition to strengthen their opinions further. Simply exposing radicalized people to alternative points of view would not undo these phenomena (O'Hara and Stevens, 2015). Quattrociocchi and Sunstein (2016) recently showed that even debunking is ignored by 99.98% of conspiracy Facebook users. Other studies, however, suggest that exposures to opposite views tend to increase political tolerance and awareness of the legitimate arguments underlying opposing perspectives. Indeed, individuals would strengthen their opinions after better understanding the other political side and, in some cases, formulating new opinions on issues that were against their former beliefs (Semaan et al., 2014).

There are several inter-connected natural human cognitive proclivities that reduce the diversity, relevance and quality of the information consumed, as well as to dismiss challenging and unfamiliar information, among these:

- *'Homophily'* is the principle that a contact between similar people occurs at a higher rate than among dissimilar people. The pervasive fact of homophily means that cultural, behavioral, genetic, or material information that flows through networks will tend to be localized. Homophily implies that distance in terms of social characteristics translates into network distance, the number of relationships through which a piece of information must travel to connect two individuals. Perhaps the most basic source of homophily is space. We are more likely to have contact with those who are closer to us in geographic location than those who are distant.[5]

- *'Selective Exposure'* means that people expose themselves to external stimuli in a selective way. When referred to the area of mass communication, this means that people choose certain types of media content and avoid other types. It is important in understanding the effects of mass communication because it is our common understanding that people can only be influenced by media messages to which they actually expose themselves. Therefore, the selective exposure concept emphasizes the active role of the individual in the selection of media content.[6]

---

[5] Zipf (1949) stated the principle as a matter of effort: It takes more energy to connect to those who are far away than those who are readily available. Yet, this ought not be the case on the web.

[6] Research of this phenomenon is undertaken in the fields of both psychology and communication studies. Basically, there are two major trends in this research. Most studies focus on factors that lead to selective exposure or that mediate this process, whereas other studies deal with the consequences of selective exposure to information processing. The

- *'Confirmation Bias'* is our tendency to actively seek information which confirms our preexisting beliefs and disregard information which might contradict it. This tendency originates from our desire to organize our world meaningfully and avoid feeling cognitive dissonance.

- *'Cognitive Dissonance'* refers to a situation involving conflicting attitudes, beliefs or behaviors. This produces a feeling of discomfort leading to an alteration in one of the attitudes, beliefs or behaviors to reduce the discomfort and restore balance, etc. Festinger's (1957) cognitive dissonance theory suggests that we have an inner drive to hold all our attitudes and beliefs in harmony and avoid disharmony (or dissonance). This is known as the principle of cognitive consistency.

Moreover, there are other explanations for group polarization and the development of echo chambers that can be summarized in three main explanations as following (Sunstein, 2009):

1. *Informational influences and persuasive arguments.* The first explanation emphasizes the role of persuasive arguments and information. In any group with some initial inclination, the views of most people in the group will inevitably be skewed in the direction of that inclination. For example, most people in a group believe that immigration should be stopped. As a statistical matter, the arguments favoring that initial position will be more numerous than the counter-arguments. Individuals will have heard of some, but not all, of the arguments that emerge from group deliberation. As a result of hearing the various arguments, deliberation will lead people toward a more extreme point in line with what group members initially believed.

2. *Social influences, namely conformism.* Sometimes people's publicly stated views are, more or less, a function of how they want to present themselves. People

---

selection processes have also been examined in different contexts, such as in political or online communication. More recently, however, has been showed that people in echo chambers often have full access to outside sources of information. They may regularly read – but do not accept – mainstream and liberal news sources. They are isolated, then, not by selective exposure, but by changes in who they accept as authorities, experts and trusted sources. They hear, but dismiss, outside voices. Knowledge, in the end, mainly depends on trusting long chains of experts.

usually want to be perceived favorably by other group members, so that some will adjust their positions at least slightly in the direction of the dominant position. Group polarization is indeed heightened when people have a sense of shared identity. People could polarize because they are attempting to conform to the position that they see as typical within their own group. If their group's identity is especially salient or important, the ingroup norms "are likely to become more extreme so as to be more clearly differentiated from outgroup norms, and the within-group polarization will be enhanced.[7]

3. The third explanation stresses the close links between *confidence, extremism, and corroboration* by others. If people lack confidence, they will tend toward the middle, and hence avoid the extremes. As people gain confidence, they usually are willing to become more extreme in their beliefs. People's opinions have been shown to become more extreme simply because their views have been corroborated, and because they have become more confident after learning that others share their views. A final observation is dedicated to the fundamental role of *identity*. In fact, will significantly increase if people think of themselves as part of a group having a shared identity and a degree of solidarity. In short, if identity is shared, persuasive arguments are likely to be still more persuasive.

Sunstein (2017) relates *political polarization* to *audience fragmentation* and analyses the risks to which they might lead. Firstly, these might have the effect of decrease (the quality of) deliberation. In fact, such phenomena can sustain so-called "partysm", which is a kind of "visceral, automatic dislike of people of the opposing political party" (p.10). It is not as bad as it is racism but, to some extent, similar to it, and nowadays it exceeds it. Sunstein provides a paradigmatic example. In 1960, just 5% of Republicans and 4% of Democrats said that they would feel "displeased" if their child married outside their political party

---

[7] This also relates to a general theory of public opinion involving a "spiral of silence" in which people with minority positions silence themselves, potentially excising those positions from society over time (Noell-Neumann, 1984). It happens in authoritarian societies as well as in democracies. Sometimes it is also a good thing; people who believe that the sun goes around the earth, or that slavery was a reasonable idea, may end up self-silencing. Interestingly, Internet gave opportunities to these groups to express themselves. As Umberto Eco famously argued: "Social media gives legions of idiots the right to speak when they once only spoke at a bar after a glass of wine, without harming the community. Then they were quickly silenced, but now they have the same right to speak as a Nobel Prize winner. (Nicoletti, 2015)" Yet, the spiral of silence is not always benign as group discussions are affected and, as a result, become more extreme. Indeed, group polarization occurs merely on the basis of *exposure* to the views of others, so that discussion is not even necessary.

(Ibid). By 2010, those numbers had reached 49 and 33%, respectively—far higher than the percentage of people who would be "displeased" if their child married someone with a different skin colour. Such a political polarization threatens progress. In fact, it is also reported that in the U.S. often politicians would like to vote in tune with the other parties but they actually risk to lose votes so that they cannot easily reach a political compromise. Even if the self-segregation involves only a small part of the electorate, they can be highly influential, not least because of the intensity of their beliefs.

Another major premise is that the truthfulness of misinformation and disinformation is often of secondary importance to polarized groups: people are attracted by disinformation not necessarily because they consider it truthful, but rather because it aligns with their worldview and it gives them a sense of community and identity.

Accordingly, the consequences to draw are the following. On the one hand, fact-checking corrections will not always be taken seriously by people who ideologically do not align with its findings. Also, because fact-checkers are often from opposed ideological and societal groups, these organizations may suffer a lack of trust among their target audience. Indeed, highlighting false information may even be counterproductive, since the excessive attention (and uptake by other media organizations) can also increase its reach. It has even been argued that being confronted with corrections could actually further strengthen the original beliefs.

On the other hand, those in most need of "education" may not be the ones who are actually reached and most receptive to it. Therefore, educating citizens may not always work in combating disinformation in society at large because the issue is not just cognitive – i.e., based on a faulty understanding of social media or personalization – but is also driven by deeper and more complex cultural and societal shifts related to the loss of trust in mainstream media, science, and other knowledge institutions.

To conclude, the capacity to counteract such phenomena is deeply controversial. Echo chambers pre-existed digital age: people naturally tend to selective exposure and homophily. Nonetheless, RSs, design choices and, more generally, personalization, have a large role in potentially worsening the phenomenon, as well as in preventing its aggravation (Hoffmann et al., 2015). To some extent, social media are indeed responsible (and at times they could even be complicit) in the potential growth of filter bubbles and echo chambers. Though there is still no sufficient empirical evidence that warrants any

strong worries at present, in case personalization technology would improve, groups polarization would increase and personalized news content would become people's main information source, problems for democracy could indeed arise (Zuiderveen Borgesius et al., 2016). Yet, indirect effects to democracy there might be already present.

2.2.6 Autonomy

In order to access social media platforms, users usually have to give up the data they generate. The profit-generating business of selling the resultant meta-data to advertisers is what supports these "free" platforms. Clearly, Internet users are not simply or even primarily hedonistic consumers to persuade or even manipulate in order to maximize data extraction and profit generation. Rather users are also often producers, citizens and unpaid workers who generate economic value for these platforms through their "digital labour" (Scholz, 2012). This is the status quo of the current data-driven society, in which sociality is turned into economic value. Users occupy a subordinate position without any reasonable possibility to opt out rather than agree or leave their social networks.

  As we will analyze in detail in Chapter 4, GDPR is likely to be the most advanced regulation in the data protection landscape as it recognizes data subjects full control of their data. However, it confers limited protection against automated decision-making in light of algorithms' manipulative potentials (Reviglio, 2019b). Of course, GDPR is granting novel rights for data subjects and duties for data controller. Along with the e-Privacy regulation draft, the GDPR actually strengthens 'data consumer protection'. Users can indeed decide whether to enter into a contract, be informed, access the data generated, receive information about the logic involved and opt not to be subject to automated decision-making based solely on automated processing. However, the right to an explanation (Art. 15) and the right to non-discrimination (Art. 22) – the most relevant in this context – are highly disputable. Also, the major problem lies in the apparent belief of lawyers (as well as designers) that input control alone is sufficient to alleviate any systemic concerns of personalization algorithms (Kohl et al., 2019). Data protection law remains indeed crucial – because personalization requires updated target profiles – but currently no legal framework addresses explicitly the outputs of personalization algorithms.

The approach of user-empowerment often relies on informed consent, but this might not be an optimal solution because opt-in approaches lead to uncertainty and context dependence. People in fact cannot be counted on to navigate the complex trade-offs involving privacy self-management (Acquisti et al., 2015). Most people in fact neither read nor understand online privacy policies. These actually have also two inherently contradictory goals; to be understandable to consumers – which requires simplicity and brevity – and say something meaningful about how data is processed – which is complicated and requires a lot of details. Moreover, dark patterns are often employed during terms of conditions and privacy updates (Moen et al., 2018). Although user-centered solutions may better foster the transparency of personalization algorithms, for example, they have significant shortcomings: they shift the responsibility and accountability for the protection of rights and utility to the users, and this usually result in inefficiency such as insufficient privacy protection.

The problem is that today's Internet (Alter, 2017) —especially social media (Deibert, 2019)— is addictive by design. This threatens individual autonomy (Gal, 2017; Zarsky, 2019). Human behavior can indeed be manipulated by priming and conditioning, using rewards and punishments. Even algorithms can autonomously explore manipulative strategies that can be detrimental to users (Albanie et al., 2017). For instance, they can use 'A/B testing' to experiment with various messages until they find the versions that best exploit individuals' vulnerabilities (Calo, 2013).[8]

Manipulation and deception become easier thanks to affective computing (or 'emotional AI'), captology—the study of computers as persuasive technologies (Fogg et al., 2002)—and the emergence of psychographic techniques focusing on demographic characteristics and 'affect detection techniques', along with diverse types of data such as location-based tracking, real-time data, or keyboard usage. As already mentioned, dozens of Facebook Likes can reveal useful and highly accurate correlations (Youyu et al., 2015). Implicated data include what the European General Data Protection Regulation (GDPR) classifies as specially protected data: sexual orientation, ethnicity, religious and political

---

[8] In an A/B test, the experimenter sets up two experiences: 'A,' the control, is usually the current system and considered the "champion," and 'B,' the treatment, is a modification that attempts to improve something—the 'challenger.' Users are randomly assigned to the experiences, and key metrics are computed and compared." From https://hbr.org/2017/09/the-surprising-power-of-online-experiments.

views, personality traits, intelligence, use of addictive substances, parental separation, age, gender and, perhaps most importantly, emotions.[9]

Eventually, persuasive techniques affect individuals' self-control, self-esteem and even self-determination. These can indeed stimulate users in a powerfully subconscious and hormonal way. Facebook's infamous, large-scale emotional contagion experiment exemplifies this point (Kramer et al., 2014), showing how mainstream social media can affect emotions and exploit vulnerabilities in human psychology. Compulsion loops are also found more broadly in a wide range of social media, and especially online games (Deibert, 2019). Research suggests that such loops may work via "variable-rate reinforcement"[10] in which rewards are delivered unpredictably. This unpredictability affects the brain's dopamine pathways in ways that magnify rewards.

Also, design facets intentionally trigger dopamine rushes or other emotional highs, stimulate popularity contest or implicit social obligations (Kidron et al., 2018)—and use a whole host of additional brain hacks, probably many of them not even publicly known. Among those known, moral outrage can be exploited to increase engagement. NYU psychology researchers found that each word of moral outrage added to a tweet raises the retweet rate by 17% (Harris, 2019). At times, users may also be captured in a spiral of ever more extreme, conspiratorial content—also known as the "rabbit hole effect."[11]

Despite apparently increasing transparency along with the resultant efforts to reform, the current algorithm-driven advertising system that lies at the core of the surveillance economy persists. And it persists in systematically promoting extreme, inaccurate and radical content—regardless of what malicious actors may do to seed it. This is the case of political microtargeting, also known as "dark ads" (Borgesius et al., 2018). Such massive power asymmetry between platforms and potentially malicious actors using them on one hand, and individual users on the other, cannot be ignored. One of the actual risks

---

[9] These techniques often rely on a hotly contested scientific paradigm that argues all humans, everywhere, experience the same basic emotions, and that all of us express those emotions in the same way. Those emotions include happiness, anger, sadness, disgust, surprise, and fear. This paradigm of universal emotions is insufficiently evidence-based and poorly regarded in the relevant scientific communities. However, it could become a self-fulfilling prophecy, as a vast range of human expressions around the world are "coarsened" into a narrow set of machine-readable bins. As a result, the many and diverse ways of emotional expression, which vary from culture to culture, may be simplified and thus impoverished.

[10] A variable ratio reinforcement schedule occurs when, after X number of actions, a certain reward is achieved. Slot machines are a real-world example of a variable ratio.

[11] This article made the definition a common reference: https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html. And even if executives from YouTube denied https://nymag.com/intelligencer/2019/03/youtube-exec-denies-the-existence-of-rabbit-hole-effect.html, it really seems that Youtube drives such conspirational content (see the project algotransparency.org).

is that a more or less certain percentage of "persuadables" – people known to be particularly vulnerable to targeted messages – can shift elections. Since the Cambridge Analytica scandals implicating manipulative and possibly illegal social media use in Brexit and Trump 2016 campaigning, challenges and more effective solutions are being discussed (Kohl et al., 2019).

Such Big Data-driven nudging is defined by Yeung (2018) as a technique of "hyper-nudging" which dynamically configures the user's informational choice context in ways intentionally designed to influence her decisions. For example, "dark patterns" define instances where designers use their knowledge of human behavior (e.g., psychology) and the desires of end-users to implement deceptive functionality that is not in the user's best interest (Gray et al., 2018). As such, hyper-nudging concerns all of the design process, not only algorithmic decision-making. This introduces a new form of power—a new "invisible hand" of behavioral prediction and modification which is unknown, pervasive and effective (Zuboff, 2015). Over time, as smart environments will likely permeate societies, users—especially young people—may be automatically plugged in and guided through life along algorithm-driven pathways acting in the best interests of whoever owns or pays to use people's data. These kinds of addictive techniques are already concerning in the case of negative effects on children's wellbeing, including increased risk of suicide and depression, conflicts with parents and adverse effects on cerebral and social development (Kidron et al., 2018).

In sum, the advertising business model of social media creates "a race to the bottom of the brain stem (…) to extract attention by hacking lower into our lizard brains— into dopamine, fear, outrage—to win" (Harris, 2019). And every day the AI system is incentivized to get better and, therefore, overall worse for individual free will and social cohesion. One resultant, paramount concern is that the Internet, especially social media, is downgrading our attention spans, a common base of facts, and capacity for complexity and nuanced critical thinking, hindering our ability to construct shared agendas to solve our problems and the epochal challenges we face. This degraded and degrading capacity for collective action arguably represents "the climate change of culture" (Ibid). Still, existing scientific evidence about these consequences remains lacking, along with easy technological fixes or political solutions.

2.3 A User-centered Perspective

Information consumption has profoundly changed, and thus research need to catch up with these changes and understand novel users' behaviors and expectations, their main causes and effects. On the one hand, news users in fact find and read the news in novel ways. They find them via social media and read it on their smartphones throughout the day, instead of at specific times, such as in the morning during breakfast or in the evening during dinner.

Consider that the time currently spent on Facebook by the average user (more than 2 billion at present) is about 1 h a day, encountering about 350 posts, prioritized out of circa 1.500 (which means that roughly 80% of the posts are being hidden) (Backstrom 2013). In the U.S. – similarly to European countries – two-thirds of users (67%) report that they get at least some of their news on social media (Shearer and Gottfried 2017). Yet, only about 36% of Facebook users intentionally tried to influence their newsfeed while just 14% believe they have a lot of control over it (Smith 2018). However, a 2016 survey of over 53.000 online news consumer in 26 countries showed that more than half agreed that more personalized news may mean they miss out challenging or important information (Thurman, 2019). Yet, in some circumstances users consider this preferable to journalistically curated choices (Thurman et al., 2018). Expectations and behaviors over news personalization are rather nuanced, ambivalent and inconsistent, and this may lead to inaction.

More generally, attitudes towards personalization need to be better understood. People across the world seem to embrace automatically generated personalized recommendations (Reuters Institute, 2016) (see Figure 9 and 10). In short, most people trust themselves more than they trust journalists. For the majority of users, news RSs are not doing anything wrong: they select neutrally and inform or recommend impartially (Gillespie, 2014). However, as we have shown, there is no objectivity in the realm of filtering and personalization. Any artefact such as algorithmic filtering embodies specific forms of power and authority (Winner, 1980). Indeed, they are using supplied criteria to determine what is "relevant" to their audiences and worth knowing, though these biases are not generally recognized.

Most users have no principal reservations against news being distributed through AI-driven tools (Thurman et al., 2018). Yet, they increasingly tend to be worried that "more

personalized news may mean that they miss out on important information and challenging viewpoints" (Neuman et al., 2016, p.113). Most users trust algorithms less and less but they enjoy every day their fundamental services. Also, there are concerns about being wrongly profiled, as well as privacy concerns (Helberger et al., 2019). Though they can have some reservations about these algorithms, many have even more reservations about journalism and editorial selection. Of course, journalism is in crisis and more often survives with click-baiting. As Internet critic and scholar Evgeny Morozov (2017) puts it "what it has gained in profitability, it seems to have lost in credibility." Yet, this sheds the light on the complex background nature of this phenomenon and, furthermore, it stresses the increasing awareness of the potential consequences of personalization.
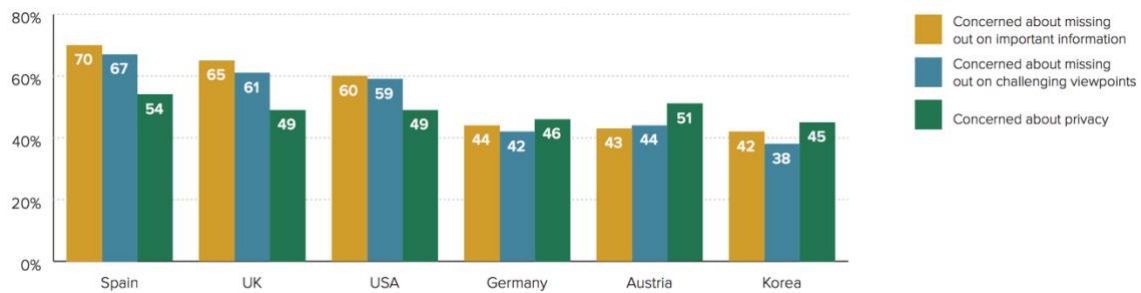
Figure 9. Proportion Who are Concerned About Each Potential Consequence of Personalization (Selected Countries) (from Newman et al., 2016).
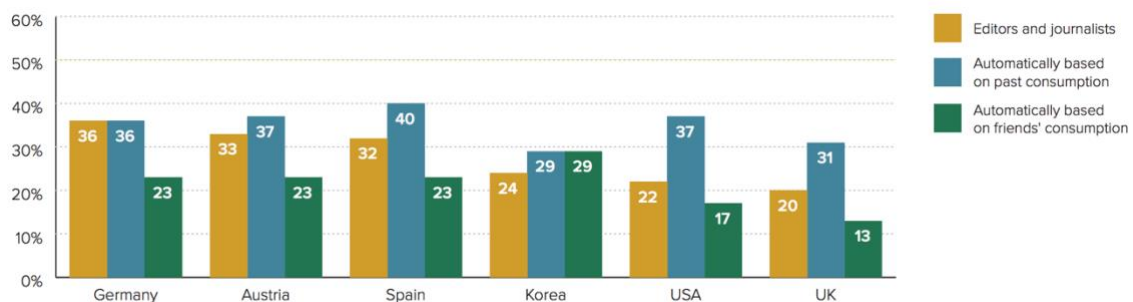
Figure 10. Proportion Who Think That Each Selection Method is a Good Way to Get the News (from Newman et al., 2016).

There are two general explanations for the widespread default trust and lack of responsiveness over platforms and the algorithms that filter one's personal information. On the one hand, in the first place, one compelling explanation for why this became possible so rapidly is that nothing similar to ICT development ever happened in the past,

so there were few institutional defensive barriers (Zuboff, 2015). Information intermediaries have indeed been overestimated and treated as "emissaries of the future," producing "an overwhelming sense of inevitability" (Ibid, p.85). As such, societies quickly came to depend on these new information and communication tools as necessary resources, and at times even as preconditions for social participation.

On the other hand, users' general behavior can be traced back to behavioral economics as well as psychology (Acquisti et al., 2015). Drawing from the famous work of Kahneman (2011), for example, Sunstein and Thaler (2009) describe two major cognitive systems from which individuals take decisions: an "automatic system" and a "reflective system". While the latter is the slow, effortful and controlled way to think, the former is efficient, rapid, largely unconscious and prone to systematic errors. This bounded rationality affects how people assess the likelihood of future events and how their individual biases and vulnerabilities might be exploited.

Also, another major issue with personalization and its future development is that they can increasingly gratify users to the extent to which they come to accept its deception as benevolent. This argument has been discussed in terms of "psychological hedonism" (Gal, 2017): if personalized systems become 'pleasure machines', able to predict our choices and simply grant them to users, are these users willing to give up their autonomous choice? And, if so, under which conditions? On the one hand, we may contend the "choice paradox" which arises from the fact that the decision to turn over one's choices to an algorithm is, itself, an act of choice: to choose not to choose (Sunstein, 2014). On the other hand, we need to think seriously about the way in which the choices we make affect our values, identities and the meaning and content of our lives. In other words, whether we as individuals actually strive for hedonism – that is the view that all human action is ultimately motivated by the desire for pleasure and the avoidance of pain – and the extent to which this is beneficial for society at large. For example, liberal democratic citizenship requires a certain amount of discomfort (Helberger, 2019).

Research elsewhere indicates that users are expecting to play a more active role in the interaction with AI-driven tools such as personalization systems (Helberger et al., 2019). More generally, the use of AI-driven tools fundamentally alters the agency of users regarding the news they consume. AI-driven tools have introduced observable measurements of user interaction allowing detailed insights into audience preferences,

impossible to obtain in non-digital media. Even if users are at the center of these processes, this increase in user agency is unidimensional as it is solely focused on observable engagement like clicks or time-spent. And as explained in the previous chapter, users do not know what information they are automatically excluded from. The challenges outlined above lead to other (open and theoretical) questions: can users be sufficiently informed to be fully responsible in such a fast-changing complex environment? Should the state override personal autonomy to safeguard personal autonomy itself? If yes, to what extent and in which cases? And where positive individual effects go along with negative societal effects: what should be done.

3. The Value of Serendipity

"A new word is like a fresh seed sown on the ground of the discussion."
- Ludwig Wittgenstein, Philosopher

In the previous chapter, we acknowledged how the value of media pluralism and diversity are indeed necessary for democratic societies, yet under-theorized. It is not clear, in fact, how much exposure to how many different contents and sources may be ever considered 'sufficient' or 'satisfying'. Also, media pluralism as a normative principle remains vague and it is not a reliable indicator of a society's level of freedom, since it may create only the illusion of content diversity. In the digital age it is indeed becoming less clear in which sense it is meaningful to speak of media pluralism if the consumption is potentially characterized by limitless choice. In this context, we now explore an alternative and additional design principle to better understand personalization in social media: serendipity.

Often undervalued because of its unpredictable nature, difficulty to operationalize it and wide or sometimes vague definition, serendipity as a design principle is attracting ever more interest, especially in the digital environment. Thus, the following questions will thus be addressed: is there a trade-off in RSs design between accuracy and serendipity? What might be the role of serendipity in the context of information filtering? What is (intended with) Serendipity? How to design for serendipity in digital environments? Why is serendipity valuable? What are the values that serendipity sustains? How to eventually implement serendipity by design? What are the limitations and potential unintended consequences of such endeavor?

Being influenced by environmental and human factors, the experience of serendipity encompasses fundamental phases of production, distribution and consumption of information. On the one hand, designing for serendipity increases diversity, quality and discoverability. On the other hand, serendipity is a capability that has a strong political value. It helps individuals to internalize and adopt strategies that increase the chances of experiencing it. As such, the pursuit for serendipity can help to burst filter bubbles and weaken echo chambers in social media. The chapter firstly analyzes serendipity and its study. Then, it discusses the role and the ethics of artificial serendipity and its design from an interdisciplinary perspective.

3.1 Serendipity and Its Study

Serendipity is the art of discovering new things by observing, and learning from encountering unexpected information. It has received attention in several fields from sociology of science to epistemology, from psychology and innovation studies to information and computer science. It is no surprise that there is no wide consensus on its meaning. It is indeed widespread to consider serendipity as a romantic ill-defined buzzword (Makri 2014). Serendipity is relatively a neologism with no equivalent in languages other than English. It originated from Horace Walpole in 1754, that was inspired by the Persian fairy tale "The Three Princes of Serendip", which narrates how these traveling princes were "always making discoveries, by accidents and sagacity, of things which they were not in quest of". Largely ignored since then, it gained popularity from mid-1900s when it was applied to various breakthroughs in scientific research, until voted as "UK's favorite word" in 2000 (Rubin et al., 2011).

Figure 11. The Percentage of Use of the Word 'Serendipity' in books (Google Books Ngram, 2019).

As serendipity became famous, however, "the vogue word became a vague word" (Merton and Barber 2006). Serendipity is indeed often incorrectly referred to as the "happy accident", thus interpreted as mere coincidence, luck and providence. This dilution of meaning challenges its application in research (McCay-Peet and Toms 2017). In fact, serendipitous discoveries can be perceived as accidentals but they are usually the result of groundwork, observation, and previous knowledge. As Louis Pasteur famously claimed: "in the field of observation, chance favours only the prepared mind." (Merton and Barber 2006). Thus, serendipity has to be considered also as a capability (de Rond 2014), intended as being capable of making use of the opportunities that the environment

provides (Sen 2005). As various definitions assign different weights to personal and environmental factors, between individual sagacity and lucky chance, this tension, eventually results in a combination of them, differing case by case. Serendipity is today often defined as "an unexpected experience prompted by an individual's valuable interaction with ideas, information, objects, or phenomena" (McCay-Peet and Toms 2013).

In the context of digital environments, serendipity can be studied under different perspectives. For information behavior research it has been approached as a quality of someone or, in research relating to recommender systems and search engines, as a quality of an event or something. While being approached as either a personal proclivity or a phenomenon or event, the most holistic approach to serendipity occurs in information science and human–computer interaction (HCI) research where serendipity is often approached as an experience or a process. As such, it has various components, dimensions and environmental factors (Kotkov et al., 2016) as well as affordances (see Figure 13) (Bjorneborn 2017).[12]

As it is a subjective experience—any individual has different strategies to experience different kinds of serendipity— in this thesis we broadly intend serendipity as the experience depending on both environmental conditions and individual capability to encounter— opportunistically (Type A, see note 11) or purposefully (Type B and C)—unexpected meaningful information that may help to solve a problem, make you discover a novel idea, interest, or even change a belief. This experience can be cultivated by design – so-called artificial serendipity – taking into consideration both the extent of profile-based information filtering (how truly personalized an item is supposed to be) and the individual and social level of serendipitous items (how an item attempts to burst your filter bubble or to challenge your echo chambers).

The ability to extract knowledge serendipitously covers all areas of human activity, including 'science, technology, art, and daily life' (van Andel 1994). According to Merton

---

[12] It is possible to generalize serendipitous encounters into three-main types (McCay-Peet and Toms 2017):
- Type A: from observations to a solution. It befalls when people are not looking for any information in particular. For instance, while scrolling a social media's feed. Thus, it is the epitome of nowadays Internet usage.
- Type B: from problem I to a solution to problem II. It occurs when people are looking for information on a certain problem and find a different but previously unsolved problem. To some extent, also this type often befalls in online information seeking.
- Type C: unexpected solutions, also called pseudo-serendipity. It usually befalls when an individual is seeking information purposefully but accidentally finds an unrelated, unexpected one.

and Barber (2006) who wrote the most extensive research about its meaning and diffusion, serendipity is one of the main forces that has steered the progress of science. It is full of renowned examples: from penicillin to DNA, from string theory to mirror neurons. The role of serendipity on epistemology of science is thus well-established. There is, however, a lack of advancement in this field. Interestingly, Campanario (1996) suggests that the reason for this lies in the mythization of scientific research and the difference between the 'context of discovery' and the 'context of justification'. Science is indeed supposed to be something that is totally under control, so that scientists may be reluctant to admit that the discoveries for which they are honoured were accidental.

Similarly, in everyday life the role of chance and serendipity is often underestimated. Though relatively less momentous, serendipitous insights occur on a daily basis, particularly on the Internet. Accidental encountering is indeed an integral part of everyday information behavior (Erdelez 2004), also called micro-serendipity (Bogers and Björneborn 2013). Even though serendipity studies have suggested that serendipity is a rare, elusive phenomenon, in today's information societies for many serendipity seems so common that they do not always reflect upon it and therefore not even recognize it.

Most of the studies on serendipity focus particularly on scientific discovery rather than on everyday serendipity. Extensive accounts on how to research serendipity (McCay-Peet and Toms 2017) and cultivate it in digital environments (Race and Makri 2016; Bjorneborn 2017; de Melo 2018) provided ground for novel studies. These literature reviews suggest how the tension between the need to manage both the quantity and quality of information has been a key driver of these research. They show how serendipity in digital environments can be reduced and cultivated through design, affordances and recommender systems.

Furthermore, it is researched the increasing role of serendipitous news consumption online (or Incidental Exposure to Online News, IEON) for the formation of individual and public opinion (Yadamsuren and Erdelez 2016). Indeed media researchers have largely ignored such phenomenon for a long time, probably due to the difficulty in capturing it. In recent years, however, more studies started to explore how individuals obtain information accidentally and opportunistically, rather than mostly purposively.

3.2 A Brief History of Artificial Serendipity

Serendipitous discovery has been a research topic for more than one hundred years. Fathers of cybernetics recognized the value of a necessary (but not sufficient) element of serendipity: surprise, which operates as a cognitive/emotional reaction to serendipitous encounters (Rubin et al., 2011). In mathematical terms, Shannon (1948) argued that information is proportional to its deviation from expectation. In other words, the amount of information is a measure of surprise. The question of whether computers are able to 'take us by surprise' was instead famously raised by Alan Turing (1950). Despite theoretical discussions, the potential of computers to trigger serendipitous encounters was already clear-cut. A famous example is the 1968's *Cybernetic Serendipity* art exhibition at the Institute of Contemporary Arts in London in which the relationship between technology and creativity was explored.
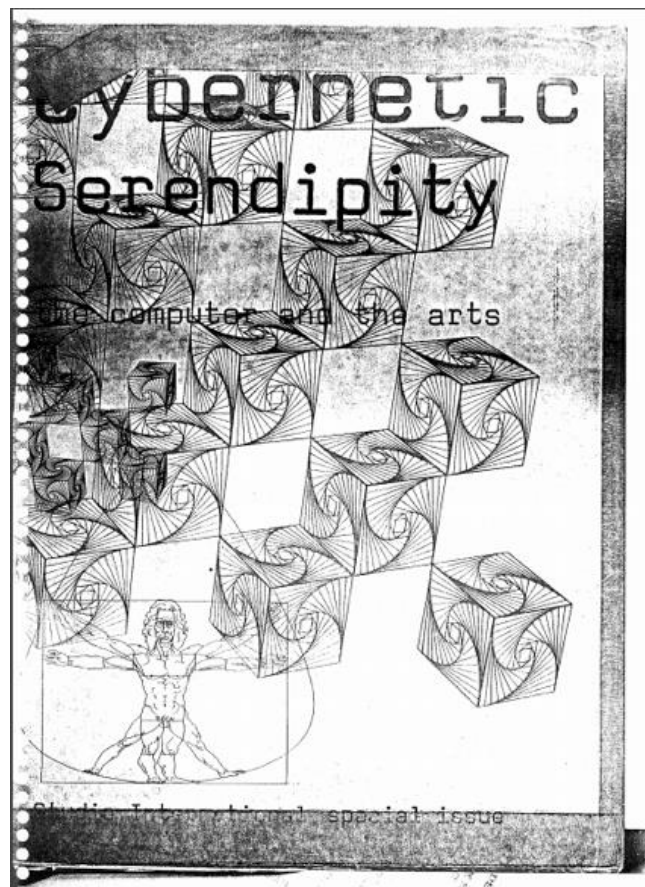

Figure 12. Cybernetic Serendipity Art Exhibition (1968)

The focus on serendipity then shifted in the realm of library and information science, with the first article by Bernier (1960). Early work on serendipity were done in relation to browsing Online Public Access Catalogs (O'Connor, 1988). Functionality for supporting serendipitous discovery included browsable search indexes and similar article citation retrieval (Rice, 1988). In more general terms, however, Van Andel (1994) argued a common argument: *"computer program cannot foresee or operationalize the unforeseen and can thus not improvise ('imprevu' = 'unforeseen'). It cannot be surprised or astonished, and has no sense for humor, curiosity or oddity" (…) The very moment I can plan or programme 'serendipity' cannot be called serendipity anymore."* Skepticism about programming 'true' serendipity is in fact widespread (Krotoski 2011; Carr 2016). Once you create an engine to produce serendipity, one may claim, you destroy its essence. If serendipity could be controlled, then an event is no longer serendipitous, but predictable or reproducible. Yet, it remains possible to cultivate serendipity (Race and Makri, 2016) and it is exactly its unpredictability in which lies its valuable capacity of increasing information diversity, discoverability and quality (Reviglio, 2019). Moreover, the concept of serendipity in digital environments – also defined as artificial serendipity, that is, not only more generally computer-generated serendipity, but also whenever an agent is able to create the necessary conditions for serendipity to occur[13] (de Melo, 2018) – has been reframed with another valence, focusing on both its preconditions and its individual experience (Erdelez, 1999). As such, there are various degrees of serendipitous encounters – it is a continuum to cover the entire spectrum of different degrees of surprise and meaningfulness.

With the advent of the Internet and the consequent change in the production and consumption of information, the concept of serendipity became more debated. It was indeed acknowledged that computer search tended to provide answers to questions facilitating purpose-driven information seeking instead of opportunistic and accidental ones – typical of browsing printed information. Still, without search engines the journey on the web was initially intended as discovering what was out there—accidentally—not on finding specific content (Hendler and Hugill, 2013). This led Gup (1997) to wonder about "The End of Serendipity", in particular the most 'random' and proactively yet

---

[13] Instead, natural serendipity—meaning the serendipity that occurs naturally in the world—is absolutely unpredictable, as the number of factors and variables that create it are impossible (at least as of now), to calculate (de Melo, 2018).

accidentally sought one. Moreover, any procedure to select and prioritize information for users – especially in newspapers, and more generally in traditional media – to some extent recognizes and seeks to solve in a beneficial manner the tension between a generalized relevance—what individuals are expected to want—and serendipity—what individuals may like—due to personalization in digital environments this balance shifted from serendipity to relevance (Thurman, 2011). In the passage from an ink economy to a link economy and from a "professional filter" to a "social filter" (Tewksbury, 2003), this ideal tension became more significant: the potential for relevant and personalized content increased as well as for serendipitous ones.

As personalization improved its accuracy there has been an increasing recognition of the value of serendipity. Media theorist Steven Johnsons (2006) described the Internet as 'the greatest serendipity engine in the history of culture'. Yet, with the development of the web 2.0 few platforms took over the market and began to steer worldwide information consumption. Thus, concerns took over on hopes. Meckel (2011) made an appeal to 'save serendipity' as a fundamental experience for individual progress. Identities can indeed be undermined by profiling algorithms that are stuck in the past, impeding the 'aspirational self'. Profiling technologies that enable personalization in fact cannot produce or detect a sense of self but they influence a person's sense of self (Hildebrandt, 2009). Hildebrandt (2017) argues that such 'incomputable nature of the self' should be somehow protected. In fact, not only no personal profile can ever entirely identify an individual, but there are always many—and sometimes radically different—ways of computing the same person. In this respect, the value of serendipity is of paramount importance as there is a trade-off between algorithmic accuracy and serendipity (Kotkov et al., 2016). In sum, its design acknowledges the natural identity reductionism of profiling techniques, the potential hedonistic redundancy users might spiral in and, more generally, the exploitation of individuals' vulnerabilities that personalization practices and design choices can unfold to maximize engagement and, therefore, it also represents the endeavor to overcome them (Reviglio, 2019).

Despite several challenges that the design of serendipity faces, there is a recognition that it has the potential to prevent the threats of filter bubbles and echo chambers (Yadamsuren and Erdelez 2016; McCay-Peet and Toms, 2017). Even if evidences on the negative consequences of these phenomena are rare and the concepts are ill-defined and

serve more as metaphors, all the models of democracy consider these intertwined phenomena problematic for particular different reasons (Bozdag and Van den Hoven, 2015). As such, the design for serendipity attempts to preserve the likelihood of encountering information one may (not) like—in opposition to individuals' hedonism and platforms' pursuit of profit— to avoid that past behaviours pre-determine and excessively strengthen future ones, and to balance the information diet. Ultimately, designing for serendipity helps the discovery process.

Zuckerman (2008; 2013) is a prominent advocate of serendipity, especially concerning the phenomenon of cyber-balkanization. For example, how the development of a cosmopolitan culture may be limited due to georeferenced algorithms. He expected that serendipity tools would become as important as search engines and social media are nowadays. Indeed, there have been over time more and more attempts to pursue serendipity as a design goal, among them, for information retrieval (e.g. Campos and Figuereido, 2003), for social media (e.g. Burkell et al., 2012), for music discovery (e.g. Zhang et al., 2012), for idea generation to contribute to interdisciplinary research (Darbella et al., 2014) business and ICT development (e.g. Saini and Khurana, 2013), for information behavior strategies to enhance online discovery, in everyday searches (e.g. Buchem, 2011) or in online academic scholarly research (e.g. Maloney and Conrad, 2016).

Nowadays Internet mainstream services such as Netflix, Amazon, Google Now and Spotify certainly provide serendipitous recommendations. Yet, they tend to exploit *convergent systems*—as the capacity to discover the right thing at the right time, to cater to the user's perceived intentions, interests, tastes—rather than *divergent systems*—as the increasing of information diversity in order to expand user's horizons and to help uncover surprising discovery (de Melo, 2018). An example of the former narrative is Google CEO Eric Schmidt (2006) which envisioned a future where people and technology come together to create a serendipity engine "where you don't even have to type" (2006). Internet, however, can be ever more serendipitous, especially in social networking sites and to the extent in which information explorability is wisely sustained. Under the current business model, personalized relevance and quantity outstrip serendipity and quality, providing only an 'illusion of serendipity' (Erdelez and Jahnke, 2018). The capacity to control curation and to navigate information filters is indeed rather limited (Harambam

et al., 2018). As such, there are two major competing conceptions of designing (for) serendipity: one is based on convergent systems and the other on divergent systems. These actually represent two diverse routes to trigger serendipitous encounters. How do we draw the line between the two, if any? Before answering this significant question, arguments on the value of artificial and digital serendipity as a design principle are now briefly discussed.

3.4 A Discussion on the Ethics of Serendipity and Its Design

Despite inevitable limitations in operationalizing complex notions, in this Section we argue how serendipity can inform designers, society and users to create more ethical platforms and algorithms. In the last decade, in fact, applied ethics has developed into an even more practical discipline. Such "design turn in applied ethics" (Van den Hoven et al., 2017) acknowledges that the needs and values of users are considered in their own right and not simply as a side constraint on successful implementation. This perspective helps to develop pragmatic methodologies and frameworks (like *Value Sensitive Design* and pro-ethical design) that help to make moral and ethical values integral parts of research and development (Friedman et al., 2002).

Fundamental questions to open the debate are thus addressed; in this context, how and why can serendipity inform designers, ethicists, users and, eventually, policy-makers? What are the values that it sustains in relation to the risks discussed in Chapter 2? And, also, how to translate serendipity into general norms and design requirements? There can be a "good measurement" or "desirable degree" of serendipity? We now discuss these questions from three main perspectives: societal and philosophical (serendipity as a metaphor), designers (serendipity as a design principle and technical challenge) and educators and users (serendipity as an educational goal).

3.4.1 Serendipity as a Metaphor

Serendipity is a fascinating and powerful concept strictly related to other philosophical ones such as Aristotle's poiesis, Jung's synchronicity, and Nietzsche's amor fati (Lupo 2012; Olma 2016). It is indeed a rather complex value that can be conceptualized in different ways. From an epistemological perspective, serendipity is an observation of a surprising fact followed by a correct abduction (Merton and Barber 2006). In short,

abduction—also called retroduction or inference to the best explanation—as a premise has a known phenomenon while as a conclusion a probable explanation. According to Peirce (1992), abduction is considered the first step of scientific reasoning, arguing that "all the ideas of science come to it by way of abduction" (Ibid). Of course, it may also lead to fallible inferences such as affirming the consequent. Yet these are amendable hypothesis. Abductive reasoning is indeed considered as a more powerful mode for navigating the uncertainties of the information age because it would better support genuine doubt and skepticism (Cunningham 2001), which is indeed associated with greater intention to seek information, more news media use and more frequent online opinion expression (Yamamoto et al., 2018).

In the current Internet landscape, anything is supposedly designed to be efficient and most of the choices are delegated to algorithms. The pursuit for serendipity can thus become a technical challenge to preserve and support a sense of freedom and mystery that is available in less networked information environments. It might indeed become fundamental to maintain in the infosphere the pleasing feelings that elevate accidental discoveries to sensations of serendipity. In future algorithms, it will be indeed necessary to leave some things open to chance (Domingos 2015).

As serendipity cannot be programmed but only cultivated, its design is more probabilistic and, therefore, it is supposed to imply a wider degree of (pseudo)random encounters compared to an efficiency-driven information architecture. Seeking serendipity would also account for what in statistical jargon are called outliers—deviant cases that do not follow the statistical model and that might compromise its predictive power—that represent the personal "long tail". Similarly, Abbott (2008) provided a theoretical basis for concerns about algorithmic filtering, arguing indeed the necessity to defend randomness from processes for search and discovery. Because of evolution and the ability to adapt to environmental changes, human beings are "antifragile" (Taleb 2012), in the sense that they actually benefit and get better from shocks, randomness and uncertainties. In other words, extremely accurate algorithms may reduce human resilience. Personalization algorithms might indeed overly reduce pain and uncertainty (Gal 2017) so as to create only an illusion of knowledge, eventually pushing people to radicalize their beliefs due to the easiness to receptively confirm them in online settings (Lynch 2016). Adding to the 'Socrates paradox' (I know that I know nothing), serendipity

helps to internalize the significance of the 'unknown unknown', what we don't know that we don't know. Thus, it preserves necessary feelings of uncertainty and mystery. Philosopher of information Floridi (2015a) argues that a low and stable degree of uncertainty leads to increased degrees of liberalism and toleration, as well as more efficient flows of information. Similarly, Marcus (2010) acknowledges that the emotional anxiety provoked by perceptions of conflict, even if unpleasant (e.g. zemblanity, see note 19), is the only means of forcing a rational re-evaluation of one's currents beliefs.

The value of serendipity can also be considered on the light of future technological developments. Advances in so-called artificial intelligence, machine learning and the semantic web have the potential to enable algorithms to make ever more sophisticated recommendations. Whereas virtual reality, augmented reality and the Internet of Things will definitely blur the anachronistic distinction between online and offline, giving rise to the vision of Ambient Intelligence, or even leading to a 'social physics' (Pentland 2015). It is a real challenge to design serendipity in physical places and other emerging contexts.

3.4.3 Serendipity as a Design Principle

Programming for serendipity sounds like an oxymoron, and to some extent it is. It is, in fact, a subjective experience that is only discernible in hindsight and for which the unexpected plays a fundamental role. Serendipity, in effect, cannot be created 'on demand'. While natural, 'pure serendipity' is not amenable to generation by a computer (van Andel 1994), artificial serendipity can actually be cultivated by creating opportunities for it through the design of physical, digital, and learning environments (Race and Makri 2016; de Melo 2018). So far, however, relatively little research has been undertaken to assess how well existing approaches to information interaction support and assess serendipity.

Even though there is no set of systematic requirements for a system that facilitates serendipity in digital environments yet, many scholars agree that the main functional requirements are to enable the "anomaly" or the chance encounter to trigger an event, support the user in "connecting the dots" so as to make him/her reach a significant surprise outcome (McCay-Peet and Toms 2017). As said, serendipity is also an affordance, an opportunity that the environment offers to the user that is able to grasp it, as an intuitive invitation (Gibson 2014). At the same time, there are affordances for

serendipity itself (Bjorneborn 2017). Clearly, the challenge is how to operationalize them, especially considering the fuzziness of the concept of serendipity.

Given the interdisciplinary and elusive nature of serendipity, different yet related disciplines have interpreted and employed the concept of artificial serendipity in different ways. For this reason, it will be now attempted an initial deconstruction of such concept in the context of personalization based on an interdisciplinary literature review.

10 SUB-AFFORDANCES FOR SERENDIPITY	3 KEY AFFORDANCES FOR SERENDIPITY	3 key personal factors for serendipity	10 personal sub-factors for serendipity
DIVERSITY [multiple potentials]	DIVERSIFIABILITY	curiosity	*interest* [regarding diversity etc.]
CROSS-CONTACTS [colliding potentials]			*playfulness* [regarding cross-contacts etc.]
INCOMPLETENESS [unfinalizable potentials]			*inclusiveness* [regarding incompleteness etc.]
ACCESSIBILITY [access to specific spot, convergently]	TRAVERSABILITY	mobility	*searching* [convergent]
MULTI-REACHABILITY [reaching anywhere, immersively]			*immersion* [both convergent & divergent]
EXPLORABILITY [inviting somewhere else, divergently]			*exploring* [divergent]
SLOWABILITY [affording slower pace, frictionally]			*stumbling* [both divergent & convergent]
EXPOSURE [highlighting broader, over longer time]	SENSORIABILITY	sensitivity	*attention* [broader sensing]
CONTRASTS [highlighting sharper, more suddenly]			*surprise* [unprepared sensing]
POINTERS [highlighting narrower, more specifically]			*experience* [prepared sensing]

Figure 13. Key affordances and sub-affordances for serendipity with coupled personal factors and sub-factors (from Björneborn, 2017).

At a theoretical level, the pursuit for artificial serendipity underpins several intertwined assumptions and values (Reviglio, 2019); acknowledging the positive value of designing for serendipity in fact means that despite advancements in computing, the outcomes of algorithms will always be actually unpredictable. There is no perfect personalization indeed (call it 'hyper-personalization') as algorithms will be more likely never able to "fully" understand the complexity of a human being. Even individuals may not always be sure of what they actually want and prefer. As such, personalization of news feeds ought not to approach any accuracy optimization and try to preserve a degree of 'random encounters' and items less profiling-based. Eventually, on the one hand, the design for serendipity helps to appraise the natural limitations of personalization – even at the cost

of users' engagement and hedonism – while, on the other hand, it invites us to accept the multiplicity of our identities and the need for a balanced and diverse diet. In any case users ought to be responsibly empowered with an effective information discovery agency and not to fully delegate information seeking to an often privatized and naturally biased algorithmic information filtering.

As it will be further outlined in this chapter, there are two main interpretations in the design of digital serendipity, particularly in RSs, namely pseudo-personalized serendipity and hyper-personalized serendipity: this is a theoretical spectrum which posit how serendipity is designed based on a user profiling, thus the extent to which designers rely on (supposedly) more accurate predictions. Even if hyper-personalized serendipity is found on the serendipity spectrum, it is less likely to be as serendipitous as pseudo-personalized serendipity. In fact, the actual difference between the two is the expected accuracy of the recommendation. More "pure" – thus more unexpected – serendipity is more likely to occur through (semi or even non) personalized recommendations. In the case of hyper-personalized recommendations, to be truly serendipitous they should be not only useful or surprising but also meaningful, memorable, insightful, information which encapsulates quality and epistemic relevance (Floridi, 2006). These should, for example, 'fairly' attempt to stimulate individual progress and societies cohesion.

Ideally, as users and citizens we need and want both hyper-personalization and pseudo-personalization. Hence, from a theoretical level it is needed to find a balance. Still, from a practical level it is hard to argue any optimal degree of pseudo- and hyper-personalization. Also, a right level of trade-off between the provider's and users' interests is debatable. Eventually, being able to identify personalization's accuracy in particular items as well as in the general consumed media content may be the basis for autonomous choice, if a proper agency is afforded by design. In order to increase serendipitous encounters, similarly to the dynamic relation between randomization, personalization and generalization suggested by Carr (2015), designers and engineers should strive to balance pseudo-personalization, hyper-personalization and non-serendipity, that is, all that information that ex-post did not unfold in any serendipitous experience. The latest represents the value of diversity; the assurance that news feeds are not all or excessively hyper-personalized, so that individual and political serendipity are cultivated so that a user actually expands his/her horizons and it is encountered information one may (not)

like, thus leaving space for generalization and diversification. Yet, it is exactly this inevitable consequence of designing for serendipity that sustains serendipity as an ethical principle in democratic societies, by diversifying information, preventing hyper-personalization and sustaining a stronger common base of facts and ideas, outside one's filter bubble and echo chambers. This framework resonates with a core function of traditional media, that is, to provide 'reliable surprises' in order to balance familiarity and chaos (Schönbach 2007) and, similarly, a balance between convergent and divergent systems (de Melo, 2018).

3.4.4 Serendipity as an Educational Goal

Serendipity embraces certain values that can represent future educational goals and political struggles of the digital era. Not only its design generally sustains creativity and innovation but its teaching helps to stimulate proactive information seeking, inter-disciplinary skills and enrich the algorithmic imaginary. As a consequence, it might help to solve the main challenges to experience diversity online (see Hoffmann et al., 2015) and, at the same time, sustain the four modalities of expressing voice (see Harambam et al., 2018).

Firstly, serendipity is a capability and, as such, can be considered in terms of virtue ethics. As sagacity and curiosity are often acknowledged as fundamental virtues, there are two other primary ones: humility and courage (Lupo 2012). On the one hand, humiliation is necessary in the face of the imponderable occurrence of chance so as to radically question one's own beliefs. On the other hand, courage is needed to follow such chance, and to turn it into an opportunity against its potential risks. It is also necessary for any individual engaged in the task of getting rid of beliefs and prejudices that curb the cognitive path. As scientists strive to embrace such attitude, individuals can also benefit from being nudged to be sagacious and critical active-seekers alike in online settings. Critically, to experience serendipity it is needed to take the risks of failure, frustration and "wasted time". Serendipity is also strictly related with randomness, chance and, therefore, uncertainty and mystery. Tolerance towards uncertainty is thus a desirable educational outcome.

Secondly, since serendipity is intimately related to research and discovery, to occur it requires an environment that stimulates platform's transparency and user control, it

implies, therefore, an empowerment of individuals. Contrary to many legal formulations in data protection which are framed in terms of harms, and thus freedom from algorithms, serendipity can be considered a positive freedom, as an individual, unexpected and valuable information. When users are treated as active-seekers, then, they can perceive their findings as triumphs of personal agency, intuition, and inspiration, and as a self-reinforcing expectation, increase their freedom from algorithms. Maturity of information societies is indeed a matter of people's expectations, not just technological or economic development (Floridi 2016a). If one analyzes the phenomenon of personalization through the lens of serendipity, it is possible to rebalance the asymmetric power between human agency and computer delegation, and between users and corporations' rights.

Thirdly, serendipity has a semantic value, in the sense that it can expand users' algorithmic imaginary (Bucher 2017). Our current conceptual toolbox indeed is no longer fitted to address new ICT-related challenges and, as Floridi (2015b) argues, this is a risk because 'the lack of a clear conceptual grasp of our present time may easily lead to negative projections about the future: we fear and reject what we fail to semanticise (p.3)'. Serendipity can actually represent a narrative for emerging information societies and enrich the 'ecology of the self', as the ethics of self-poiesis (Floridi 2011). For instance, it can strengthen individual's reactance to bias-driven algorithms. Either, it might help to recognize one's own responsibility in the "spreadability" of serendipitous information, given that users are actually one another's curators (Race and Makri 2016, p. 5.2) (Table 1).

Finally, managing serendipity can generally help users to navigate our hyper-complex world more resiliently. Scholars have remarked that the researcher of the future has to be able to work in several fields temporarily and to find similarities between apparently heterogeneous concepts (Lupo, 2012). Serendipity, in fact, more often comes out among interdisciplinary scholars (Edward Foster and Ellis, 2014) and it boosts interdisciplinary research (Darbellay et al., 2014). In the informal learning that derives from HCI, a serendipitous proclivity represents a powerful soft-skill to tackle the challenges of the information era and a new fundamental idea that contribute to rethink the current context of education (Cobo and Moravec, 2011). Indeed, it can help to develop abilities of self-learning through accidental learning (Kop, 2012; Yadamsuren and Erdelez, 2016).

Favours	Disfavours
(Pseudo)Randomization	Predictability
Chance	Efficiency
Noise	Signal
Autonomy	Delegation
Information seeking	Passive consumption
Uncertainty	Security
Pluralism	Political polarization
Belief change	Radicalization
Interdisciplinarity	Specialism
Divergency	Convergency
Realism	Hedonism
Weak ties	Homophily
Hard news	Soft News
Innovation	Redundancy
Education and self-learning	Engagement
Long-term satisfaction	Instant gratification

Figure 14. Summary of the main opposing phenomena and antithetical values that the design of serendipity tends to favour and disfavour (from Reviglio, 2019a).

3.3 A Taxonomy for Designing Digital Serendipity

In the last article written on serendipity (Reviglio, 2019b), it was finally attempted to extrapolate implicit meanings of serendipity as a design principle, especially in the context of social media. To begin with, two general approaches that can be taken by designers are default rules and active choosing (Sunstein, 2017b). As previously mentioned, a user might take both a non-purposive or passive information behaviour (which generally supports Type A serendipity) and a purposive or active (mostly Type B and C) (Kotkov et al., 2016). On the one hand, *active* serendipity occurs both when designers provide serendipitous tools by design and when users proactively seek information and eventually experience serendipity with these. On the other hand, *passive* serendipity occurs when engineers inscribe serendipity into algorithms and when user is passively exposed (and nudged) to potentially serendipitous information. These intertwined conceptualizations share another paramount distinction: *hyper-personalized serendipity* and *pseudo-personalized serendipity* which ideally differ for the extent to

which they rely on individual profiling and, therefore, on their (presumed) accuracy. Furthermore, we identify *individual* and *political serendipity*, which respectively aim to increase the exposure to new ideas and potentially new interests, outside one's filter bubble, and facts, views and arguments, outside one's echo chambers. These can be both hyper- and pseudo-personalized.

All these different interpretations of serendipity need a more detailed analysis. In fact, they emerged in literature but they may be difficult to grasp as they are often implicitly intended and their meanings often overlap. Therefore, we now propose a preliminary taxonomy of the main interpretations of serendipity in digital environments – especially in personalized social media – according to different design processes and disciplines' perspectives.

3.3.1 Passive Serendipity – Serendipitous Recommender Systems

Passive serendipity is mostly attained in personalized information flows. Personalized recommender systems (RSs) have presented since long the problem of excessive homogeneity. RSs research in fact began to move beyond accuracy and experimenting with serendipity (Ge et al., 2010). Indeed, user satisfaction does not always correlate with very accurate RSs. This, however, does not guarantee that mainstream platforms will take it into consideration. RSs have already been criticized to not sufficiently account for serendipity (Matt et al., 2014). Optimizing current RSs for more serendipitous recommendations is not a trivial task of course. Promising attempts have been already done long ago. Campos and Figuereido (2002) famously pioneered an information retrieval software that resulted in 52.7% of (pseudo)serendipitous suggestions. Despite the subjective character of the results, they showed that explicitly programming for serendipity is possible. Yet there is still no consensus on the definition of serendipity in RSs (Kotkov et al., 2016). It is, however, a compound of other concepts like novelty, utility, diversity, unfamiliarity and unexpectedness.

Even though design a RS is a multi-criteria decision making problem, there is a major theoretical design trade-off between two main criteria: accuracy and serendipity (Ibid). Accuracy-driven RSs can lead to more engagement and satisfaction for users (thus revenues for information intermediaries) than serendipity-driven RSs. More often, in fact, serendipitous content may not be liked. Designing for serendipity, however, might

increase user satisfaction in the long-term. Furthermore, seeking and experiencing serendipity is time-consuming. Potentially-serendipitous content (e.g. a long article or a documentary) may bring the user to another website, pulling off profitable scrolling. More generally, it implies a slower consumption of information as it may require an 'incubation period' (McCay-Peet and Toms 2017). Thus, such time may be given away to scrolling, and so to engagement within the platform. The profit-driven model of mainstream SNS, for example, tend to prioritize ephemeral content to durable one, short videos (snippets) to long ones and snapshots (casual images) to written text as they better lock-in users in their "walled-garden" (Derakhshan 2016). Eventually, other relevant criteria need to be considered to create serendipity-driven RSs, in particular challenging ethical and political information.

3.3.2 "Active Serendipity" – Serendipity by Design

There are actually countless ways to empower users, extract value from their profiles meaningfully, to illustrate blind spots, increasing diversity, and to stimulate creative associations. First of all, visualization and discovery tools such as the use of real-time transparency dashboards help to increase users' awareness, autonomy and, as a consequence, the potential to access and encounter more diverse and, eventually, serendipitous information (for example, see Nagulendra and Vassileva 2016). There are indeed several visualization tools that can increase control and awareness about a user's information diet. In this context, a significant design choice might be to afford more profiles per user. The management of multiple filtering can in fact nudge users to subtract from the determined path offered by algorithms (Bozdag and Timmermans 2011; Hildebrandt, 2017). Similarly, to afford users to "empathize" with—namely browsing— the information flow of other's users, friends, and even groups may be serendipitous. Either, to exploit others' profile to filter your own feed (what is called algorithmic recommender personae, see Harambam et al., 2018). Yet, it must always be acknowledged the paradox of choice, that is, the more to choose from, the less that is chosen. These design choices and visualization tools provide a general idea of serendipitous design choices and tools that could supplement user experience. Its aim would be to provide users with 'institutionalized serendipity' (Merton and Barber, 2006).
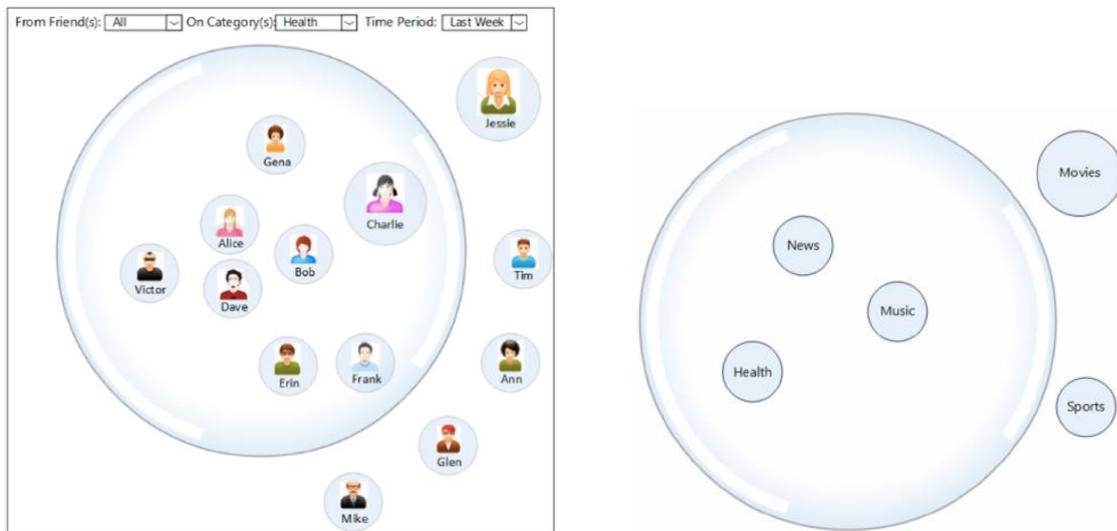
Figure 15. A dashboard that displays users their filter bubbles (from Nagulendra and Vassileva, 2016)

3.3.3. Hyper-personalized Serendipity

Hyper-personalization is generally intended as the use of data to provide more personalized and targeted products, services and content. Hyper-personalized serendipity can thus be intended as the attempt to recommend serendipitous items based on an individual data profile. In recent years there has been growing interest in this philosophy, and in how it can be applied to a variety of disciplines and issues. This endeavor could follow the principle of kairos[14] (one of the principles of captology) that refers to the ability to provide the right content, at the right person in the right moment (Fogg et al., 2002). Hyper-personalized recommendations could in fact be timely, useful, surprising and persuasive, yet not necessarily serendipitous. When this is the case, hyper-personalized serendipity could represent what has been termed as 'serendipity on a plate' (Makri et al., 2014), namely highly personalized and passive serendipity. It is the conception of serendipity more akin to research on RSs, marketing and nudging practices (Kotkov et al., 2016). Thus, it is primarily driven by the principles of delegation and efficiency. To some extent, it is indeed a necessary feature of RSs and other information systems. In theory, however, encounter only (or clearly mostly) hyper-personalized information, particularly in personalized news feeds, may only create an illusion of serendipity (Erdelez and Jahnke, 2018) and become detrimental both to individuals and societies at large (Reviglio, 2019). Yet, hyper-personalization is also an ideal tool for news

---

[14] Traditionally kairos is a Greek divinity, personification of the "opportune moment".

organizations to regain control of the news distribution process and reconnect with audiences.

Hyper-personalization is the paradigm of convergent systems. To give an example, one watches a movie at the cinema and then stumbles upon a review of that particular movie online. One may experience this as serendipitous, and the content related to the review may well be indeed. Its "serendipitousness" may, however, be limited. These recommendations may be less unexpected by users – they can realize they depend on data surveillance – and, therefore, it becomes naturally less serendipitous while, relatedly, it is problematic from a privacy perspective. Moreover, the conflicting interests of platforms that provide such services are questionable; in fact, the natural feeling of privacy loss can undermine trust over the system. For this reason, it is clear how the deployment of hyper-personalized serendipity may be even limited in order to preserve profitable surveillance practices. Also, it might be employed in a crafty and deceitful way. For example, by creating compulsion loops that are actually found in a wide range of social media. Thus, hyper-personalized content could be purposefully offered among uninteresting content in a way that increases engagement (intended as the more scrolled posts, the more ads sold and thus revenues). In this case it is fake serendipity (as further explained). Moreover, user's privacy preferences can constitute informative metadata which the system could use to make sensitive inferences about the user and ultimately influence such exposure to hyper-personalized content. Despite speculations, the business model of mainstream social media is nevertheless based on surveillance practices yet individuals are even risking of not getting a deserved pay-off, that is, high quality (personalized) information.

3.3.3.1 Illusory Serendipity

Illusory serendipity is a sub-group of hyper-personalized serendipity. Yet, while hyper-personalization may also imply serendipitous encounters, illusory serendipity is surprising but does not lead to any valuable insight. It is merely fake, ephemeral and/or even addictive. It usually exploits users' inferred vulnerabilities regardless of the quality of information. Illusory serendipity is indeed inherently persuasive and, at times, manipulative. It drives click-bait, misinformation and disinformation. Further research on the techniques to design it and the individual reactions to it is certainly needed to better understand its expression.

3.3.4 Pseudo-personalized serendipity

Pseudo-personalization refers to all those recommendations that are either statistically unlikely, either less likely – that may lead to more unexpected serendipitous encounters. To some extent, it is already an implicit by-product of personalization practices as they are indeed inherently probabilistic. This, however, depends on the extent of the likelihood of a recommendation to be liked (recommendation's accuracy). As such, it refers to the spectrum in which recommendations are less profile-driven. On the one hand, pseudo-personalization might be highly unlikely to be liked (low accuracy of recommendation) – or at times even be random or – from a design perspective – random in an 'exploratory way'. On the other hand, it might be more accurate, that intersects a user's profile but, still, less accurately than hyper-personalization.

This design attempt is significant to trigger serendipitous encounters because in order to make sense of chance information exposure "information must resonate with some prior knowledge, interest, or experience for the user" (Helberger, 2011). Still, contrary to hyper-personalization, engineered pseudo-personalization attempts to provide the most unexpected serendipity – as being less probable and less predictable – it thus represents a more hazardous attempt to design for serendipity, particularly from an economic perspective (i.e. engagement and thus profit may be threatened). Also, it is time-consuming, risk-taking, sometimes distracting and frustrating. In order to properly sustain pseudo-personalized serendipity, information discoverability –the possibility to find and access content with a specific quality– is also fundamental, along with the semi-randomizability of information filtering. In the case of active serendipity, its role becomes even more significant; it could in fact provide the possibility to subtract from any potentially (biased or excessively pre-determined) algorithmic curation and be able to be exposed to more casual, diverse and, eventually, serendipitous content. The main aim of pseudo-personalization is to actually overcome the potential determinism implicit in profiling – which can result in redundancy and limited pool of information – as well as platforms' information selection power. Since long this limitation of recommender systems was acknowledged and it originates from the analogous "over-specialization problem", that is, when the system can only recommend items scoring highly against a user's profile, the user is restricted to seeing items similar to those already rated.

3.3.5 Individual Serendipity

'Individual serendipity' is related to the 'filter bubble problem' and, thus, particularly related to personal identity, particularly affected by ICTs that are indeed technology of construction of the self (Floridi 2011). It is individual exactly because it focuses on the individual information consumption. It is thus intended as the design attempt to expose users to information that intersects a user digital profile, so that an individual can discover new ideas and interests, indeed outside one's filter bubble. As such, this conception is the most akin to the concept of 'content diversity' by design (Helberger, 2011) and it implies a degree of generalization and diversification of information selection. It can be sustained through active serendipity that means, generally speaking, to provide serendipitous affordances (Gibson 2014). Similarly, discoverability is fundamental to sustain cultural diversity (Burri, 2019). In this context, a significant design choice might be to afford more profiles per user. The management and access to multiple filtering can in fact nudge users to subtract from the determined path offered by algorithms (Bozdag and Timmermans 2011). This conception of serendipity also sustains what Hildebrandt (2017) defines as 'agonistic machine learning', that is, demanding companies or governments that base decisions on machine learning to explore and enable alternative ways of datafying and modelling the same event, person or action.

Pseudo-personalization may actually be not particularly problematic as long as the system adapts to the individual serendipitous disposition – the more diversity you consume the more diverse the information selected. There are, however, two main issues to consider. On the one hand, growing 'epistemic inequality' (Lynch 2016). Certain privileged group of users, that have higher (digital) literacy, are effectively able to reach a good balance between relevance and serendipity, and the recommender systems would indeed adapt to their serendipitous proclivity. Instead, a larger group of users risk to be exposed only to a minimum, qualitatively inferior range of information. On the other hand, there is an asymmetry between first and second order preferences. What if a user would ideally prefer to be more open and curious while in practice actually prefers her or his comfort zone if this is (more often subtly) offered? In other words, individual serendipity also attempts to fill the mismatch between the necessity for the aspirational self to uncover and the intrinsic identity reductionism of profiling. Finally, one may feel

to enjoy a diverse media diet while it is ignoring the actual potential diversity and serendipity the Internet may offer. Therefore, it is essential to provide 'serendipity agency', namely the capability to use tools to control and be aware of the production and one's own consumption of information. Such endeavor embraces 'pro-ethical design' or so-called "tolerant paternalism" (Floridi, 2016) that aims to modify the level of abstraction of the choice architecture by educating users to make their own critical choices.

3.3.6 Political Serendipity

Political serendipity concerns the risks of echo-chambers and, thus, it is particularly related to certain democratic principles, especially media pluralism. It is intended not only as a metaphor but also as the proactive attempt to expose users to content that is politically challenging in order to balance user's political news diet. It is typically intended as such in discussions related to media ethics and media pluralism (Helberger, 2011; Hoffmann et al., 2015; Burri, 2019). Theoretically, it is sustained by a vast literature and all democratic models (Bozdag and Van den Hoven, 2015), in the liberal one particularly for the concept of 'marketplace of ideas' while in the deliberative one to promote open-mindedness and mental flexibility (Helberger, 2019). More generally, it sustains the fundamental right to autonomously seek and be exposed and eventually assess alternative fact, viewpoints, worldviews, with the chance to raise that emotional anxiety that may trigger belief change (Marcus, 2010) or simply to preserve a common base of facts to discuss.

Basically, political serendipity it might be achieved either through the attempt to balance the political news offer by design, either to proactively attempt to expose to challenging and alternative views in a highly personalized manner. On the one hand, however, being actively exposed to challenging information—eventually serendipitous—may not necessarily result in experience of diversity, de-polarization and/or more tolerance. In some cases, the opposite may be true. Polarized individuals radicalize further while exposed to challenging information (Quattrociocchi et al., 2016). Experiencing serendipity especially requires attitudes of open-mindedness and sagacity that are, above all, an educational issue. Yet, political serendipity by design could, in theory, help to prevent people to begin to radicalize (Reviglio, 2017). On the other hand, there may be

problems related to political hyper-personalization and illusory serendipity. In other words, persuasive or even manipulative political micro-targeting. Therefore, it may be better to provide users affordances. In practice, Sunstein (2017b) proposed that social media could create a "serendipity button" for news and opinions, allowing people to opt in, especially during elections. Related stories at the bottom of a "post" also seem to help in counteracting misinformation or simply enriching a user perspective in a serendipitous way (Bode and Vraga, 2015). Another example to sustain political serendipity as an agonistic approach to political news may be *Flipfeed*, a MIT Lab-made Twitter plug-in that provides to the users the possibility to scroll the feed of a random individual which resides in a far ideological spectrum from one's own.

Serendipity Category	Level	Explicit Goal	Latent Goal
Individual	Personal Identity	Increase 'Diversity Exposure and Experience'	Identity Development
Political	'Public Sphere'	Balance Political News and Issues	Preserve the 'Marketplace of Ideas'
Illusory	Cognitive	User Satisfaction	Engagement Maximization

Figure 16. Summary of the Serendipity Categories' Features.

Personalization Serendipity	Hyper-	Pseudo-
	← *continuum* →	
Individual	Accuracy-driven (Profile-Convergent)	Discovery-based (Profile-Divergent)
Political	(Persuasive/Manipulative) Microtargeting	Balance of the 'Marketplace of Ideas'
Illusory	"Kairos"	Mostly "virality-driven"

Figure 17. Summary of the Relationships among the Categories.

3.5 Limitations and Unintended Consequences

There are several arguments that problematize serendipity as a design principle. As many other principles, indeed, serendipity works well in theory while in practice shows many limitations. Several challenging issues indeed lie on its definition, its normative translation and, eventually, in its measurement.

First of all, one of the main challenge is indeed the measurement of serendipity, which mainly depends on the subjectivity of the phenomenon. Individuals have very different thresholds for considering something serendipitous. Serendipity is not a purely discrete concept (Makri and Blandford 2012). In its assessment, serendipity is a continuum to cover the entire spectrum of different degrees of surprise. Therefore, serendipity cannot be considered as a 'good measurement' that is valid, reproducible and accurate, thus it can become object of dispute (see Kroes and van de Poel 2015). It can be established, nonetheless, a reasonable or justified consensus on how to operationalize it and measure it. On the base of which evaluation criteria, then, information architectures and RSs should be designed remains open-ended. Measurements of serendipity can be both quantitative and qualitative scales (McCay-Peet and Toms 2017). It is indeed possible to observe serendipity in controlled research environments, throughout surveys or by directly observing information encountering behaviour (Kotkov et al., 2016). Quantitative measurements may then be based on mixed method approach and other evaluation criteria. Eventually, these might complement each other to potentially come up with an additional key performance indicator. Nevertheless, a "desirable degree" of serendipity remains subjective; in the case of passive serendipity, how much potentially-serendipitous information should be provided both in the quantity of un/expected information recommended and the degree of such un/expectedness? Perhaps, somewhat counter-intuitively, an optimal balance may be found personalizing serendipity, albeit there might be significant privacy implications (McCay-Peet and Toms 2017).

A related-risk lies in potential manipulations of serendipity (particularly passive serendipity), due to its relatively conceptual fuzziness and, again, its subjective character. Paradigmatically, as already mentioned, Google CEO Eric Schmidt (2006) envisioned a future where people and technology come together to create 'a serendipity engine' "where you don't even have to type" (2006). Instead of create serendipity and offer it 'on a plate', however, it is more sustainable to create opportunities for it by supporting strategies that

may increase its likelihood, thus empowering users to create their own personal 'recipes' (Makri et al., 2014). Mere simulations of serendipity, in fact, may produce vague, washy attempts to make the world a better place (Olma, 2016). One risk may be to legitimize whatever recommended content with serendipity and to muddle it with sensationalism, propaganda or even disinformation and misinformation (so-called fake news).

From a regulatory perspective, optimize social media for serendipity does not appear promising. Above all, it might be considered as a form of stealthy paternalism naturally suitable to handling. Yet epistemic inequality might persist, if not increase. Challenge-averse individuals who would ideally benefit more from an architecture for serendipity, are naturally those less willing to play with it. On the contrary, those already prone to it are those more willing to benefit from it. Given the heterogeneity of users, to avoid further inequality a regulation may be required. As said, the actual loss for platforms might be users' engagement and, consequently, it would threaten companies' revenues. Therefore, it also represents a political and legal challenge.

Still, the beneficial effects of an architecture for serendipity are uncertain. Looking for serendipity is often time-consuming, risk-taking, sometimes distracting. As explained above, being actively exposed to challenging information—eventually serendipitous—may not necessarily result in experience of diversity, de-polarization and/or more tolerance. In some cases, the opposite is true. Polarized individuals radicalize further while exposed to challenging information (so-called backfire effect) (Quattrociocchi et al., 2016). Any intervention should eventually be measured accordingly. Experiencing serendipity especially requires an attitude of open-mindedness that is, above all, an educational issue. Therefore, while from a theoretical perspective serendipity can be advocated as a beneficial design principle, from a pragmatic perspective it is a highly problematic endeavor. Not only there are still potential improvements in the understanding of serendipity from a theoretical perspective but there is also significant design, psychological and educational research gap. Yet, we argued that serendipity has to be conceived both passive and active. In other words, an architecture for serendipity should balance personalization, generalization and randomization, and may be beneficial as insofar it is both implicit and explicit, convergent and divergent.

3.6 Conclusions

Serendipity is a polysemantic notion and polymorphic phenomenon. As such, different yet related disciplines have interpreted and employed the concept in different ways. The chapter outlined the research history of serendipity in digital environments to seeks to map the conceptual space of 'artificial serendipity'. This was done carrying out an interdisciplinary literature review and defining and discussing relevant terms in the area, in particular the distinction between hyper-personalized and pseudo-personalized serendipity. These have been made to further develop the study of serendipity in digital environments. In fact, artificial serendipity proved to be a technically volatile and a conceptually nuanced ethical principle. Though, it is also a beneficial design principle for both individuals and democratic societies. In fact it could help to burst filter bubbles and echo chambers, particularly in personalized news feeds. Cultivating serendipity is individually desirable and socially beneficial. Its design can sustain media pluralism and strengthen human rights, such as freedom of expression and the right to receive information.

As of now, many research gap remain in the study of serendipity in digital environments. Discussions and, eventually, consensus on the definitions, the design and the assessment of serendipity need to be further reasoned. Also, the practical effectiveness in bursting filter bubbles and soften echo chambers is questionable. Digital literacy, critical thinking and high quality media remain fundamental preconditions for a beneficial expression of serendipity in digital environments. There is certainly room for research on serendipity in learning environments, conceived as a soft-skill to master. Nonetheless, under the current business model of mainstream social media, for example, serendipity may be limited, and often mingled with what we termed 'illusory serendipity'. In fact, when designed for profit-making alone, or when they are mainly profit-driven, "algorithms necessarily diverge from the public interest" (Benkler, 2019). In order to increase serendipity, designers and engineers should recognize the nuances of designing for serendipity – especially the trade-off with accuracy – and, from an ethical standpoint, attempt to balance hyper-personalized and pseudo-personalized recommendations, even throughout design affordances and information discovery tools. Our current conceptual toolbox is in fact no longer fitted to address new ICT-related challenges and "the lack of a clear conceptual grasp of our present time may easily lead to negative projections about

the future: we fear and reject what we fail to semanticise" (Floridi, 2015, p.3). These concepts can eventually better inform and inspire educators, designers and engineers to help individuals find the information they need, not just the information they think they need, or that algorithms presume, and sometimes even deceive, they need.

4. Personalization and Its Regulation

In this Chapter it is analyzed the phenomenon of personalization from a regulatory (4.1), ethical (4.2) and governance (4.3) perspective. The analysis takes into account the interdisciplinary nature of personalization and it is focused on the European legal context. To begin with, modes of regulation and an overview of social media regulation are broadly discussed. Then, an analysis of the European Legislation is provided. After having introduced a human rights perspective, there are briefly discussed the E-Commerce Directive, E-privacy Directive, the General Data Protection Regulation and, more generally, media law. In the second part, an ethical approach to personalization is discussed, introducing the role of ethics in European legislation and Artificial Intelligence and discussing certain issues such as group privacy and nudging. In the third part, it is discussed the governance of personalization as a synthesis of regulation and ethics. Also, it is questioned the role of Public Service Media (PSM) in cultivating healthier social media.

4. An Introduction to the Normativity of Personalization

The regulation of the digital, the governance of the digital and the ethics of the digital (whether computer, AI, information or data ethics) are different normative approaches, complementary, but not to be confused with each other (Floridi, 2016) (see Figure 18). In this chapter, the normativity of personalization will be discussed from these different but intertwined perspectives. Prior to this, necessary clarifications are briefly given.

To begin with, 'digital regulation' has to be basically intended as "a system of rules elaborated and enforced through social or governmental institutions to regulate the behaviour of the relevant agents in the infosphere (p.3)". Instead, 'digital governance' is "the practice of establishing and implementing policies, procedures, and standards for the proper development, use and management of the infosphere (Ibid)". For example, through digital governance, a government agency or a company may (a) determine and control processes and methods used by data controllers and (b) devise effective procedures for decision-making and for the identification of accountabilities with respect to data-related processes. As such, digital governance may comprise guidelines and recommendations that overlap with, but are not identical to, digital regulation. Similarly, not every aspect

of digital regulation is a matter of digital governance and not every aspect of digital governance is a matter of digital regulation.

Then, there is digital ethics, understood as "the branch of ethics that studies and evaluates moral problems relating to data and information (including generation, recording, curation, processing, dissemination, sharing and use), algorithms (including AI, artificial agents, machine learning and robots) and corresponding practices and infrastructures (including responsible innovation, programming, hacking, professional codes and standards), in order to formulate and support morally good solutions (e.g. good conduct or good values) (Ibid, p.4)". Digital ethics shapes digital regulation and digital governance through the relation of moral evaluation.
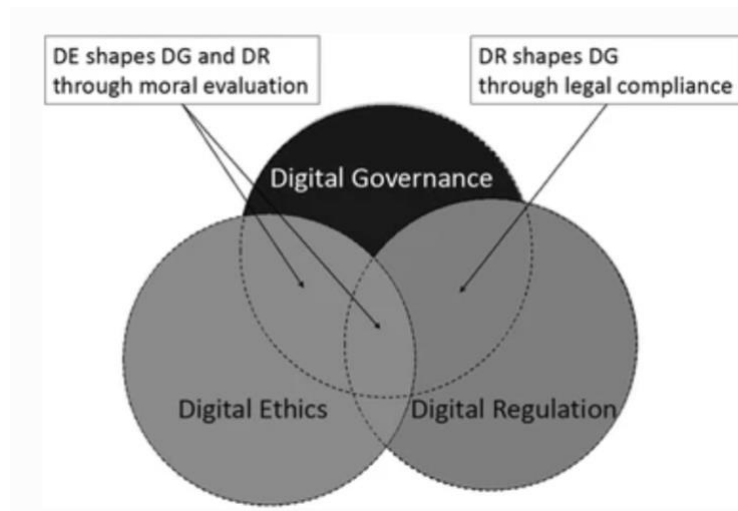
Figure 18. The relationship between digital ethics, digital regulation and digital governance (from Floridi, 2018).

Furthermore, digital ethics may be understood now in two ways, as *hard* and *soft ethics* (Floridi, 2018). Hard ethics is what we usually have in mind when discussing values, rights, duties and responsibilities—or, more broadly, what is morally right or wrong, and what ought or ought not to be done—in the course of formulating new regulations or challenging existing ones. In short, hard ethics is what makes or shapes the law. Thus, in the best scenario, lobbying in favour of some good legislation or to improve that which already exists can be a case of hard ethics. Instead, soft ethics covers the same normative ground as hard ethics, but it does so by considering what ought and ought not to be done over and above the existing regulation, not against it, or despite its scope, or to change it, or to by-pass it (e.g. in terms of self-regulation). We will further discuss this distinction

in Section 4.2.

Now, in the following section we firstly analyze *personalization regulation* (Section 4.1). This will be done in the context of European legislation, with a special focus to Human Rights, the E-Commerce Directive and the GDPR. Then, we explore relevant issues for both *personalization ethics* (Section 4.2) and *personalization governance* (Section 4.3) to provide a more comprehensive approach to the government of the subject of research. In the former case, we advocate, again, the importance of serendipity as a design principle for personalization algorithms and design, thus discussing it in the broader ethical debate. Moreover, an analysis of the relevance of inter-cultural ethics and, eventually, a critical approach to nudging are also introduced. In the latter case, we analyze specifically the role that public service media had in Europe, and what role it may have in the present day, especially regarding personalization practices and, more generally, in sustaining the traditional principles historically sustained by public service broadcasting.

4.1 Personalization Regulation

In this section, an overview over the debate on regulating platforms is introduced. For many reasons, in fact, the Internet – and more specifically the information intermediaries that enable personalization – is not an easy area to regulate. Since its inception, it has often been conceived as the epitome of decentralization. Internet, in fact, is more decentralized than any other communications systems that has occurred previously. Some of its features undermined the benchmarks that a century of international cooperation helped to build; the principles of territoriality, universality of values and effectiveness collide with the fluidity of information and data flows. This denaturalising effect suggests fundamental differences between the Internet as an object of regulation and any other domains that have been regulated internationally. Also, it is an area of constant fast-pacing changes not only for its infrastructure and architecture but, above all, for its capacity to trigger social, economic, and (geo)political changes.

As we have seen throughout the thesis, many Internet intermediaries have dominant positions in their particular markets – above all, Facebook for social networking sites and Youtube for video sharing – and utilize design to satisfy (or even nudge) users, and RSs to disseminate and determine personalized content, thus giving them immense power and

influence. This should come with certain responsibilities and, potentially, with certain liabilities. Yet, these platforms argued that they are technology companies rather than media companies, thereby avoiding media regulation and editorial responsibilities (Napoli and Caplan, 2017). How they design their platforms, how they allow content to flow, and how they agree to exchange information with competing platforms have also direct implications for both communication rights and innovation. The intermediation of content driven by personalization is related to several conditions of democracy; how people receive news, the articulation of relationships and associations; access to knowledge, and spaces for deliberation about issues of public concern. And as the Internet transcends national borders, the norms and laws in various jurisdictions are often in conflict and these private intermediaries must determine which government requests to comply with and which to ignore.

In this context of complex and global socio-technical systems like social media platforms it is therefore discussed what are the best regulatory approaches. Regulators are thus faced with two interrelated questions in this context: *who should regulate platforms* and *how they should be regulated*. In general, there are three main approaches that are usually discussed; traditional regulation, self-regulation and co-regulation who set and enforce different regulatory goals, standards and justifications (Hirsch, 2010: Finck, 2018):

- *Regulation*. Basically, proponents of government regulation argue that the desire for profits, coupled with the economic value of personal information, will inevitably lead private firms to collect a great deal of personal information online.

- *Self-regulation*. Self-regulation is a regulatory system in which "business representatives define and enforce standards for their sector with little or no government involvement" (Ibid, p.458). In general, critics of government regulation often argue that the market, either alone or in combination with self-regulation, will do a better job of protecting privacy as well as human rights, owing to; a) Industry members know better than anyone else and therefore are uniquely positioned to identify the most effective and efficient means of

protecting public values. b) If users ask for more privacy, for example, individual Internet businesses will enhance their competitive positions by responding to customer preferences for greater privacy, thereby leading to a more privacy friendly Web. Those who object to such a market solution focus on information asymmetries. Users are indeed often unaware of the collection of their personal information online or about complex technological dynamics and their social consequences.

- *Co-regulation*. "Co-regulation" is intended as a system that combines state- and non-state regulation. It is therefore a hybrid approach that has the potential to combine the strengths of the former two approaches. Co-regulation encompasses initiatives in which government and industry share responsibility for drafting and enforcing regulatory standards. Co-regulation is not a new phenomenon and can be found at various places in the regulatory landscape. American legal and policy scholars often use the term interchangeably with "*collaborative governance*." Proponents of co-regulation claim that it provides the flexibility of self-regulation while adding the supervision and rigor of government rules. Its critics, however, assert that co-regulation a) lacks transparency and accountability as compared to traditional notice-and-comment rulemaking, b) industry will not reveal insider knowledge to regulators but will instead use its informational upper hand to obtain weaker standards, so that c) enforcement responsibilities will often result in deals that favor industry and disfavor the public interest. In any case, there remain also serious questions about whether detailed government regulation – and any other approach – is appropriate for a fast-paced, complex part of the economy or even feasible in light of major political obstacles.

For further clarification the above denominations point towards various approaches to regulation, yet they operate on a spectrum. Also, all of them are already set in place to some extent. On the one hand, platforms are already self-regulating entities. They determine the terms and conditions of their intermediary function and define online and offline standards of behaviour. On the other hand, there is an information asymmetry

between platforms and policy-makers that naturally requires some forms of co-regulation. So, what is the best approach to regulate platform and its personalization systems, if any? This question will be further addressed after having analyzed the current European legal framework in Section 4.3.

Another significant issue is which jurisdiction and law are more suited for regulation and also whether there is a need for modification of existing legal frameworks or to create new legislation. Significantly, in the past years policymakers have undertaken several different initiatives to regulate social media RSs. This has included *horizontal instruments*, such as competition law and data protection law, which are not tailored to social media's personalization in particular but may still have some *spillover benefits* for its purposes (Leersen, 2020). Antitrust and consumer protection law already place some limits on platforms, and for the sake of this thesis, on the risks of personalization. Antitrust law could place limits on the ability of platforms to prioritize their own services and therefore discriminate other actors. Consumer protection law in most jurisdictions forbids covert advertising, which means that platforms have a duty to disclose whether content is being sponsored – i.e., whether it constitutes an advertisement rather than organic content.

Yet, more recently, some innovative sectoral proposals have been discussed and drafted. For example, the *Medienstaatsvertrag*, proposed in 2018 by the German broadcast authority, requires social media platforms to disclose the selection criteria that determine the sorting and presentation of content. The Dutch State Commission on the Parliamentary System has proposed an "independent entity" to monitor social media RSs to maintain diversity and avoid bias and, finally, the *EU Code of Practice on Disinformation* which requires signatories to "[d]ilute the visibility of disinformation by improving the findability of trustworthy content" and to "invest in technological means to prioritize relevant, authentic, and authoritative information where appropriate in search, feeds, or other automatically ranked distribution channels". Most of these instruments are rooted in media pluralism policy, but they also target other public interest considerations such as the combating of online disinformation.

Nevertheless, considering the special position of media and the preference for *self-regulation* following from media freedom, for some time it has been considered unlikely that the normative principles in the media context would directly translate into legal obligations. Such context has enabled social media mainstream platforms to actually

privatize internet governance (DeNardis and Hackl, 2015). So far, governments have been reluctant to take a strong lead. As of now, (supra)national governments have tried some forms of co-regulation, such as the above mentioned *EU Code of Practice on Disinformation*, in which the major tech companies – Facebook, Google and Twitter – pledged to work more actively to lessen the spread of disinformation and hate speech online. As previously discussed, all of this is non-binding and the rules can be interpreted rather loosely by these companies. More and more critics in academia and civil society are arguing that such forms of regulation do not provide sufficient incentives. Others argue that in designing and implementing interventions, one way to mitigate concerns regarding social media regulation is to incorporate *multistakeholder,* polycentric and co-regulatory elements. This perspective will be discussed in Section 4.3.1.

To better understand social media regulation and governance, in the next sections we attempt to answers the following questions; what is the role of human rights and secondary law on personalization and its practices at the European level? How is it ensured that profiling, processing and decisions made on their basis are legal, fair and non-discriminatory? Is the "terms and condition/terms of service" a sufficient and fair contractual paradigm? How the European legal approach tackles the challenges of personalization and the risks coming from the employment of increasingly sophisticated techniques of manipulation?

4.1.1 European Legislation

In this Section, we will elaborate on the European legal system's attempt to – or should try to – regulate personalization and its related technologies and practices, in particular referring to legal debates on (news) RSs, behavioural advertising and political micro-targeting, throughout a comprehensive legal lens.

4.1.2 A Human Rights Perspective

In order to start with the analysis of the existing legal framework for the regulation of online personalization, a number of provisions and legal acts in the areas of fundamental human rights as well as within primary and secondary law of the EU need consideration. As Leenes et al. (2017) argue, the common heritage of European fundamental rights and values can serve as an anchor point for regulatory discussions. Fundamental rights as the highest legal assets within democratic states -both in their (subjective) expression as

defensive rights of natural and legal persons and partly as (objective) guarantees- must be safeguarded as the basis of every legal framework and in every legislative and regulatory activity. In some cases, they can oblige the states bound to them to (certain) actions in order to counteract existing circumstances that cannot be reconciled with fundamental and human rights and to eliminate existing impairments. Fundamental human rights involved in personalization and its practices are, above all; dignity, the right to privacy, freedom of thought, freedom of expression, right to information, freedom of media and others.

In this context, the European Convention of Human Rights (EHCR) has an important role and impact on the EU in two respects; on the one hand, the Member States as Convention States are bound to the ECHR as a source of international law, also in the implementation of Union law, which at the same time means that the Member States are guarantors of measures taken by the Union. On the other hand, even after the adoption of the Charter of Fundamental Rights of the European Union (CFR), the ECHR is still one of the most relevant legal references within Union law. Indeed, the Court of Justice of the European Union (CJEU) recognizes in its decisions that "the principles on which the European Convention for the Protection of Human Rights and Fundamental Freedoms is based must be taken into consideration in community law"[15], and it states that "fundamental rights form an integral part of the general principles of the law, the observance of which is ensured by the Court. In safeguarding those rights, the Court has to look to the constitutional traditions common to the Member States, so that measures which are incompatible with the fundamental rights recognized by the constitutions of those States may not find acceptance in the Community"[16]. This means that the CJEU therefore incorporates both the norms of the ECHR and the case law of the European Court of Human Rights (ECtHR) into its decisions, in particular within the framework of the justification of infringements of fundamental rights guaranteed in the CFR. In principle, the ECHR, the CFR and national constitutional law complement each other cumulatively.

An overriding principle and foundation of all other fundamental rights is human dignity - Article 1 of CFR – which is inviolable and must be respected and protected. Although

---

[15] CJEU, judgement of 15.5.1986, C-222/84, Marguerite Johnston v Chief Constable of the Royal Ulster Constabulary, para. 18.
[16] CJEU, judgement of 13.7.1989, C-5/88, Hubert Wachauf v Bundesamt für Ernährung und Forstwirtschaft, para. 17.

the ECHR does not explicitly mention human dignity, the ECtHR assumes that the principle of respect for human dignity underlies all Convention guarantees and that "[t]he very essence of the Convention is respect for human dignity and human freedom"[17]. Human dignity is both a fundamental right with subjective guarantee content and a principle under objective law. However, it has always been difficult to solidify the content of human dignity or define it. It can be stated here that at least the minimum core of human dignity consists in the fact that every human being possesses an intrinsic worth, merely by being human, which should be recognized and respected by others. In the area of online content, there are many conceivable possibilities for violating human dignity. This applies, for example, to algorithmic manipulation or content that contains pornography or depictions of violence.

The right to freedom of thought, conscience, and religion is protected by Article 9 of the EHCR and Article 10 of the CFR. Various philosophies and belief systems fall within the ambit of the right. Freedom of thought strengthens other fundamental rights. Without freedom of thought, freedom of expression is meaningless. One cannot speak freely if one cannot think freely. At the same time, freedom of thought is reinforced by other fundamental rights. Freedom of thought is possible only with effective freedom to receive information. Freedom of thought also overlaps with freedom of opinion; if a belief is not sufficiently serious and coherent to attract protection of freedom of thought, at least it is protected by freedom of opinion. For beliefs to attract the protection of freedom of thought, conscience, and religion, they should attain a certain level of cogency, seriousness, cohesion, and importance – and conspiracism? The Strasbourg Court held that freedom of thought, conscience, and religion protects against "indoctrination of religion by the State" (Angeleni v. Sweden, 1986, p. 48). So, what about potential indoctrination by algorithms? This issue will be discussed further.

The freedom to hold opinions as a component of freedom of expression, Article 10 of the ECHR and Article 11 of the CFR. In general, news personalization systems have a significantly lesser effect on opinion formation than, for example, ideology conversion systems because personalization systems apparently do not punish users for not changing their opinion. Furthermore, it is part of the public task of the news media to inform people, and it is insurmountable that they influence public opinion. The question is indeed at what

---

[17] ECtHR, judgement of 29.4.2002, no. 2346/02, para. 65.

point news RSs interfere with free opinion formation and ultimately become coercive. The information and technologies that news media use for personalization give them the power to influence our opinions. For example, if you are presented with a personalized news feed, you might think that "what you see is all there is" and jump to conclusions (Kahneman, 2011, p. 85). One factor to distinguish legitimate influence on public opinion from coercion of opinions is transparency. Users often ignore that their social media news feeds are tailored to their interests and preferences (Eslami et al., 2015). The lack of transparency reinforces the risks of selective exposure. Overall, we can claim that most people are not in a "filter bubble" or "echo chamber". Still, some groups might be more vulnerable to end up in a filter bubble, and others are more susceptible to receive less diverse online news on certain issues (Bodó et al., 2019). In any case, Article 10 ECHR may nevertheless prove an important point of departure to realize democratic values in the personalized media landscape (as we will discuss in Section 4.1.5.1) (Eskens et al., 2017).

Last but not least, privacy is protected by Article 8 of the European Convention on Human Rights and Article 7 of the Charter of Fundamental Rights of the EU. The Convention and Charter use the term "private life" instead of "privacy," but these terms are commonly assumed to be the same. Instead of providing a static definition, the Court has brought different interests and rights under the notion of privacy. Overall, the right to privacy has a wide scope of application and "goes beyond concealed personal information" (Purtova, 2010, p. 186).

Together, these rights protect what Eskens (2020) calls the *personal information sphere*. The personal information sphere is the domain where people can determine for themselves how they interact with information about the world and how other people may interact with information about themselves. This is a form of control that is different from the kind of control enabled by data protection law, which focuses on consent, transparency, and data access rights. The notion of a personal information sphere resembles the concept of "intellectual privacy" (Richards, 2015). Intellectual privacy is "a zone of protection that guards our ability to make up our minds freely" so that we can prepare ourselves to exercise our freedom of expression rights (Ibid, p. 95). The difference between the two concepts is that the personal information sphere arises from European fundamental rights, whereas intellectual privacy is built on the U.S. First

Amendment. Furthermore, due to the elaborate European fundamental rights framework, the personal information sphere encompasses more rights and freedoms and is more inward-looking than intellectual privacy, which is more outward-looking.

Following Eskens (2020), we agree that the notion of a personal information sphere shows how privacy interferences affect other fundamental rights that enable people to develop their sense of self and relate to the world. The notion is reminiscent of personality rights. Yet, the difference between personality rights and the personal information sphere is that the latter encompass all kinds and directions of communication about all kinds of private and public matters, whereas personality rights are mainly about the communication of one's image from the rights holder to the outer world.

4.1.3 E-Commerce Directive

The diversity of content, dissemination channels and problems on the Internet is matched by the diversity of issues covered by legislative acts. The E-Commerce Directive (ECD) is the main secondary legislation to be dealt with in the context of this study because of the horizontal approach it follows. Adopted in 2000, the ECD was intended to create, for the first time, a coherent framework for Internet commerce. As key areas of regulation, the proposal identifies the responsibility of intermediaries, electronic contracts, commercial communications, transparency and enforcement, and the country-of-origin principle. The primary objective of the liability rules is to prevent distortion of competition between cross-border services through different civil and criminal responsibilities.

In the EU, the E-Commerce Directive provides for certain liability protections in relation to user-generated content. Specifically for providers of information society services who are acting as intermediary service providers (as defined in the Regulations Directive as "any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services") (Cobbe and Singh, 2019). To be acting as an intermediary service provider, it must be neutrally providing a service by automatic, technical, and passive means. The E-Commerce Directive, in recital 42, states that the Directive's liability protections are limited to situations where the service provider has engaged in an activity of a "mere technical, automatic and passive nature, which implies … neither knowledge of nor control over the information which is transmitted or stored." As such, in relation to content recommending of user-generated

content, service providers are unlikely, in most cases, to have knowledge of the content itself. Rather, they will have knowledge of metadata about the content (e.g. 'likes' or 'shares'). Yet, service providers do exercise control over content, and this is the very point of recommending; service providers exercise such control in order to show people what they want them to see in order to drive engagement, profit, and market position. The normative nature of recommender systems – the fact that they enable platforms to exercise control over content distribution in pursuit of their own goals – is the reason for their use.

Consequently, in relation to recommending, platforms cannot be intermediary service providers and are operating beyond the limits of the liability protections provided by the E-Commerce Directive. The effect of this analysis would be that recommending is not an activity that is covered by any of the E-Commerce Directive's liability protections. Recommending is an activity engaged in by service which developed in the years after the E-Commerce Directive was passed. However, this distinct activity of recommending is also not otherwise covered under any particular regulation beyond the general law. This means that recommending falls into a significant and consequential gap in the current legal regime. Following this analysis, service providers would have, under the E-Commerce Directive or any other current law, no special protection against liability for recommended illegal content or activity on their platforms.

The straightforward response would be to provide for a substantively similar regime to that already established for hosting, whereby service providers are shielded from liability provided they expeditiously remove illegal recommended content once they become aware of it. However, recommending is a much less neutral, more involved, and a more active form of service provider activity than simply hosting. While liability protection for recommending is perhaps desirable, this should come with some responsibilities beyond simply removing illegal content expeditiously upon knowledge of it's illegal nature. These responsibilities would sit alongside and complement other applicable legal frameworks, such as around data protection or non-discrimination. Regulating the usage of recommender systems in social media therefore provides an opportunity to establish a more inclusive framework of principles, requirements, and limitations within which the discrete activity of recommending can responsibly be undertaken.

A new relevant horizontal instrument in this context is the Regulation on Promoting Fairness and Transparency for Business Users of Online Intermediation Services (Platform-to-Business Regulation) (Leersen, 2020). This Regulation does not affect consumers of social media content but rather producers, who are granted certain notice rights in relation to recommender systems under Article 5. This provision requires platforms to disclose, inter alia, "the characteristics of the goods and services offered to consumers through the online intermediation services or the online search engine". For sophisticated content providers who rely on social media, such as newspapers and other media outlets, this could be an additional way to adapt to changes in recommendation algorithms, and potentially to detect unlawful or abusive forms of discrimination.

4.1.4 E-privacy Directive

In previous years, the EU has adopted some provisions that give consumers the power to manage their personal data and not to be subject to automated decision-making such as in personalization practices. The right to data portability[18] envisaged in the new General Data Protection Regulation (GDPR), as well as the e-Privacy regulation[19], and also the "retrieve them all" provision of the proposed Digital Content Directive, are all tools whereby digital consumers will supposedly have the chance to decide who should use their data to offer them the goods and services that they want. These regulatory interventions bring to the fore a reshaping of the traditional landscape of the consumer protection rules, thereby providing a more comprehensive vision of "data consumer law". In actuality, they grant consumers several rights, such as the right to transfer data from one controller to another and the right to retrieve any data produced or generated through their use of a platform. They are expected to re-balance the relationship between data subjects and data controllers and to encourage competition between companies. They represent a new paradigm that abandons a purely protective and paternalistic regulation

---

[18] Data portability refers to the ability to move, copy or transfer data easily from one database, storage or IT environment to another. To make an example, move one's Facebook profile to another social network.

[19] Notice that the E-Privacy Regulation should be treated as *lex specialis* in relation to the GDPR. However, the enforcement mechanisms of GDPR and E-privacy Regulation remain the same. In a further step the question about the relationship between the GDPR and the ePrivacy Directive is to be raised – in the ordered brevity. Art. 95 GDPR answers this question expressly in such a way that the GDPR "shallnot impose additional obligations on natural or legal persons in relation toprocessing in connection with the provision of publicly available electroniccommunications services in public communication networks in theUnion in relation to matters for which they are subject to specific obligationswith the same objective set out in Directive 2002/58/EC". This meansthat there is a lex generalis-lex specialis relationship where the special provisionsof ePrivacy Directive prevail over the general rules of the GDPR inareas which they specifically seek to regulate.

focused only on consumers' weaknesses, with the aim of experimenting with a more proactive approach (Colangelo et al., 2019).

Without delving into detail on the scope of the ePrivacy Directive, it should be noted that – in the light of the objective of this Directive, to ensure the protection of fundamental rights and freedoms of the public when they make use of electronic communication networks – at least some of its provisions also apply on providers of information technology networks (in particular Art. 5(3), 9 and 13). Regarding the dissemination of content in the online environment this means that Art. 5 para. 3 and 13 of the ePrivacy Directive apply to website operators (e.g. for cookies[20]) or other businesses (e.g. for direct marketing). For example, website operators have to obtain the (active) consent of the user if they want to store (and subsequently use for e.g. personalized advertising) certain cookies on the end devices of website users, which in particular enables web tracking and is therefore essential for many advertising-based business models on the Internet.

Discussions about the legal requirements for behavioural targeting, for example, often focus on the e-Privacy Directive's consent requirement for tracking cookies and similar technologies. But in many cases, the Data Protection Directive and the right to protection of personal data also applies, namely when behavioural targeting entails the processing of personal data. The European Union Charter of Fundamental Rights only allows personal data processing if the data controller has a legal basis for the processing, such as consent. However, Borgesius (2015) argues that Article 5(3) of the e-Privacy Directive does not provide a legal basis for the processing of personal data. Sometimes called the 'cookie provision', it requires consent for storing or accessing information on a device of a user or subscriber. Some companies suggest that they can use an opt-out system to comply with Article 5(3) of the e-Privacy Directive. However, even if companies could obtain consent for cookies, companies would generally be required to obtain the data subject's unambiguous consent if they process personal data for behavioural targeting. The e-Privacy Directive's preamble says that a user's device and its contents are part of the user's private sphere. 'Terminal equipment of users of electronic communications networks and any information stored on such equipment are part of the private sphere of

---

[20] Ads publishers often use tracking cookies, small text files that are stored on a user's computer to recognize that computer. With a tracking cookie, it is possible to follow an Internet user across all websites on which it serves ads. Through almost every popular website, tracking cookies are stored; through some websites dozens of them. Behavioural targeting companies use many other tracking technologies as well, such as flash cookies and device fingerprinting.

the users requiring protection under the European Convention for the Protection of Human Rights and Fundamental Freedoms.'

In actuality, the ePrivacy Directive contains no provisions on law enforcement, sanctions and the establishment of supervisory authorities. Rather, it is up to the Member States (who must ensure that the requirements are implemented) to decide how they are to be implemented. In addition, there is no forum under the ePrivacy Directive, such as the ERGA or EDPB, where an exchange could take place on the implementation and enforcement of the requirements of the Directive. Yet, the ePrivacy Directive is also currently undergoing a reform process. Outside the regulatory framework of the ePrivacy Directive, the GDPR applies. Due to its character as a regulation, it is directly applicable in all Member States and has thus strongly harmonized the data protection laws throughout Europe.

4.1.5 General Data Protection Regulation

The General Data Protection Regulation (GDPR) grants platform users a bundle of individual rights. Article 5 lists 'transparency' as one of the Regulation's key principles, and users are granted information and notice rights about personal data processing under Article 12-14. More specifically, under Article 22, data subjects may under certain circumstances have the right to opt out of such automated decisions, and also enjoy a bundle of information rights collectively known as the "right to an explanation". Van Drunen, Helberger and Bastian (2019) argue that these provisions should be interpreted contextually as a means to empower data subjects in their capacity as news consumers. In fact, users of RSs are entitled to a range of information, for example, the parties able to influence editorial decisions, the profiles that the algorithms construct about them, and the algorithm's metrics and factors. In such a reading, the GDPR could in theory be a source of insights about platform gatekeeping decisions. It remains debatable whether such access rights would find much usage with the average end user.

Furthermore, critics suggest that the GDPR – one of the most lobbied pieces of EU legislation to date[21] (Edwards and Veale, 2017) – delivers personalization to social media platforms with the utmost ease (Kohl et al., 2019). Firstly, by shifting the prerequisite for

---

[21] For example, Facebook lobbying of GDPR has been analyzed. It seems that even if in most areas not, on some very specific issues, the GDPR was shifted towards Facebook's ideals (Johansen, 2019).

more expansive (re)uses of personal data from anonymization to 'pseudonymization' which still allows for some form of re-identification.[22] In fact, although anonymized data is effective in protecting privacy, much analytical value of the data is lost through anonymization (which is relevant for personalization purposes). Secondly, the GDPR facilitates personalization by making the collection and processing/use of personal data essentially a matter of informational self-determination. This emphasis suggests to users that all that is at stake in data protection is their own personal interest, whereas also fundamental collective public goods are actually at stake, such as deliberative democracy. Moreover, the GDPR lacks a precise language, in addition to explicit and well-defined rights and safeguards (Wachter et al., 2017). A number of provisions may lead to confusion, enforcement gaps or asymmetrical interpretations. This is understandable given that the reform of EU data protection is ongoing and needs further guidelines.

Furthermore, clarification is required, as the word personalization contains the adjective personal, this implies that personalization should, at least in part, be based on personal data. Yet, researchers in personalization have largely overlooked the crucial connection between personalization and legal definitions of personal data (Greene and Shmueli, 2019). Clearly, a source of complexity is that different legal regimes define personal data differently. In this case, the GDPR defines personal data as "any information relating to an identified or identifiable natural person ('data subject')" (Article 4). Such a broad definition means personal data could constitute anything from browser cookies, to location data, to even a combination of non-sensitive measurements, if there are sufficiently few or unique observations to single out individuals. Ultimately, it is context that determines whether data are personal data (Greene et al., 2019).

The focus of the following section is specifically on the most relevant GDPR's articles affecting fundamental principles to guarantee a fair personalization dependent, above all, by 'profiling' which is a relatively novel concept in European data protection regulation (Art 4(4)). It refers to both the creation and the use of profiles. By virtue of deriving, inferring or predicting information, practices of profiling generate personal and sensitive data. The rights to erasure (Art 17) and restriction of processing (Art 18[23]) are then useful

---

[22] Pseudonymization is a technique that is used to reduce the chance that personal data records and identifiers lead to the identification of the data subject whom they belong

[23] As many other articles, more guidance is needed in practice. For instance, the UK's Information Commisoner Office recently released a code that provides more guidance to appropriately safeguard children's personal data. In the case

forms of redress in the context of unlawful profiling techniques. Further guidance, however, is needed to clearly set out the articles' scopes of application. This is also true for highly debated articles that we are going now to briefly analyze, namely Articles 13-15 and Article 22.

4.1.5.1 Transparency

In the context of RSs, concerns over transparency often refer to the specific algorithm used to produce recommendations, the outputs (i.e. recommendations) and the inputs (e.g. user content, behavioral data, etc.). In addition, transparency can also refer to individuals and organizational structures involved in designing and operating this system. As Ananny and Crawford (2018) argue transparency in algorithmic systems should take into account "not just code and data but an assemblage of human and non-human actors (p.983)." Given that research on mainstream social media's human agents involved in personalization are rather scarce, in this section we focus only on the formers, specifically from a data protection perspective.

In the present day, personalization RSs – and, more generally, algorithmic systems – have the potential to be significantly more complicated than traditional formalized decision-making systems. This led the requests for greater transparency from lawyers, social and computer scientists, and this has been reflected in both legislative developments and proposals across the world. Transparency is indeed often assumed to be an ideal for democratic oversight and political discourse in democracies and it is generally defined with respect to "the availability of information, the conditions of accessibility and how the information…may pragmatically or epistemically support the user's decision-making process" (Turilli and Floridi, 2009, p.106). This is significant when regarding decisions that are extremely complex, black-boxed and cause several individual and social concerns. Interestingly, there has also been considerable debate as to whether more complex algorithmic systems are actually needed or useful. Some research has indeed focused on creating simpler systems with comparable utility to more complex ones.

---

of profiling, it calls for a differentiation between different types of profiling for different purposes so as to offer separate privacy settings for each different type of profiling.

Burrell (2016) distinguishes three barriers to transparency: 1) Intentional concealment on the part of corporations or other institutions, where decision making procedures are kept from public scrutiny. 2) Gaps in technical literacy which mean that, for most people, simply having access to underlying code is insufficient and 3) a mismatch between the mathematical optimization in the high-dimensionality characteristic of machine learning and the demands of human-scale reasoning and styles of interpretation. In addition, it has to be considered whether individuals can tell that a decision or personalization measure is taken algorithmically at all, especially in a world of 'ambient intelligence' it can be difficult for an individual to know when data relating to them is being processed, or when their environment is being altered. Yet, transparency can be reached in many ways. Above all, throughout the explicability of the system and the process, and this can take many forms as we will highlight in the next sections.

4.1.5.2 Explicability

Strictly related to transparency is the right to an explanation of the decision made by personalization algorithms. Harsh debates have been conducted not only as to whether this right is contained in the GDPR but, also, whether this right can even exist. In other words, the decisions of algorithms are not always clear-cut, explicable or even meaningful to users. While a right to an explanation appears attractive, and has significant public support, the utility of it is spurious. Some authors have expressed concern that it may provide a meaningless, non-actionable form of explanation that does little more to help deal with algorithmic harms than the privacy policies individuals often ignore (Edwards and Veale, 2017).

Especially relevant to profiling, there is the right to be informed (Art 13) and the right of access (Art 14). In particular, Articles 13(2)(f) and 14(2)(g) require data controllers to provide specific information about automated decision-making, based solely on automated processing, including profiling, that produces legal or similarly significant effects, namely: 1) the existence of automated decision-making, including profiling; 2) meaningful information about the logic involved; and 3) the significance and envisaged consequences of such processing for the data subject.

Article 15(1)(h) uses identical language as of the above articles and entitles data subjects the right of access to information about solely automated decision-making,

including profiling. However, some key expressions in Articles 13-14, specifically "meaningful information about the logic involved" as well as "the significance and the envisaged consequences" (Art 13(2)(f)), needs to be interpreted to provide data subjects with the information necessary to understand and challenge profiling and automated individual decision-making. As a result, the right to explanation has been interpreted in two drastically different ways: as an ex-ante general explanation about system functionality or as an ex-post explanation of a specific decision (Art 15). Yet, in the interest of strong consumer protection, meaningful information must be sufficient to answer questions that the data subject might have before they consent to the processing (notification) and after a decision has been made (right of access).

A right to explanation is thus not explicitly mentioned in the GDPR. Legal basis has been detected (Wachter et al., 2017). In particular, Recital 71 states that data subjects have the right 'to obtain an explanation'. Yet, the legal status of recitals is debated but, in general, they only provide guidance to interpret the articles so they are not considered legally binding. This is a critical gap in transparency and accountability (Edwards and Veale, 2017).

4.1.5.3 Non-discrimination

In the Article 22(1) of the GDPR there is additional safeguards against one specific application of profiling, namely the case of automated individual decision-making that fulfils is "based solely on automated processing" and produces "legal effects concerning him or her or similarly significantly affects him or her". Profiling can indeed form the basis of decision-making that is both automated and produces significant effects, in particular discrimination. A right to non-discrimination is, in fact, deeply embedded in the normative framework that underlies the EU and the use of algorithmic profiling for the allocation of resources is, to some extent, inherently discriminatory (Goodman and Flaxman, 2016). In this sense, Article 22 is set. There are, however, several ambiguities that must be discussed.

Firstly, the wording of the "right not to be subject to automated decision-making" can be interpreted as either a prohibition or a right to object. This ambiguity has existed since the Data Protection Directive 1995 (Wachter et al., 2017), but resolving it is nowadays critical. Since profiling and automated decision-making often occur without the

awareness of those affected, data subjects may not be able to effectively exercise their right to object. Moreover, Article 22 only applies to decisions that are "based solely" on automated processing, including profiling. Since "based solely" is not further defined in the regulation, the regulation allows for an interpretation that excludes any human involvement whatsoever. This would render the article inapplicable to many current practices of automated decision-making and there is the risk that the controller may fabricate human involvement. Ultimately, the complexity and multifaceted nature of algorithmic discrimination suggests that appropriate solutions will require an understanding of how it arises in practice, and this highlights the need for human-intelligible explanations of algorithmic decision making.

Furthermore, paragraph 71 and Article 22(4) specifically address discrimination from profiling that makes use of sensitive data. Goodman and Flaxman (2016) broadly questioned the interpretation of the wording 'sensitive data' and that clarification is significant.

Finally, the wording of both "based solely on automated processing" and "legal and "similarly significant effects" leaves room for interpretation and should be clarified to offer the strongest possible protection for data subjects. So far, the Working Party has opted for a nuanced subjective interpretation of "significant effects" that runs the risk of placing the burden of proof on the data subject (Kaltheuner and Bietti, 2018). This guidance raises several important questions on who defines whether a targeted data subject is vulnerable.

4.1.5.4 Auditing

Auditing is one promising mechanism for achieving transparency (Mittelstadt, 2016). For all types of algorithms, auditing is a necessary precondition to verify correct functioning. For platforms that mediate political discourse, auditing can create a procedural record to demonstrate bias against a particular group. Auditing can also help to explain how citizens are profiled and the values prioritized in content displayed to them. It allows for prediction of results from new inputs and explanation of the rationale behind decisions. Yet, many epistemic, technical, and practical challenges must be overcome (Burrell, 2016). Firstly, a political right to transparency may undermine the privacy of data subjects and autonomy and competitive advantage of service providers, or even national security.

Secondly, the rationale of an algorithm can be epistemically inaccessible, rendering the legitimacy of decisions difficult to challenge. Nonetheless, algorithm auditing may be approaching at pace and the belief that highly complex algorithms are incomprehensible to human observers should not be an excuse to surrender robust political discourse.

Furthermore, developing practical methods for algorithmic auditing is highly needed. For example, Tutt (2016) suggests that a regulatory agency for algorithms may be required, and this agency can "classify algorithms into types based on their predictability, explainability, and general intelligence" (p.15) to determine what must be regulated. Some researchers even see promise in the collective use of individual rights to achieve greater societal transparency (e.g. group privacy). In any case, GDPR requires data processors to maintain a relationship with data subjects and explain the logic of automated decision making when questioned (Art 13, 14 and 15). The regulation might prove a much-needed impetus for algorithmic auditing. However, with opacity, implementing transparency and the right to an explanation in a practically useful form for data subjects is extremely difficult, necessary yet likely insufficient.

4.1.5.5 Public Disclosures

Another source of transparency could be represented through 'public disclosures' that could include the documentation of recommendation outputs and their audiences; content-specific ranking decisions and other interventions by the operator's in RSs performance and the organizational structures that control RSs (Leersen, 2020). These might be fundamental sources for external accountability.

The main drawback of public records, compared to confidential disclosures such as data sharing partnerships and government auditing, is their limitations in sharing sensitive data: public disclosures would indeed require a trustless design that pre-empts abuse by malicious actors, to contend with threats to user privacy, and, according to platforms, the integrity of the service (i.e. by enabling third parties to 'game' the algorithm). Privacy-by-design techniques such as anonymization and differential privacy can aid in mitigating these concerns. Ultimately, publicity places hard limits on what can be disclosed and thus on the ultimate research utility of public disclosures. However, public disclosures also have an important advantage over data partnerships: by simply making information publicly accessible, public disclosures would be available to every researcher with the

time and interest, even to those without institutional means of accreditation. In terms of format, public disclosures about recommender systems should include real-time, high-level, anonymized data access through public APIs and browser interfaces. Therefore, public records could be instrumental for purposes of real-time, high-level monitoring by media watchdogs such as journalists, activists, and NGOs. As Leersen argues (2020), these may not suffice to conclusively demonstrate bias or discrimination in RSs, but at a minimum they could offer a starting point for such investigations and serve as a first-warning system for more targeted efforts. An example of a similar endeavor is the project Algotransparency.org (see Figure 19), which covers YouTube recommendations made by 1000 selected channels and on a limited set of keywords. It has already led to important insights about social media RSs.
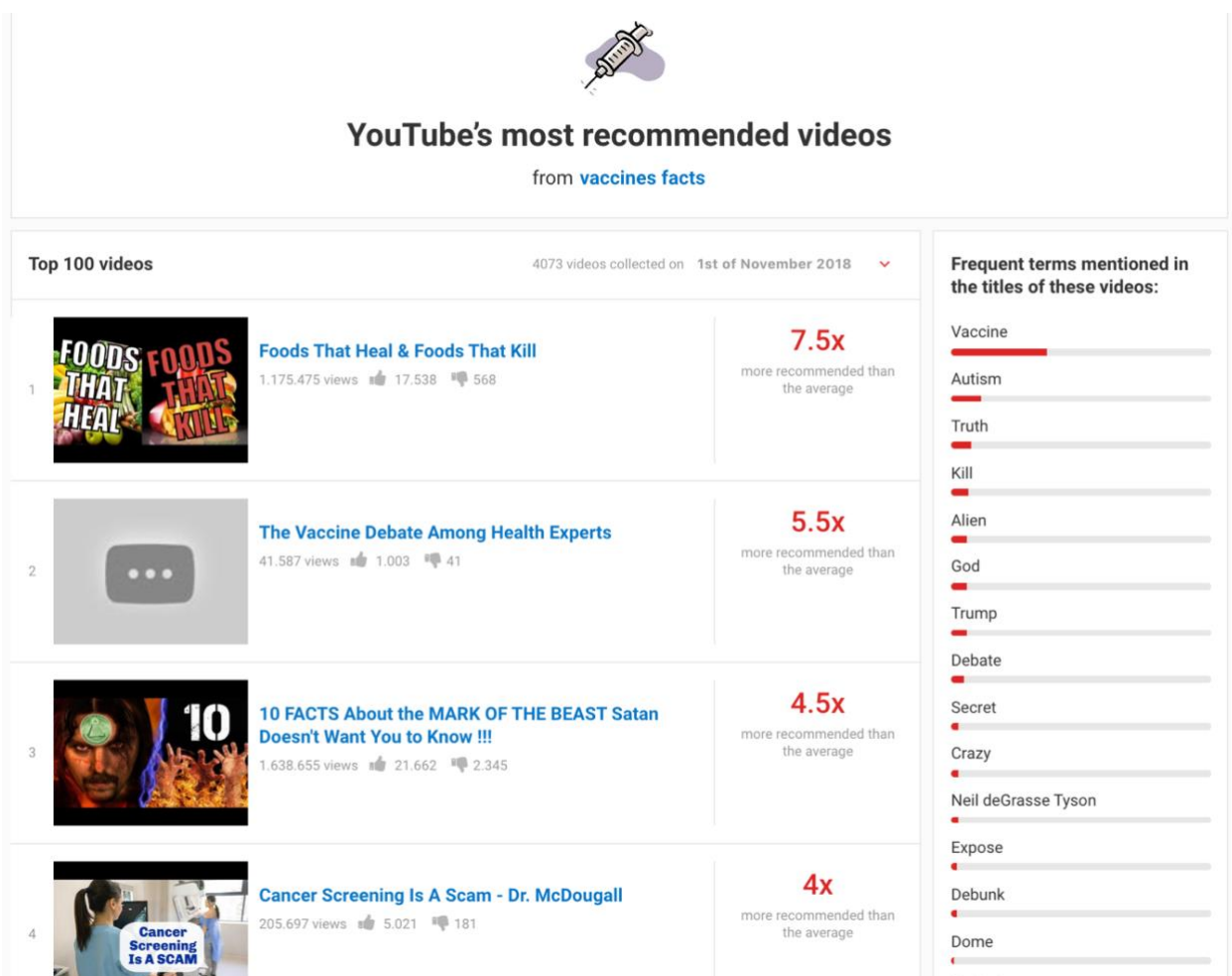
Figure 19. Youtube's most recommended videos on the website Algortransparency.org. (screenshot, May, 2019).

Given the complexity of designing privacy-compliant disclosure standards, rules for public disclosure will be difficult to capture exhaustively in one-size-fits-all legislation. Not only are social media RSs technically complex, but they are also heterogeneous; each has unique features (types of posts and formats, engagement metrics, et cetera) which may require unique forms of documentation and privacy safeguards. Therefore, a more feasible model may be to set broad disclosure standards in legislation, enforced by an independent public authority with the power to offer case-by-case guidance for particular platforms on what types of disclosures are necessary.

4.1.5.6 A Short Discussion on the GDPR

To conclude, GDPR defines novel rights for data subjects and duties for data controllers. Along with e-Privacy regulation it strengthens 'data consumer protection'. Users can decide whether to enter into a contract, be informed, access the data that is generated, receive information about the logic involved and not to be a subject to automated decision-making based solely on automated processing. When a data subject enters into a contract, however, gives up to some of these rights. As a member of the European Parliament claimed: "The GDPR is not only a legislative piece, it's like a textbook, which is teaching us how to understand ourselves in this data world and how to understand what are the responsibilities of others and what are the rules which is governing in this world" (in Hasselbalch, p.9, 2019). Although, users seem to be insufficiently empowered by the GDPR with regards to the logics involved in the personalized (news) services. At first glance data-driven personalization may appear to be only a matter of data protection law, however, the analysis of automated inferences, predictions or decisions more often lies outside of it (Kohl et al., 2019). In essence, data protection law focuses on 'inputs' rather than 'outputs', that are mostly out of its scope.

As such, data protection law shows some limitations when it comes to the actual consumption of information. Yet, the application of consumer protection law to data-related commercial practices can certainly add to the protection offered by data protection law (Helberger et al., 2017). The complex interplay between data protection and consumer law needs to be further analyzed in order to understand whether and how they might complement each other so as to be able to prevent the risks of media personalization. There is indeed a fundamental need for interdisciplinary work, not only

across academics and practitioners, but also between different legal jurisdictions and across different disciplines. GDPR, for example, does not provide any duty for the data controller towards the information a data subject might consume. In this respect, it is media law that is directly responsible.

4.1.5 Media Pluralism

Throughout the twentieth century, state institutions in Western Europe were primarily responsible for the organization of public space and for safeguarding public values. This societal arrangement has since come under increasing pressure due to economic liberalization and the privatization of public institutions and services. The rise of digital intermediaries in the form of online platforms is accelerating and further complicating this trend. In many cases they intensify the pressure of the market on important public values, such as the diversity of media content.

The concepts of media pluralism and diversity are indeed well-established ideas in Western Europe's media research and policy since the 1960s (Hoffmann et al., 2015). According to McQuail (2007), diversity is "the most potent concept in communication policy in modern times" (p. 41). Media pluralism and diversity are considered a crucial foundation for both democratic societies and an enlightened public. The European Commission (2007) explicitly recognized diversity as a policy goal that "embraces a number of aspects, such as diversity of ownership, variety in the sources of information and in the range of contents available" (p. 5). Art. 3 para. 3 subpara. 4 TEU plays a role in the context of media policy as a whole and therefore in the context of this study. According to this, the EU shall respect its rich cultural and linguistic diversity and shall ensure that Europe's cultural heritage is safeguarded and enhanced. This provision therefore addresses the role of media as economic and cultural heritage in safeguarding diversity. Thereby the EU's objective is not to create a uniform European culture but rather to preserve existing cultural diversity, which draws its strengths precisely from its historically grown diversity.

Pursuant to Art. 85 para. 1 GDPR, EU Member States are obliged to balance the right to the protection of personal data with the right to freedom of expression and information, including processing for journalistic purposes, by means of legal provisions. This balancing act i.e. the assessment of the extent to which the balancing of data subjects'

interests and media/information interests requires modifications with regard to data protection provisions, is the responsibility of the Member States, which, however, have a wide margin to act. More generally, a number of Council of Europe recommendations emphasize the importance of informing and empowering users. For instance, Recommendation 2018-1 on media pluralism and transparency of media ownership calls on states to encourage platforms to "provide clear information to users on how to find, access and derive maximum benefit from the wide range of content that is available". The Council of Europe has also developed standards on the need for government oversight of content recommenders, emphasizing diversity or pluralism as a guiding principle. Their Committee of Ministers has recommended that "[s]tates should encourage social media, media, search and recommendation engines and other intermediaries which use algorithms … to engage in open, independent, transparent and participatory initiatives that seek to improve these distribution processes in order to enhance users' effective exposure to the broadest possible diversity of media content."

While the terms pluralism and diversity are frequently used interchangeably, the difficulty of distinguishing one from the other also hints at the complexity of the underlying phenomena: media diversity and pluralism can both describe objects as varied as media outlets or platforms, ownership, sources, content, ideas, or forms. The term pluralism is commonly used when describing various media outlets or diverse ownership structures. Diversity, in turn, more frequently describes the variety of content available to or consumed by citizens. The latter is assumed to be an outcome of the former, although this relationship can still be considered contentious in empirical research. From the first perspective, diversity is generated by a plurality of available media or content, and the latter conceptualizes media diversity as a result of user behavior (as we already analyzed in Section 2.2.5).

Recent debate on the public responsibility of social media platforms pivots on the question of whether or not platforms can be held accountable for the content shared through them, legally and morally (Helberger et al., 2018). From the legal point of view, this discussion is grounded in the host-editor, namely either social media qualify as hosts, with the consequence that they fall under the European e-Commerce regime, or they are categorized as editors, having full legal responsibility for what is shared through their platforms (Council of Europe 2011, para. 29–36). Yet, the actual role and capacities of

social media's platforms to prevent certain undesirable outcomes or to contribute to their realization. Of course, platforms fundamentally shape user activity, yet they do not determine this activity. Many of the problems with media pluralism and diversity are, to some extent, user-driven. For similar reasons, at least part of the solution to potential public policy challenges lies with the users. The current focus in law on allocating responsibility to one central party – editor, data controller, or the supplier of a service – is primarily due to the fact that this central actor is the source of potential risk or harm, or the controller of a resource that can give rise to legal responsibilities. Yet, multiple actors are effectively responsible.

4.1.5.1 The 'Right to Receive Information'

Taking into account the unique role of the media in democratic societies as a source of information, a platform for deliberation, and a critical watchdog, it becomes evident that freedom of expression is a central human right to consider in the deployment of personalized recommender systems, next to the right to privacy and the prohibition of discrimination (CoE, 2020). The news consumers' fundamental right to receive information guaranteed by Article 10 ECHR[24] may prove an important point of departure to realize democratic values in the personalized media landscape (Eskens et al., 2017). Information consumption has changed dramatically and needs to be reconceived. Given the vast amount of information produced and consumed, to some extent users are necessarily passive actors which have to delegate information filtering to algorithms and, therefore, to platforms. Thus, the right to information is, in effect, a right to receive information. How this would eventually translate is difficult to argue. It is indeed an under-theorized right, lacking a framework to fully understand the rights of news consumers or the obligations of states regarding news recipients.

Article 10 entails positive and negative obligations for states (CoE, 2020); a negative obligation to abstain from interference with the freedom of expression rights of

---

[24] ARTICLE 10 of the EHCR, Freedom of expression:

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

2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary

journalists, editors and users, including the use of AI in, for example, forms of automated censorship. And positive obligations for member States to create the conditions for a favorable environment for the exercise of freedom of expression, also in the relationship between private parties. In effect, the exercise of this freedom carries with it 'duties and responsibilities' (Article 10 ECHR, para. 2). The scope of someone's duties and responsibilities depends on the situation and the technologies used for communication.

Media (and in particular news) personalization invites us to reconsider subjective rights to receive information. In traditional one-to-many media, people have a subjective right to receive information that others are willing to impart, but they don't have a right to receive information that the media is not willing to impart. In fact, the media would lose its editorial freedom if people could demand specific news stories and distribute these to them and, at the same time, if these were conflicting, it would be difficult to decide whose right to receive information should prevail. By enabling one-to-one communication, personalization technologies could, in theory, resolve conflicts between subjective rights to receive information and the media's or other parties' freedom of expression. Such a type of subjective right to receive information could help to establish what news consumers legitimately may expect from the news media with respect to the diversity or relevance of personalized recommendations.

The right to receive information is also intrinsically connected with the rights to privacy and data protection. Users who are aware of the use of algorithmic tools by news media, and know that they involve personal data collection, may fear the consequences of such personal data processing. The risk can be *chilling effects,* namely they may hesitate to consult online news and recommended content and affect the exercise of their expression and information rights (Büchi et al., 2019). This connection between freedom of expression and privacy rights has been called 'intellectual privacy' (Richards, 2015). The possibility of state access to data on individual reading patterns, collected by newsrooms and internet intermediaries to create or improve AI-driven tools, also poses a broader, latent threat to democracy.

Actually, media personalization may enable or hinder the exercise of this largely institutionally protected right. There are many different values and interests at stake especially with news personalization, which can lead to conflicts (prominently truth finding versus social cohesion) that are not likely to end up in court but must be discussed

in public. There is a need to further discuss what the right to receive information should mean nowadays, how it relates to data protection, and to empirically study how people's information seeking strategies and privacy attitudes influence the exercise of this right.

Considering the special position of media and the preference for self-regulation following from media freedom, it is unlikely that the normative principles in the media context should directly translate into legal obligations (Harambam et al., 2018). Negative and positive obligations of member States following from Article 10 ECHR, in fact, do not hold against these 'technology companies'. At the very least, from Article 10 ECHR can follow some obligation, for example that media actors have a responsibility to develop professional rules regarding the risks of personalization algorithms for bias and media diversity. Nevertheless, this raises the following question: how the state could dispose of its obligations? Eskens et al. (2017) unveil five perspectives from which to understand the right to receive information and answer this question:

1. *Political debate*. receiving information is essential for people to participate in political life. Eskens et al. broadly intend political participation as encompassing participation in the electoral process, but also 'discovering and forming an opinion of the ideas and attitudes of political leaders', forming opinion about the public and private activities of political representatives, and discussing governmental actions with others. Such perspective leads to a concrete obligation for states. In particular, as Eskens et al. (2017) stress, the ECtHR held that states have a duty to ensure that the public has access to accurate information and varied opinions through audiovisual media.

2. *Truth finding*. This perspective means that the quest for truth may legitimize a claim to receive information. It is a broader perspective than political participation, since people may aim to find out the truth about non-political issues. This also carries the implication that the public is entitled to receive a diversity of information, because truth will only emerge out of the marketplace of ideas. Diversity has for a long time been a central tenet of European media law and policy, implementing the value of pluralism. The ECtHR has indeed established that democracy demands pluralism, and that the state is the ultimate guarantor of

pluralism. Subsequently, the Court imposes on states the positive obligation to put in place a legislative framework to guarantee pluralism in the media system.

3. *Social cohesion*. According to Eskens et al. (2017), the Council of Europe, as well as the ECtHR, recognize the importance of receiving information for the creation and maintenance of social cohesion. In this sense, social cohesion refers to high quality and strong social relations, including people a sense of belonging to a social group, and group orientation towards the common good. This necessitates a certain degree of trust, while social cohesion also enables mutual trust amongst people. ECtHR case law supports the connection between social cohesion and exposure diversity. The Court recognizes that public debate about complex issues furthers social cohesion by ensuring that representatives of all views are heard, particularly minority voices.

4. *Avoidance of censorship.* Censorship clearly has implications for exercising the right to receive information as it prevents people from receiving information. In this case, the Court has established that the right to receive information prohibits the government from restricting a person from receiving information that others wish or may be willing to impart to him or her.

5. *Self-development*. Various legal judgements or policy instruments are based on the idea that receiving information is necessary for people's self-development. This perspective covers two types of information: personal information in the sense of personal data, and information that is not personal but relevant to someone's private life for another personal reason. The Court established that access to information is also important for private and cultural issues, in addition to public interest issues. Besides, this case also relates to the perspective of social cohesion.

Users share a 'cooperative responsibility' with social media platforms as content editor (Helberger et al., 2018). Concerning current individual control over information,

Harambam et al. (2018) identifies four ways in which people can influence the algorithmically curated information they encounter, and these are:

1) *Alternation*, that is, switching between different news outlets and media forms, and also by using multiple or different recommenders. Yet, it requires effort, skills and, also, it is limited in working around hidden biases in algorithmic curation.

2) *Awareness*, that is, being aware of algorithms functioning. In this respect, the GDPR, which raises the bar on transparency and user control over personal data processing, and may have a positive impact.[25]

3) *Adjustment*, that is, adjust algorithms according to personal interests and wishes. Most news outlets, however, have not developed formal ways to influence their curating algorithms.

4) *Obfuscation*, that is, mobilizing against the data-driven processes through the deliberate addition of ambiguous, confusing, or misleading information to interfere with data collection. Yet, this may run against some of the goals and benefits of personalization.

The above techniques, however, are not particularly effective and are difficult to pursue for the average user.[26] Yet, more generally, the ability to intervene in data-driven processes can be challenging and costly to implement, in particular at scale, considering the nature of the required possibility to intervene. Then, what forms of intervention at the level of data inputs and processing can be achieved in the context of algorithmic news recommenders to further the value of voice must be discussed further. This leads to a related issue which might help to better define strategies to tackle the current limitations of data protection law previously outlined.

---

[25] This is the case with Facebook which is implementing a feature "why I am seeing this" to provide users a better understanding of the reasons why a post has been recommended.

[26] Recently, it is even questioned whether the actual 'horizontal approach' based on the notion of 'average consumers' is fit to protect all consumers in a highly personalized digital environment.

4.1.6 Manipulation

The European Commission (EC) has recognized the exposure of citizens to large scale online disinformation (2018a) and micro-targeting of voters based on the unlawful processing of personal data (2018b) as major challenges for European democracies.

Bradshaw and Howard (2018) have found that governments and political parties in an increasing number of countries of different political regimes are investing significant resources in using social media to manipulate public opinion. Political bots, as they note, are used to promote or attack particular politicians, promote certain topics, fake follower bases, or to get opponents' accounts and content removed by reporting it on a large scale. Micro-targeting, as another tactic, is commonly defined as a political advertising strategy that makes use of data analytics to build individual or small group voter models and to address them with tailored political messages (Bodó et al., 2017). These messages can be drafted with the intention to deceive certain groups and to influence their behaviour, which is particularly problematic in the period that leads up to voting in elections when decisions of paramount importance for democracy are made. During elections the tensions are high and the time for correction or reaction is scarce.

In a response to these challenges, and to ensure citizens' access to a variety of credible information and sources, the EC has put in place several measures which aim to create an overarching "European approach" (Nenadić, 2019). The key instruments contained in the approach and drew out the key principles upon which it builds: data protection; transparency; cooperation; mobilizing the private sector; promoting diversity and credibility of information; raising awareness; empowering the research community. The principles of traditional media regulation in the electoral period are impartiality and equality of opportunity for contenders.

The European approach certainly requires further elaboration, primarily to include additional layers of transparency. This entails transparency of political parties and of other actors on their actions during election campaigns, as well as further transparency about internal processes and decision-making by platforms, especially on actions relevant to pluralism, elections and democracy. This approach faces two constraints. Namely that the power of global platforms is shaped in the US tradition, which to a significant extent differs from the European approach in balancing freedom of expression and data

protection. Conversely, the EU approach confronts the resilience of national political traditions in member states, in particular if the measures are based on recommendations and other soft instruments.

In the online environment, the principles applied to traditional media require further elaboration as the problem of electoral disinformation cuts across a number of different policy areas, involving a range of public and private actors. Political disinformation is not a problem that can easily be compartmentalized into existing legal and policy categories. It is a complex and multi-layered issue that requires a more comprehensive and collaborative approach when designing potential solutions. Lastly, the ability to manipulate is becoming as effective as widespread. Thus, new forms of regulation and control are discussed.

4.1.6.1 Towards a "Right Not to Be Deceived"?

Human behavior can be manipulated by priming and conditioning, using rewards and punishments. Algorithms can autonomously explore manipulative strategies that can be detrimental to users (Albanie et al., 2017; Lanzing, 2018). Basically, they exploit human biases and vulnerabilities to affect self-control, self-esteem and personal beliefs. As a consequence, autonomy is threatened (Kohl et al., 2019; Zarsky, 2019; Deibert, 2019). This is even more concerning in the case of negative effects on children's wellbeing, including increased risk of suicide and depression, conflict with parents and adverse effects on cerebral and social development (Kidron et al, 2018). Recital 38 of the GDPR in fact states that: "Children merit specific protection with regard to their personal data, as they may be less aware of the risks, consequences and safeguards concerned and their rights in relation to the processing of personal data…".

Social media already act as addictive machines (Deibert, 2013). As such, users are tempted to give up their rights to benefit from such 'hyper-personalization'. In theory, using such techniques goes against the 'fairness' and 'transparency' provisions of the GDPR (Zarsky, 2019). In this sense, GDPR proves to be a necessary yet insufficient step. In fact, as smart environments will permeate societies, users (especially young people) will be automatically plugged in and guided through life along algorithmically determined pathways, and the boundary between legitimate persuasion and deception will become increasingly blurred. This is why more reflection is needed.

The right most clearly implicated in big data-driven hyper-nudging is the right to informational privacy. As such, data controllers are obliged to follow the principle of data protection by design and by default. This might go beyond the individual to focus a priori on the creation of better algorithms (Edwards and Veale, 2017). For example, privacy might be fundamental to enable what Hildebrandt (Hildebrandt, 2019) defines as 'agonistic machine learning', that is, demanding companies or governments that base decisions on machine learning to 'explore and enable alternative ways of datafying and modelling the same event, person or action'. In the case of personalization, the value of serendipity is thus implied.

Of course, also consumer law could actually help to protect consumers against unfair profiling and persuasion practices (Helberger et al., 2017). However, the extensive uncertainty and context dependence implies that people cannot be counted on to navigate the complex trade-offs involving terms of services and privacy self-management (Acquisti et al., 2015). There is overwhelming evidence that most people neither read nor understand online privacy policies. According to behavioral sciences as well, existing notice and consent model cannot be relied upon to protect the right to informational privacy (Yeung, 2017).[27]

In addition to privacy, online digital users should have a separate and distinct right not to be deceived, rooted in a moral agent's basic right to be treated with dignity and respect given that deception violates the autonomy of the person deceived, involving the control of another without that person's consent. Basically, this might unfold throughout several practices:

1. First of all, an ideal aim to achieve is 'content neutrality' by design, meaning platforms is agnostic towards content. Content neutrality might be also achieved with a 'right to profile transparency', through an infrastructure that supports what the philosopher Mireille Hildebrandt defines as 'counter-profiling' (2015), meaning "to conduct data mining operations on the behaviours of those that are in the business of profiling whether 'those' are humans, computing systems or hybrid configurations" (p.223). This is a general strategy to figure out how one is actually targeted.

---

[27] Also the European Commission recognizes that individual decision-making occurs subconsciously, passively and unreflectively rather than through active deliberation (2016).

2. Platforms should routinely disclose to their users and the public any experiment that the users were subjected to.[28] Appropriate information and specific consent to the use of techniques of deception ought to be given. Yet, given the complexity and subtleness of online deception the choice may not even be sufficiently informed and conscious even with consent.

3. There could also be established independent boards to review and approve experiments in advance. A specific agency responsible to detect manipulative techniques, prevent them and protect fundamental human rights may be required. The current vast power asymmetry between global digital service providers and individual service users cannot be ignored. In the Big Data era any kind of personalization is a form of nudging with unintended (and unsettling) consequences. As it currently stands, the EU legal framework seems to be insufficient in facing this challenge. In fact, this would be particularly in line with two intertwined newly discussed human rights: the 'right to not be measured, analysed or coached'[29] (van Est et al., 2017) and a 'right to cognitive sovereignty' that ought to protect individuals' right to mental self-determination (Yeung, 2018b).

4. As for political microtargeting, platforms should clearly identify political advertisements and who paid for them and offer a transparent repository of all the material used in the campaign provided by candidates and political parties. Furthermore, a stronger requirement for political parties to report on the amounts spent on different types of communication channels is desirable, as well as the requirement for platforms to provide more comprehensive and workable data on sponsors and spending in political advertising.

4.2 Personalization Ethics

Personalization ethics is by no mean an abstract exercise. It can (and should) guide the development of personalization technologies, practices and usage. Numerous data ethics public policy initiatives, for example, have been recently created, moving beyond issues

---

[28] In U.S. a bill about this issue (called "Deceptive Experiments To Online User Reduction") has been recently proposed and discussed.

[29] The right is intended as "the right to decide whether or not they want to participate in experiments carried out by other actors (which usually goes hand in hand with surveillance) or other activities that involve registering or otherwise observing people's lives and influencing their behaviour with technological means." (van Est et al., 2017, p.43).

of mere compliance with data protection law to increasingly focus on the ethics of big data, especially concerning private companies' and public institutions' handling of personal data in digital forms (Hasselbalch, 2019). These should not be viewed as solutions in their own right. Indeed, they do not replace legal frameworks. Rather, they "complement existing law and may inspire, guide and even set in motion political, economic and educational processes that could foster an ethical "design" of the big data age, covering everything from the introduction of new laws, the implementation of policies and practices in organizations and companies and the development of new engineering standards, to awareness campaigns among citizens and educational initiatives (p.2, Ibid)." In other words, data ethics (but also, more broadly, digital and information ethics) become means to make sense of emerging problems and challenges of a rapidly evolving information and data infrastructure, and to evaluate various policies and solutions.

To better understand such processes, there will be now introduced several fundamental concepts and notions that will be used throughout the Section. By doing this, a critical overview of the ethical debate is presented, and the relation between personalization regulation and personalization ethics is understood.

4.2.1 Soft and Hard Ethics

To begin with, it is important to acknowledge a meaningful distinction (already introduced in Section 4); the one between hard and soft ethics (Floridi, 2018):

- *Hard ethics* occurs when we discuss values, rights, duties and responsibilities— or, more broadly, what is morally right or wrong and what ought or ought not to be done—in the course of making choices or taking decisions in general and of formulating new legal norms or challenging existing ones in particular. Therefore, hard ethics is what may contribute to making or shaping the law.

- *Soft ethics* covers the same normative ground as hard ethics, but it does so by considering what ought and ought not to be done over and above the existing norms, not against them, or despite their scope, or to change them, or to by-pass them, e.g. in terms of self-regulation. In other words, soft ethics is post-

compliance ethics. This last approach has a dual advantage (Floridi, 2018); on the one hand, an 'opportunity strategy', thus it enables actors to take advantage of the social value of digital technologies, seizing new opportunities that are socially acceptable or preferable and balancing any precautionary principle with the duty not to omit what could and ought to be done. On the other hand, the advantage of a 'risk management solution', thus it enables organizations to anticipate and avoid costly mistakes.

It is really within the EU that post-compliance soft ethics can currently be exercised, to help individuals, companies, governments and other organizations to take more and better advantage, morally speaking, of the opportunities offered by digital innovation. Because even in the EU, legislation is necessary but insufficient. It does not cover everything (nor should it), and agents should leverage digital ethics in order to assess and decide what role they wish to play in the infosphere, when regulations provide no simple or straightforward answer, when competing values and interests need to be balanced (or indeed when regulations provide no guidance) and when there is more that can be done over and above what the law strictly requires.

To better understand these two different approaches, let me make a paradigmatic example: recitals in the GDPR. As it is often the case with complex legislation, the Articles leave grey areas of normative uncertainty uncovered, are subject to interpretations and may require updating when applied to new circumstances, especially in a technological context where innovation develops so quickly and radically. Therefore, to help understand their meaning, scope and applicability, the Articles are accompanied by 173 Recitals. In EU law, Recitals are texts that explain the reasons for the provisions of an act, but are not legally binding. Yet, they are used by the Court of Justice of the European Union (CJEU) in order to interpret a directive or a regulation and reach a decision in the context of a particular case. In addition, they can also be used by the European Data Protection Board (the EDPB, which replaces the Article 29 Working Party), when ensuring that the GDPR is applied consistently across Europe. The Recitals themselves will require an interpretation, namely the ethical framework that can contribute to interpret the Recitals. Thus, while hard ethics is what contributed to the process leading to the elaboration of the GDPR, soft ethics is the framework that

contributes to enable the best interpretations of the Recitals. The space of soft ethics is bounded by human rights and compliance. Yet, its potential is an ever expanding one over time.

4.2.2 AI Ethical Principles

To fill the gap between the prescriptions derived from general theories and the results of the prescriptions in the world of policy making and professional practice it is fundamental to acknowledge that technology is not neutral with respect to values (Winner 1980). Artefacts can influence and alter our moral behavior through a reciprocal process, especially personalization that is an egopoietic technologies or technology of construction of the self (Floridi 2011). This is particularly true in the case of recommender systems that are indeed AI-driven.

As of now, EU can become a global leader in developing and using AI for good and promoting a human-centric approach and ethics-by-design principles (Commission, 2018). The increasing demand for reflection and clear policies on the impact of AI on society has produced several initiatives that state principles, values, or tenets to guide the development and adoption of AI. The risk is unnecessary repetition and overlap, if the various sets of principles are similar, or confusion and ambiguity, if they differ. In either eventuality, the development of ethical AI may be delayed.

Floridi and Cowls (2019) analyzed six sets of ethical principles from initiatives with global scope, or from within western liberal democracies, to summarize the effort and provide an overview of the debate. To begin with, by comparing the sets of principles there is a convergence (at least in Europe) with the four core principles commonly used in bioethics: beneficence, non-maleficence, autonomy, and justice. Yet, they are not exhaustive. In fact, a new principle is needed in addition: explicability, understood as incorporating both intelligibility and accountability.

Let me now briefly introduce these five principles and contextualize them into the issue of personalization:

1. *Beneficence*. It means promoting well-being, preserving "human dignity" and "sustainability." Its principle of "sustainability" articulates perhaps the widest of all interpretations of beneficence, arguing that "AI technology must be in line

with … ensur[ing] the basic preconditions for life on our planet, continued prospering for mankind and the preservation of a good environment for future generations."

2. *Non-maleficence*. It means privacy, security and 'capability caution'. Though 'do only good' (beneficence) and 'do no harm' (non-maleficence) may seem logically equivalent, they are not, and represent distinct principles. Non-maleficence assumes the responsibility by proactively working against the risks arising from technological innovations.

3. *Autonomy*. It is the power to decide (to decide). When we adopt AI and its smart agency, we willingly cede some of our decision-making power to technological artefacts. Thus, affirming the principle of autonomy in the context of AI means striking a balance between our decision-making power and the delegation to artificial agents. This introduces a notion we might call 'meta-autonomy,' or a 'decide-to-delegate' model: humans should retain the power to decide which decisions to take: exercising the freedom to choose where necessary, and ceding it in cases where overriding reasons, such as efficacy, may outweigh the loss of control over decision-making.

4. *Justice*. It means promoting prosperity, preserving solidarity, avoiding unfairness. The consequences of pre-existing disparity in autonomy are addressed in the principle of justice. Elsewhere 'justice' has still other meanings (especially in the sense of fairness), namely to eliminate unfair discrimination, promoting diversity, and preventing the rise of new threats to justice.

5. *Explicability*. It means enabling the other principles through intelligibility and accountability. The situation is indeed inherently unequal: a small fraction of humanity is currently engaged in the development of a set of technologies that are already transforming the everyday lives of almost everyone else. Different terms express this principle: "transparency", "accountability", "intelligibility", "understandable and interpretable".

Significantly, explicability complements the other four principles. As Floridi and Cowls (2019) put it "for AI to be beneficent and non-maleficent, we must be able to understand the good or harm it is actually doing to society, and in which ways; for AI to promote and not constrain human autonomy, our 'decision about who should decide' must be informed

by knowledge of how AI would act instead of us; and for AI to be just, we must know whom to hold accountable in the event of a serious, negative outcome, which would require in turn adequate understanding of why this outcome arose."
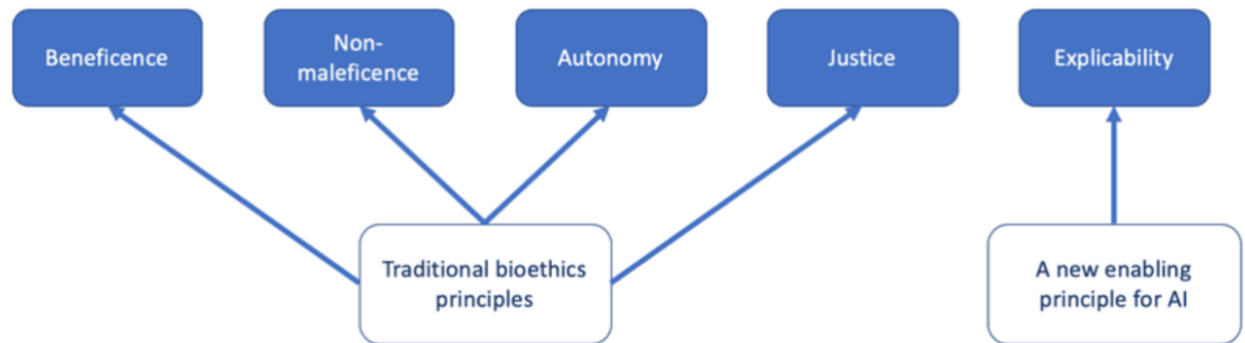
Figure 20. An ethical framework of the five overarching principles for AI (from Floridi and Cowls, 2019).

Concerning RSs, in the past 20 years they have been developed focusing mostly on business applications, and the emphasis has tended to be on commercial objectives (Milano et al., 2019). Thus, research into the ethical issues posed by RSs is still in its infancy. However, as we have seen in Chapter 2, RSs have several moral implications, as well as individual and social undesirable consequences. They need to be ethical (Paraschakis, 2017; Tang and Winoto, 2015). To make RSs ethical and operational one needs foremost to specify three parameters (Milano et al., 2019): a) what the space of options is; b) what counts as a good recommendation; and, importantly c) how the RS's performance can be evaluated. In news RSs, a good recommendation may be defined as a news item that is relevant to the user (Floridi, 2008). This is clearly not sufficient as the consequences of RSs go often beyond mere news relevance. Generally, the main concerns for ethical RSs are privacy, autonomy, inappropriate content, opacity, fairness and social effects (Milano et al., 2019) and these are all embraced by the five AI principles explained above. These need, however, to be complemented with ethical personalization in its broadest (and concrete) applications (as we will do in Chapter 5).

4.2.3 From Theory to Practice

In any case, enabling the dual advantage of 'ethical personalization'— so that the opportunities are seized, while the harms are foreseen and minimized or prevented —

requires asking difficult questions about design, development, deployment, practices, uses and users, as well as the data that fuel the whole process. The hope is that the principles outlined above can act as normative constraints. Here it lies the distinction between the 'what' (ethical principles) and the 'how' (technical requirements). Truly, in the last decade applied ethics has developed into an even more practical discipline. Such "design turn in applied ethics" (Van den Hoven et al., 2017) acknowledges that the needs and values of users are considered in their own right and not simply as a side constraint on successful implementation. This perspective helps to develop pragmatic methodologies and frameworks (like Value Sensitive Design and pro-ethical design) that help to make moral and ethical values integral parts of research and development (Friedman et al., 2002). However, the mere existence of these principles does little to bring about actual change in the design of algorithmic systems, leading to accusations of 'ethics washing' and feelings of 'ethics fatigue' (Floridi, 2019b).

In this sense, Bozdag and Timmermans (2011) have initially highlighted how to unfold the principles stressed to make personalization ethical:

1) To make sure different identities are allowed per user, which might differ per context.

2) To design for autonomy, so that the user can customize the filter, and change the identity that is formed on basis of his previous interactions, and

3) to design for transparency, so that the user is aware that a filter is taking place. The user must be able to see which criteria is used for filtering, and which identity the system has of the user.

Similarly, Hakala et al. (2014) wrote a foresighted article titled "*Ethical Personalization Act 2025*" in which they stressed the importance of education and, therefore, of the awareness of the educators over personalization practices. These solution, however, are still too broad. As such, this debate will be widened in the last final chapter in order to provide a more detailed analysis of personalization ethics.

Still, the gap between principles and practice is as dangerous as too large (Morley et al., 2019). Especially in a digital context, ethical principles are not simply either applied or not, but they are regularly re-applied or applied differently, or better, or ignored as algorithmic systems are developed, deployed, configured, tested, revised and re-tuned. Complexity, variability, subjectivity, and lack of standardization, including variable

interpretation of the 'components' of each of the ethical principles make this endeavor challenging. This is also risky as the lack of guidance may (a) result in the costs of ethical mistakes outweighing the benefits of ethical successes; (b) undermine public acceptance of algorithmic systems; (c) reduce adoption of algorithmic systems; and (d) ultimately create a scenario in which society incurs significant opportunity costs (Cookson, 2018).

Another significant risk is that the tech industry, especially in Silicon Valley, 'owns ethics', in the sense that it does not bear sole responsibility for it, but rather oversees integration of ethical projects across organizations. For example, using corporate resources hiring staff for 'ethical' roles. This recent widespread commitment to ethics "in tension with—and at risk of being absorbed within— broader and longer-standing industry commitments to meritocracy, technological solutionism, and market fundamentalism (Metcalf and Moss, 2019, p.11)." In this context, ethical words like "fairness," "accountability" and "transparency" can be used as "weasel words" whereas, more pragmatically, words like "justice, liberation, and rights" actually remain frameworks for material improvement of human lives.

Furthermore, during the application of ethical principles there are still significant concerns or malpractices that might limit the potential to implement good ethics. Floridi (2019) identifies five major ones that deserve to be introduced:

(1) *ethics shopping* as "the malpractice of choosing, adapting, or revising (mixing and matching) ethical principles, guidelines, codes, frameworks, or other similar standards (especially but not only in the ethics of AI), from a variety of available offers, in order to retrofit some pre-existing behaviours (choices, processes, strategies, etc.), and hence justify them a posteriori, instead of implementing or improving new behaviours by benchmarking them against public, ethical standards (p.2)".

(2) *ethics bluewashing* as "the malpractice of making unsubstantiated or misleading claims about, or implementing superficial measures in favour of, the ethical values and benefits of digital processes, products, services, or other solutions in order to appear more digitally ethical than one is (p.3)".

(3) *ethics lobbying* as "the malpractice of exploiting digital ethics to delay, revise, replace, or avoid good and necessary legislation (or its enforcement) about the design, development, and deployment of digital processes, products, services, or other solutions (p.4)."

(4) *ethics dumping* as "the malpractice of (a) exporting research activities about digital processes, products, services, or other solutions, in other contexts or places (e.g. by European organizations outside the EU) in ways that would be ethically unacceptable in the context or place of origin and (b) importing the outcomes of such unethical research activities (p.6)".

(5) *ethics shirking* as "the malpractice of doing increasingly less ethical work (such as fulfilling duties, respecting rights, and honouring commitments) in a given context the lower the return of such ethical work in that context is mistakenly perceived to be (p.7)."

The above one are the most obvious and significant ethical risks one might encounter when navigating from principles to practices in digital ethics, thus they need to be kept in mind. These will indeed be highlighted when necessary during the following analysis.

There are, then, also concerning research gaps. Currently, most attention for all the ethical principles is focused on interventions at the early input stages but very few tools or methods during the middle building and testing phases (Morley et al., 2019). Few of the available tools surveyed provide meaningful ways to assess, and respond to, the impact that the data-processing has on an individual, and even less on the impact on society as a whole. Also, while ethics and values have been studied extensively, these conversations are frequently bound to the academic community and related discourses, making practitioner access to these conversations difficult. Comprehensive ethics education is critical to ensure that future generations of practitioners, designers and engineers take their role as creators of futures seriously.

Furthermore, legislations may even prove to have unintended consequences in terms of incentivizing some types of research over others and promoting minimum adherence over best practice, given that being legally compliant may be seen to be good enough. For

example, GDPR has encouraged a focus on privacy and explicability over the promotion of autonomy in design choices and done very little to encourage competition to be the most ethical system (Floridi, 2018). As such, societies need to come together in communities comprised of multi-disciplinary researchers, innovators, policymakers, citizens, developers and designers to foster the development of common knowledge and understanding and a common goal to be achieved with innovative tools and methodologies for applied AI ethics.

Finally, there are opportunities and (ethical) challenges of ethics by design. It is possible indeed to nudge users to rely on empowerment nudges, which promote decision-making in the interests of citizens, as judged by themselves, without introducing further regulation or incentives' or manipulative measures (Helberger et al., 2016). A similar approach is embraced by Floridi (2016), which suggests what he calls "tolerant paternalism" that, in short, aims to modify the level of abstraction of the choice architecture by educating users to make their own critical choices and to assume explicit responsibilities. Certainly, there are also ethical concerns about ethics by design and nudging (Helberger et al., 2016). This issue will be analyzed in Section 4.2.6. Then, two other related issues are now introduced, namely the significance of intercultural information ethics (Section 4.2.4) and the role of group privacy (Section 4.2.5).

4.2.4 The Relevance of Inter-cultural Ethics

Where does morality begin and end in a socio-technical infrastructure that extends across jurisdictions and continents, cultural value systems and societal sectors? Evidence from different parts of the world shows that the relationship between users and RSs, societies and privacy, and journalists and automation can differ greatly, and that the socio-cultural context as well as contextual factors such as the media and political system should be taken into account in the development of personalization, its governance and its regulation. This is particularly true regarding a fundamental and paradigmatic value of information societies: privacy.

Importantly, intercultural information ethics study shows how the concept of privacy is associated with the Western perspective of individual autonomy (Capurro 2005), even though forms of privacy are found in every culture throughout the world. More generally, many experts highlight the dangers of the dominance of "Western" ethics, even in other

strictly related debates such as the one on Artificial Intelligence (AI). For example, not even one African country has drafted an AI ethical guidelines independently from international or supranational organizations (Jobin et al., 2019). This reveals a power imbalance in the international discourse. As pointed out by a number of scholars (Capurro, 2005, 2008; Ess, 2006; Olinger et al., 2005; Makulilo, 2016), privacy discussions have been dominated for a long time by Western cultures, to the exclusion of non-Western ethical and socio-cultural perspectives. Some have even erroneously argued that the concept of privacy is completely absent in many Eastern cultures (Ess, 2006). This limited approach carries risks not only because the ethical norms and values can collide with those of the communities in which they are deployed, but also because it might represent a missed opportunity to enrich and further develop the current paradigm of privacy protection so as to effectively face new emerging technological challenges.

Yet, in intercultural information ethics there is a danger not only to group together all Western traditions under a single banner but also to negatively designate such influence. Only because there exists a monopoly of influence by one system over another does not mean that said monopoly is not valuable. Instead, it can be stressed that those same liberal democratic values arose out of conflicts of cultural and subcultural difference and are specifically designed to be accommodating enough to include a rather wide range of differences. Therefore, as Ess (2006, 211) pointed out, a global information ethics that seeks to avoid any homogenization must "conjoin shared norms while simultaneously preserving the irreducible differences between cultures and peoples."

An initial response to the encounter with apparently irreducible difference is conceived as a *modus vivendi* pluralism, one that 'simply lets these differences stand and accepts that there is no further common ground to be found between diverse views, approaches, norms, etc' (Ess, 2006, 216). Such an approach seems insufficient. As Luciano Floridi pointed out (in Ess 2006), such modus vivendi pluralism is pragmatically not to be found, in the sense that in the absence of some larger form of shared community life and values, and as the basic tensions are not resolved, differences seem rather to sustain constant conflicts that might bring to violent confrontation.

Charles Ess (2006) argues that a global information ethics could overcome such conditions by taking up an *ethical pluralism*, specifically Aristotles *pros hen*.[30] Away from ethical relativism and ethical dogmatism, this model highlights connection in the face of irreducible difference, a complementarity in which an active engagement across these differences works to enrich both participants. This is a promising endeavor; according to Ess (2006, 218), conceptions of privacy and data privacy protection laws emerging in countries such as China, Hong Kong, Thailand, and Japan "cohere and resonate with Western notions of privacy and data privacy protections – and in ways that compose a robust, pros hen pluralism." Yet, intercultural dialogue is difficult, so that excessive optimism is not reasonable (Hiruta 2006). Discussions of privacy in non-Western cultures are indeed scattered and sometimes conflicting. Whereas some even contend that the concept of privacy is completely absent in many Eastern cultures (Ess 2006).

Following Ma (2019), considering other ethical perspectives jointly could spark research contributing to the relational understanding of privacy in three ways: i) to understand how cultural factors influence people's understanding of privacy, ii) to assess whether there are some underlying shared social norms that might be more salient in one context/culture than another and iii) to find empirical evidence of how prevalent control-based belief of privacy is among individual, corporate and/or government documents in other cultures other than Western ones.

To conclude, current working conceptualizations of privacy – as well as other related values – are rooted in the Western concept of individual autonomy as having control over personal information. This is increasingly unsuited in helping answer many questions people could face in everyday life, even in the ethics of personalization. For example, algorithmic systems which measure, count and profile groups of individuals create knowledge that is not (only) private to an individual, but which reveals something about a group of individuals. Being targeted as part of a group could indeed affect one's online personalized experience. This is, in a nutshell, the problem of protecting 'group privacy' (Taylor et al., 2016). As such, researchers have moved on and proposed a contextual understanding of privacy (Nissenbaum, 2011) as well as relational privacy (Stoljar, 2015;

---

[30] 'Pros hen pluralism' is a three-layered doctrine, consisting of one requirement for diversity at a cultural level, another for diversity at an ethical theoretic level, and the other for unity at the ethical theoretic level (Ess, 2006).

Bannerman, 2018; Ma, 2019). Cultures that embrace a conception of privacy as (more) relational have already been analyzed in this context, notably Indian (Basu, 2012) and Japanese cultures (Capurro, 2005), Buddhism (Hongladarom, 2016), Confucianism (Ma, 2019) and, finally, also the sub-Saharian etno-philosophy of Ubuntu (Reviglio and Alunge, 2020).

4.2.5 The Importance of Group Privacy

Another particularly significant theoretical and practical issue to protect privacy and guarantee fair personalization practices is represented by 'group privacy'. As a matter of fact, nowadays most people are not profiled as individuals but, instead, as a member of a specific group. Data are not gathered about a specific person or group, rather, to borrow van der Sloot's terms (2016), they are gathered about an 'undefined number of people during an undefined period of time without a pre-established reason'. These data are then processed on a group or aggregated level through the use of statistical correlations to provide personalized services or make targeted decisions. Thus, the 'individual' element is mostly lost, yielding biased or discriminatory results. Despite the focus on personal identifying information is still useful for more traditional data processing activities, a number of scholars (Taylor et al., 2016) indeed argue that in the big data era these should be supplemented by a focus on identifying information about categories or groups.

Yet, harm to specific individuals is very difficult to demonstrate. These data do not directly identify a person, and consequently fall outside the scope of the data protection regulations, although they may affect him or her as being part of a specific group.[31] To argue what a notion of group privacy might imply both theoretically and practically, it is necessary to firstly understand what is intended as 'group', and how profiling threatens such 'group privacy'. And being targeted as part of a group could affect algorithmic decisions made about us and for us, for example credit scoring or online personalization.

Kammourieh et al. (2016) note that a 'group' refers to a 'class or unit made up of more than one person or thing' and the term also leads us to imagine 'people who have formed

---

[31] Interestingly, Van der Sloot (2016) notes that "one could focus on the initial moment when personal data are gathered and not yet aggregated, but this may only concern the split second which it takes to aggregate data. The same counts for the moment at which group profiles are applied and used to affect a specific person. This only concerns the very end of the data process. By focusing on the individual, his interests and rights, one loses sight of the larger part of the data processing scheme and the general issues concerned with that."

explicit ties, whether legal or otherwise, that bind them together' like a social, religious or ethnic group, or a structured organization like a company or political party. Traditionally, groups can be active (members proclaim the existence and willingly identify themselves as members of the group) or passive (members are attributed membership by society and may not willingly wish to be identified as such e.g. categories of vulnerable persons in a country or society). This conventional definition changes with new technological practices. Through machine learning, it is now possible to infer information and draw knowledge from vast amounts of unstructured data - usually without anyone's knowledge - while pattern recognition tools facilitate the discovery of previously imperceptible interrelations within datasets and, as such, creates new means for identifying and grouping individuals (Kammourieh et al., 2016).

These new abilities to group individuals poses significant ethical and legal problems. Taylor (2016) raises the problems of misuse and misinterpretation or inaccuracy of such information collected in a Big Data context: inability of privacy legislations and ethical frameworks to keep up with the pace of, and therefore regulate the use of such data (misuse), and the eventuality of such data being interpreted by data scientists rather than social scientists (misinterpretation). There is equally the fundamental problem of consent: how can informed consent be collected from each member of such groupings (Taylor 2016; Bernal 2012).

These developments triggered discussions on whether a legal notion of collective privacy is necessary and whether this collective dimension requires the granting of specific collective rights. In privacy literature, the idea of group privacy is not absent (e.g. Bloustein, 1978). The philosopher of information Luciano Floridi (2014, 1) defines it as a "right that is held by a group as a group rather than by its members severally. It is the group, not its members, that is correctly identified as the right-holder." Since long, the *group* has not been conceptualized in terms of privacy beyond a collection of individuals with individual interests in privacy (Bloustein 1978). In other words, the rights holder has always been the data subject and the rights regarding informational privacy have been mainly exercised by single individuals. Even in the new often-lauded GDPR, specific reference to group privacy is absent[32].

---

[32] European data protection law however appears to provide legal grouns for the consideration of networked and group privacy with the EU Data Protection Directive Article 29 Working Party's Opinion 4/2007, which considers data to be personal data if it impacts the rights of an individual.

Historically, the legal position of the 'group' has been a complex one and disputed one. Groups are usually dynamic entities as they come in an endless number of sizes, compositions, nature, and they are fluid (Taylor et al., 2016). So, one may argue, with groups acting as moving targets and no clear or fixed ontology for them, it may be difficult to develop a theory of group privacy. The only fixed entity is therefore the individual, so group privacy is nothing more than the sum of 'individual privacies'. Groups, however, are not "given" but are determined by algorithms through clustering. In other words, the activity of grouping comes before its outcome, the group. This different approach helps to explain why profiling a standard kind of grouping may already infringe on the privacy of the resulting group, if such profiling is oriented by a goal that in itself is not meant to respect the privacy of the group. This is true even in cases in which the members of the group are not aware of this: a group that has been silently profiled, in fact, need not be aware of this to have its privacy protected.

In big data contexts, the common attributes of the group often become evident only to the data gatherer. A telling example is what Nathaniel Raymond (2016, 48) calls "demographically identifiable information," or DII, which he defines as: "either individual and/or aggregated data points that allow inferences to be drawn, enabling the classification, identification, and/or tracking of both named and/or unnamed individuals, groups of individuals, and/or multiple groups of individuals according to ethnicity, economic class, religion, gender, age, health condition, location, occupation, and/or other demographically defining factors." From a traditional perspective, such information, provided it relates to no specific individual but to a given population, is not subject to privacy law. However, there are reports of such information being (hacked and) used to study, locate and launch attacks on targeted populations in warzones, or combined with other data points to (sometimes erroneously) infer the Muslim faith of taxi drivers in New York city (Kammourieh et al., 2016). In a nutshell, individuals could still get affected by the processing of such information.

The implication of the technical issues related to group privacy is that our legal, philosophical and analytic attention to the individual may need to be adjusted, and possibly extended (Taylor et al., 2016). The fact that the individual is no longer central, but incidental to these types of processes, challenges the very foundations of most currently existing legal, ethical and social practices and theories related to privacy.

4.2.6 To Nudge or Not to Nudge?

One of the hottest ideas in current policy debates is "libertarian paternalism," the design of policies that push individuals toward better choices without limiting their liberty (Sunstein and Thaler 2008; Hausman and Welch, 2010; Floridi, 2016). This debate is not only occurring in policy debates but it is also reflected in personalization systems. Thus, to better understand the ethics of personalization in social media, we now discuss the ethics of nudging and, given that nudging is a form of persuasion, we begin with an introduction to the ethics of persuasion to understand both the difference between persuasion and manipulation, and between persuasion and nudging.

Persuasion is a very broad and diffuse phenomenon and, therefore, the concept is notoriously difficult to characterize. To some extent, it is an umbrella term for 'influence'. It can be defined as "the action or process of persuading someone or of being persuaded to do or believe something (Oxford Dictionary)." From the serpent in the Garden of Eden to modern mass-media society, persuasion is a natural part of human interaction. There have always been human persuaders in society capable of changing people's minds and behaviors. Aristotle considered persuasion an art of rhetoric. Examples of persuaders abound—cult and political leaders, mothers, car salesmen, journalists. Also teachers can be considered as institutionalized persuaders. Nowadays persuasive efforts evidently manifests in media, especially through propaganda, advertisement, and opinion leaders. Since long, media technology has played a significant role in facilitating the delivery of persuasive messages to buy, donate, vote or act – from megaphones to podcasts. Media have always been a vector for persuasion and, at times, manipulation. Critical Theory – in particular the Frankfurt school – since long denounced the role of media in shaping contemporary societies (FONTE). This also is caused by the intertwined relationship between media and advertisement. Marketing has in fact always attempted to understand individuals' desires, emotions and thoughts. Eventually, it is important to recognize that persuasion is a pharmakological phenomenon, that is, it can be used either for the good (e.g. persuasive learning or health behavior change) or for malicious ends.

The standard definition identifies two possible objects of change, that is, attitudes (such as opinions and beliefs) and behaviors (such as decisions). In this sense, change can be intended in three main ways: formation, alteration, and reinforcing (Oinas-Kukkonen,

2010) and it has traditionally been associated with two routes (Pettty and Cacioppo, 1986): 1. central route, whereby an individual evaluates information presented to them based on the pros and cons of it and how well it supports their values and 2. peripheral route, in which change is mediated by how attractive the source of communication is and by bypassing the deliberation process. Of course, when the latest rules it is ethically disputable. Persuaders have indeed always stood on uneasy ethical ground. To provide the most paradigmatic example: if a serpent persuades you to eat a fruit, to what extent the culpability fall upon you or upon the serpent? Ethicists have struggled with such questions for thousands of years.

Persuasion apparently distributes responsibility between the persuader and the persuaded. It is therefore fundamental to distinguish (whether and when possible) whenever persuasion falls into manipulation. In order to bypass human resistance in being deceived, the manipulator has to hide the fact that he or she is manipulating the person in some way. In other words, there must be misinformation, disinformation and/or information asymmetry. Stated more generally, manipulation can proceed in three distinct ways, all of which violate the principle of persuasion: 1) providing false or too incomplete information (deception) 2) providing no comprehensible information at all, but only some ambiguous cues that likely lead to the target change of the manipulator 3) providing information, but manipulating the context in such way that the central route is blocked and the peripheral route will likely lead to the target change of the manipulator. The principle underlying manipulating and deception is thus controlling individuals in such way that their awareness of the very fact that they are manipulated is prevented. Manipulation and deception are ruled out by the principle of persuasion.

So, what is the difference between persuasion and nudging? This distinction is essential as it helps to recognize how nudging and persuasion are different, with the former merely targeting a particular behaviour in a particular setting, and the latter targeting a person's underlying beliefs and preferences, with a view to secure more lasting identity, attitude and behaviour change. To appreciate the difference between nudging and persuasion, imagine a shop in which customers are nudged onto a particular path to encourage them to buy particular products. This strategy may work as long as the customer does not realize they are being nudged. However, the moment that choice architecture is no longer present, old habits will reappear instantly.

In sum, persuasive technologies (PTs) can be used in policies which apply nudging techniques. These belongs to the more general class of 'behavior-influencing technologies'. Proponents of this class of technology argue that in a society that is increasingly shaped by technology, we should delegate part of our moral decision-making to this technology. Some scholars like Achterhuis argued for this stance in developing his idea of 'moralization of technology' (Veerbek, 2011). Others like Skinner (1972) went even further arguing that we should exploit these PTs for social control and engineering. Since John Stuart Mill's *On Liberty* (Hausman and Welch, 2010), many have vigorously attacked such paternalistic proposals by arguing that it would amount to an unwarranted interference with our autonomy and plea to technocracy instead of democracy (Brey, 2006).

In nudging, however, designers are supposed to leave the "choice set" while taking advantage of imperfections in human decision-making abilities. As Suntein and Thaler (2008) argue: "it tries to influence choices in a way that will make the choosers better off, *as judged by themselves*" (p. 5, [their emphasis]). Individuals are free to choose, despite the architecture of choice has been modified for their good. A core idea is that when choice shaping is not avoidable then it must be permissible. Yet, paternalism may still set limitations on freedom and this is why it is morally problematic. It is, of course, also open to abuse, and this is why Sunstein and Thaler (2008) set the condition of *publicity*, that is, any nudge must be transparent, open to alternative choices and not deceptive. However, there remains a 'paradox of nudging', namely that it works more effectively as long as information is not explicitly given or set in a certain manner.

In any case, paternalism can take different forms: soft, libertarian, tolerant, perceived. Libertarian paternalism is a relatively weak and nonintrusive type of paternalism, because choices are not blocked or fenced off. Such libertarian and 'softer' approach to paternalism and nudging is nowadays embraced by many scholars such as Sunstein and Thaler (2003) and, more recently, by Floridi (2016), who advocates what he calls "tolerant paternalism" that, in short, aims to modify the levels of abstraction of the choice architecture by educating users to make their own critical choices and to assume explicit responsibilities. Instead of being a threat to autonomy, tolerant/libertarian paternalism can in fact also enhance autonomy. In particular, all those persuasive technologies that intend to promote attitude-behavior consistency seems powerful support for one's autonomy

when applied to oneself. The debate, however, is still arguably open-ended. To some extent, it has to do with what we mean by freedom. In any case, to paraphrase Winston Churchill's evaluation of democracy as a form of government, persuasion and nudging may be the worst method of social control—except for all the others, *in primis* coercion.

4.3 Personalization Governance

In this last section we present reflections on the potential governance of mainstream social media's personalization to observe past experiences and alternative approaches to traditional regulation and ethical steering. There has been much debate in the last years on how to approach social media's governance. As such, we discuss in this section the most significant approaches to provide a critical and comprehensive overview of the debate. In the first part we discuss innovative approaches and its debate. In the second part, we focus specifically on the (potential) role of the European Public Service Media in influencing personalization practices and development as well as to provide standards and tools to influence or even help social media mainstream platforms in governing personalization.

4.3.1 A Polycentric Cooperative Governance

The question of regulating personalization is part of a complex regulatory system involving many different legislative acts. This also results from the fact that a large number of different stakeholders are involved in the development, production, distribution, exploitation and marketing of personalized content. The three main categories are users, content providers/producers and distributors/platforms, which in turn are split into a number of different types of actors – much more diverse in the digital environment than in the analogue environment. The regulatory space has also increased dramatically due to the borderless nature of the digital world, as has the technical expertise needed to create effective, appropriate and enforceable rules. For these reasons, and as already argued, the involvement of various stakeholders in regulatory approaches has become much more important.

Together with the centrality of subnational authorities, there is need to move away from homogenous top-down model towards a decentralized, reflexive, effective, collaborative and cooperative framework that is *'polycentric'* (Finck, 2018). Regulatory conversations

on social media platforms are already polycentric in that they are transnational and multi-sectoral. Polycentricity is indeed inherent to new governance models. Unlike traditional conceptions of law that rely on a unitary source of authority, new governance is predicated upon a dispersal and fragmentation of authority, and rests upon fluid systems of power sharing. A polycentric regulatory network, encompassing additional stakeholders, would be preferable for it is likely to generate better results and echoes the nature of social media's platform economy itself.

Social media governance in fact does not depend exclusively on a single source of accountability or regulation, but rather on more complex and multistakeholder systems of governance. Helberger, Pierson and Poell (2018) argue that "the realization of public values in platform-based public activities cannot be adequately achieved by allocating responsibility to one central actor (as is currently common practice)", and instead envisage pursues a "dynamic interaction between platforms, users, and public institutions." Social media policy is in fact moving towards a *cooperative responsibility*, that is, the result of the dynamic interaction between platforms, users and public institutions such as PSM. Thus, they identify four key steps to organize the (re)distribution of responsibilities:

1) To collectively define the essential public values at play in particular economic activities and modes of public exchange.

2) Each stakeholder (platforms, governments, users, advertisers, and others) to accept that they have a role to play in the realization of these values. The definition of the public value(s) and the specific contribution of each stakeholder are dependent on the context – the sector, type of service, regulatory situation, and socio-cultural sensitivities.

3) To develop a multi-stakeholder process of public deliberation and exchange, in which agreement can be reached between platforms, users, and public institutions on how important public values can be advanced. In this case, governments would need to give some space to online platforms to experiment and operationalize workable solutions, without putting the realization of public policy objectives entirely at the mercy of self-regulatory initiatives.

4) To translate the outcome of public deliberation and agreements into regulations, codes of conduct, terms of use and, last but not least, technologies (e.g. 'by

design').

Furthermore, Helberger et al. (2018) have also indicated that this can take various forms: a) organizational and, more importantly in the context of personalization, design responsibility for platforms, b) active participation, empowerment, and real responsibility for users, and c) creating frameworks for shared responsibility and shared values for governments, considering platforms and users as partners in regulation rather than subjects.

One of the main problem with self-regulatory measures in fact is that they often lack means that would effectively help to measure, evaluate and audit the actions which the industry stakeholders have committed to. For example, they are at the moment unable to harmonize reporting and takedown mechanisms or shed light on the decision-making processes of both automated algorithmic and human content review systems. There are further issues at stake. For example, government oversight of social media RSs faces many significant challenges, both practical and principled (Leersen, 2020):

- To begin, government authorities are capacity-constrained, particularly with regard to the technical expertise required to perform complex algorithmic auditing. This is especially true for horizontal agencies such as competition and data protection authorities, for whom RSs risk being overshadowed and overlooked.

- Sectoral proposals, instead, would in many cases require the creation of entirely new oversight bodies. As said above, what makes this particularly challenging is that in Europe media policy is largely a national affair, without a clear institution at EU level capable of performing a monitoring role. Indeed, EU governments have repeatedly shot down proposals for creating a supranational media authority. National-level action could, on the other hand, result in a duplication and fragmentation of efforts.

- Government auditing powers also raise issues of 'second-order accountability': is the governance system itself sufficiently open to outside scrutiny? If government determinations rely on privileged access to confidential data, which is not accessible to broader publics, it may be difficult for citizens to fact-check and second-guess government policy in this space. Without broader forms of second-

order transparency and accountability, the legitimacy of a technocratic, command-and-control approach in such a politically sensitive, value-laden context can therefore be called into question.

Despite mainstream social media's platforms have devised a number of self-regulatory measures, these have generally failed to overcome criticisms. The instruments deployed so far can be grouped into three general categories, pursuing accountability towards three different sets of stakeholders (Leersen, 2020):

1) *User-facing disclosures*, which aim to channel information towards individual users in order to empower them in relationship to RSs (e.g. Facebook's 'Why Am I Seeing This' feature). The aim of such transparency is to inform users about their available options so as to help them form their own preferences, appealing to values such as individual autonomy, agency and trust. If platforms fail to do so, users can take their activity elsewhere. This user-facing approach to transparency can be seen in several European instruments, notably the GDPR.

2) *Government oversight*, which appoints a public entity to monitor RSs for compliance with publicly regulated standards (e.g. auditing agency, public disclosures).

3) *Partnerships with academia and civil society*, which enable these stakeholders to research and critique RSs (esempio data sharing for researchers). Yet, these are often met with skepticism for several reasons. Above all, creating meaningful transparency arguably runs counter to platforms' incentives: they have a commercial interest in monetizing traffic data and insights, and thus in keeping this information exclusive, and a political interest in avoiding negative publicity.

Instead of self-regulation and regulation, it is co-regulation of social media's platforms the likely most effective approach to the governance of personalization systems and practices. This would lead to a situation in which the EU defines legislative standards that are subsequently implemented by platforms, and it would be characterized by a number of features (Finck, 2018), including:

(i)   participation and power sharing, as power is not monopolized at supranational level but shared by those participating in the exercise (Helberger et al., 2018);

(ii)   multi-level integration as innovative regulatory solutions have been adopted

by subnational actors across the Union that now serve as blueprints for regulation elsewhere;

(iii) diversity and decentralization given that the impossibility of uniform regulation is acknowledged;

(iv) deliberation among multiple stakeholders takes place as the EU doesn't regulate in isolation; the resulting rules would moreover be characterized by

(v) flexibility and revisability as they are constantly evaluated and rapidly adapted; and

(vi) experimentation and knowledge creation as the various concrete applications of the general standards will reveal several indicators as to the suitability of a given standard. These features of co-regulatory approaches would be further developed if a wider variety of actors were involved. As such, polycentric decision-making allows for the concentration of knowledge, which is naturally dispersed across society. This would remedy the currently prevailing information asymmetries and allow to make regulation fit for purpose.

In addition, multi-stakeholder bodies that include governments at various scales, industry, consumers and providers but also social scientists and other stakeholders provide room for deliberation whereas platform self-regulation and regulation by code do not. As such we might argue that co-regulation can be more consistent with democratic, participatory, and representative ideals, especially where it operates as a polycentric process of cooperative responsibility. A process fashioned in this manner recognizes pluralism and allows for decentralization in addition to facilitating experimentation.

Co-regulatory solutions bear the potential to marry the benefits of both regulatory paradigms in harnessing the effectiveness of platform's involvement in the regulatory process with public oversight. In this process, which relies on cooperation and dialogue, platforms and public authorities are collaborators rather than adversaries. Such a process can allow for more informed decision-making, easier enforcement, and continuous review and assessment. The experimental nature of this process allows for mutual learning and the identification of best practices as well as for a dynamic adaptation of the relevant rules over time.

4.3.2 Towards Personalization 'Algorithmic Sovereignty'

After having analyzed the politics of mainstream social media over personalization practices, it is easy to realize how these explicitly counteract the possibility for its participants to gain more sovereignty. Facebook, in particular, denies all possibilities of participation in the decision-making process of its own algorithms, as well as strictly regulates the opportunities of interoperability for the data it gathers. This is why it is also important to speak of personalization algorithmic sovereignty, as the moral right of a person to be the exclusive controller of one's own algorithmic life and, more generally, the right and capacity by citizens as well as democratic institutions to make self-determined choices on personalization algorithms and related design choices. Self-determination is essentially at issue here, and the battlefield is whether or not users will be able to exercise their actual will when it comes to personalization. Until now this premise – essentially that it matters whether people control personalization and its algorithms – has never been seriously commented, stated or even discussed by social media platforms.

The notion of 'algorithmic sovereignty', based on a constructionist approach that posits that metaphors organize users' perceptions and contribute to creating new realities (Krippendorf, 1993), might be a fruitful notion to develop further. To effectively address new ICT-related challenges it is indeed needed to enrich our conceptual toolbox (Floridi, 2015). Taking the notion of "algorithmic sovereignty" as a metaphor – firstly explored by Roio (2018) – we are interested in what could be emphasized or neglected in its use and the extent to which its use may contribute to reiterate or deconstruct mythologies of digital infrastructures, with a specific focus on personalization algorithms in social media.

In the last years, similar related concepts have been introduced and discussed in academia and media (Couture and Toupine, 2017), notably data sovereignty (e.g., De Filippi and McCarthy, 2012), digital sovereignty and technological sovereignty (DIEM25, 2019). Such increasing interest in the notion of sovereignty has the ultimate goal to assert some form of individual and collective control empowerment over digital technologies. In particular, the notion of technological sovereignty is used to refer to initiatives that create alternatives to commercial and/or military technologies, often with a great emphasis on free and open source software and hardware. As such, it is also framed as an opposition to a hegemonic power, namely the United States and its biggest private tech companies (also called GAFAM: Google, Amazon, Facebook, Apple,

Microsoft). In its broad sense, technological sovereignty is often meant as a rupture from "state sovereignty", while other times – similarly to the notion of "computing sovereignty" by Richard Stallman (Couture and Toupine, 2017) – as the duty of public services to keep total control of their computing tasks in the benefit of citizens.

We can ideally distinguish between two poles of algorithmic sovereignty – weak sovereignty and strong sovereignty – between which different approaches may be situated: "weak algorithmic sovereignty" would refer to the negotiation between states and social media platforms to enact certain technical conditions, while "strong algorithmic sovereignty" would refer to the actual common ownership of algorithms by the people. This latter interpretation widely overlaps with the concept of technological sovereignty. Algorithmic sovereignty can thus be considered as a sub-group of the broader notion of technological sovereignty but, at the same time, it can provide more specific conditions on what the individual and social control over algorithms should be.

In this context, it is easy to think of the free software movement and the open source business models as the paradigms that enable the scrutiny of algorithms by thousands of experts, as well the freedom to modify them and distribute modifications (Roio, 2018).[33] Yet, even when the access to algorithms complies to these major ethical concerns, would that be enough to verify a condition of sovereignty for all living participants? Is transparency, as opposed to secrecy, a sufficient condition to make algorithms functional to the creation of an intelligible society? And how to effectively translate these principles in the context of personalization algorithms in mainstream social media?

Nowadays, the only way for participants to gain more sovereignty in mainstream social media is to create an entirely new territory.[34] This is hardly suitable due to *network effects* which occur when the value of a platform to any user increases exponentially with the number of already present users. Normally, this leads to monopolies (Lovink and Rasch, 2013). The history of alternative media is indeed a history of enormous challenges, as the consequence of a political economy that limits the possibilities for civil society because hearing alternative voices is a matter of money and political resources that afford

---

[33] More than 30 years ago the free software movement put forward an ethical and legal framework that establishes four fundamental freedom: 1) The freedom to run the program as you wish, for any purpose. 2) The freedom to study how the program works and to change it as you wish. 3) The freedom to redistribute copies so you can help your neighbour. 4) The freedom to distribute copies of your modified versions to others.
[34] A directory to all alternative platforms can be found on switching.social.

visibility (Fuchs and Marisol, 2015). What has changed in the last decade is the vast societal dependency on information intermediaries' networks.

The only opportunities for social media users to gain algorithmic sovereignty lies in two possible lines of action (Roio, 2018): either to "fork"[35] the project, with great expense of resources and effort and with an uncertain outcome – this might even be facilitated with the right to portability afforded by the GDPR – either to rely on a strong regulatory framework that opens Facebook's governance to its participants. Yet, both options are very uncertain and unlikely to succeed. On the one hand, the 'critical mass effect' does not make data portability a useful mechanism for migrating to other networks. On the other hand, the focus of the GDPR on data consumer protection does not actually guarantee the protection of collective rights.

The pursuit for more individual agency and social control of platforms is receiving increasing attention both in theory and in practice. For example, the *European observatory on Algorithmic Sovereignty*, is a new collective place of documentation gathering research, developments, events and projects related to the topic.[36] Also, the project *Fediverse* – with roughly 5 million users[37] – is an example of a social media where users approach an algorithmic sovereignty. The name is a portmanteau of two words "federation" and "universe", a common name for federated social networks running on free open software on a myriad of servers across the world. The 'Federation' refers to a global social network composed of nodes that talk to each other. Each of them is an installation of software which supports one of the federated social web protocols. Fediverse networks are designed to be run by anybody, free to choose and register on any server and choose the person who will be in charge of its data - the administrator of your server. Selecting a server for its politics of data, however, is an expert choice one cannot expect from the large public.

Another significant approach is taken by the project *tracking.exposed.* By scraping and collecting the data individually recommended to the user on supported platforms (Facebook and Youtube above all) it collects and stores evidence which is usable in new ways. A result of such data reuse is the comparison of personalized experiences. By

---

[35] In software engineering, a project fork occurs when developers take a copy of source code from one software package and start independent development on it, creating a distinct and separate piece of software.

[36] See https://algosov.org/

[37] See https://fediverse.party/en/post/fediverse-in-2019

assuming everyone has an individualized perception of the public discourse the project extrapolates the topics and let users compare, which might allow for improved critical judgment regarding the quality of information received. Also, this data is shared for research purposes, with NGOs and academic teams. By creating new profiles, with as many of the possible variables (age, friends, pages followed, likes) under control, it is possible to test how algorithms lead to divisive perceptions. As a side effect, it has collected a library of public information. Inspired by the political concept of the European Data Commons (DIEM25, 2019), the project keeps data in the collective interest, as far as it does not expose any individual without his or her consent. Yet, full anonymity cannot be offered at the current state of research. Therefore, the team has to make a privacy impact assessment for each form of data reuse.[38]

Now, considering the above experiences and experiments, we summarize some fundamental preconditions for achieving what we defined as personalization algorithm sovereignty;

4.3.2.1 Education and Media Diet Awareness

First of all, and more obviously, there is a general need for 'algorithmic literacy'. We can define it as the basic knowledge on how filtering mechanisms and design choices function and what their impact on one's own life is and may be. This can be cultivated with regular updates on personalization research, visualization tools and dashboards. Some paradigmatic experiments have already been done: i) in general through browser extensions (Reviglio, 2019b), ii) with interactive sliders (e.g. gobo.social) or iii) showing users their filter bubbles and help to 'burst' them (Nagulendra and Vassileva, 2017). Clearly, in order to gain more control algorithmic literacy need to go hand in hand with digital and media literacy.

Relatedly, an updated and comprehensive media diet awareness is preconditional. The development of 'natural metrics' usable to evaluate the informative experience produced by the algorithm, for example, is a critical issue; the percentage ratio of photos/videos/text, the time spent on the platform, the number of posts encountered, or which sources have been consumed. This information should be collected on the device and given back to the user to primarily increase self-awareness. We might assume that

---

[38] See https://facebook.tracking.exposed/data-reuse

negotiating these parameters and design choices should be an effort made and/or supported by the current democratic institutions and expert network bodies, such as the w3c, IEEE, or IETF.
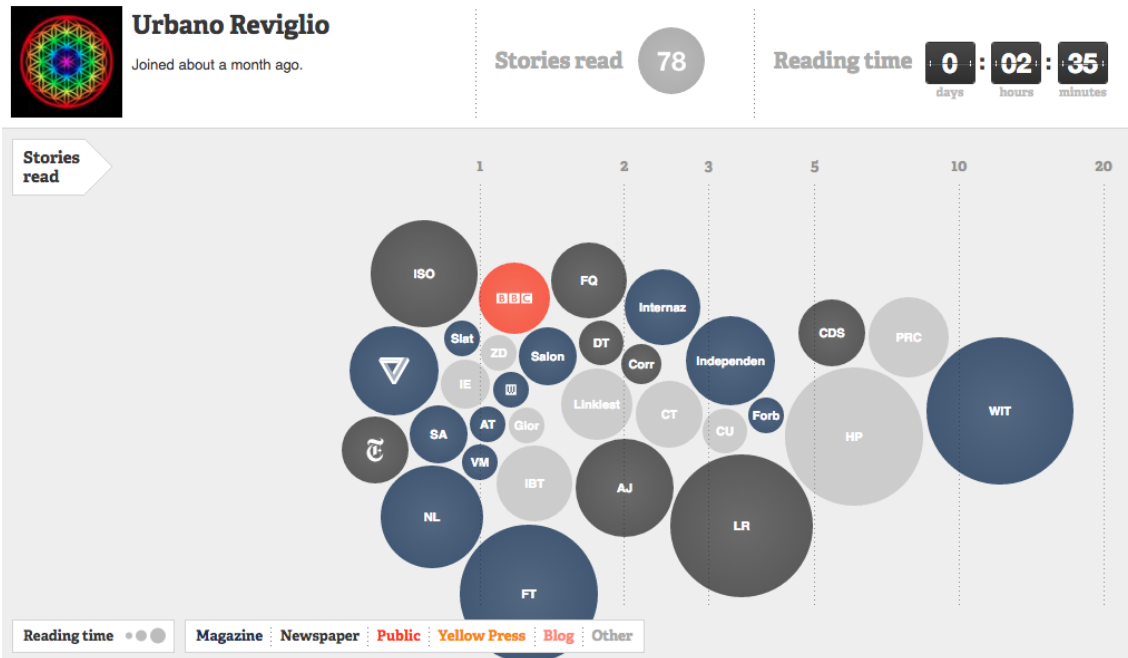
Figure 21. Scoopinion, a browser add-on that displays user's news consumption habits. Larger circles are news outlets that the user consumed the most items (screenshot, May, 2019).

4.3.2.2 Neutrality and Transparency

Ideally, another significant precondition is 'content neutrality', meaning platforms should become agnostic towards content. This is extremely difficult to realize, if not impossible. A potential way to approximate that is the following; platforms should return machine-readable, unfiltered, chronologically-ordered data. Then, a reader device would receive these data, and an algorithm would filter and prioritize. This could be a fundamental architectural requirement for a strong algorithmic sovereignty. By making sure that the filtering happens on the client side, it is ensured that the platform becomes effectively neutral, and nobody, except the individual, will end up knowing what was watched and for how long. Such neutrality would also help to prevent deception by design and, eventually, to cultivate a related aspirational goal: 'algorithmic neutrality', which is the idea that content should spread freely without biases.

Content and algorithmic neutrality might be also achieved with a 'right to profile transparency', through an infrastructure that supports what the philosopher Mireille Hildebrandt has defined as 'counter-profiling' (2015). In addition, platforms and algorithm developers could routinely disclose to their users and the public any experiment that the users were subjected to. An independent board to review and approve experiments in advance could also be established. This would be particularly in line with two intertwined newly discussed human rights: the 'right to not be measured, analysed or coached' (van Est et al., 2017) and a 'right to cognitive sovereignty' that ought to protect individuals' right to mental self-determination (Yeung, 2018b).

More generally, algorithm auditing need to be supported as well as data sharing. In particular, it is fundamental to guarantee a strong separation between filtering mechanism and data. Filtering mechanisms should be, as much as possible, stateless and idempotent, that is, capable to return the same result if executed twice. This is a formal way to describe a more easily auditable machine. On the other hand, personalization algorithms that use neural networks keep updating their internal state, making it impossible to perform certain analyses either ex-post or ex-ante. If the same result can be accomplished using a simpler computing system, it would also be simpler to explain, standardize, teach and assess outcomes. If it is possible to keep a simpler computing system as standard, the technical debt caused by the lack of explainability of algorithms might be avoided. Moreover, metadata which describes the content produced is fundamental to control. Functionalities of research, analysis and prioritization might perform better with metadata usable by end-users. Even if it is a laborious process mainly for skilled individuals, metadata represent an opportunity to differently index and retrieve content so as to eventually increase content quality over quantity.

4.3.2.3 Design Requirements and Information Discovery

There are, then, user-centered preconditions to negotiate. Being essential to personalized information consumption, design choices and affordances should be to some extent adjustable, especially for what concerns attention management. For example, disabling the auto-loading of posts (similarly to the auto-play function for videos which is usually set by default) and decide to scroll n posts anytime one logs in. Another interesting example is the possibility to hide metrics in order not to be influenced by comparative

quantifications (such as the number of likes and shares). The quantification and maximization of social interactions, in fact, contribute to create a "culture of performance" (Castro, 2016) that seems to be negatively correlated with well-being (Verduyn et al., 2017). In essence, all those basic design choices that ultimately affect information behavior and personalize one's information diet ought to be socially negotiated, implemented and, ultimately, adjustable.

Furthermore, there are information self-determination and media pluralism concerns to cope with. In particular, the capacity to reach a balanced information diet from both an individual and a political perspective (Eskens et al., 2017) and, at the same time, make sure that a range of viable 'perspective widening' tools are provided (Delacroix and Veale, 2019).

On the one hand, a 'right to receive information' – as guaranteed by Article 10 of the European Court of Human Rights – recontextualized in the digital context could help to establish what news consumers legitimately may expect from the news media with respect to the diversity or relevance of personalized recommendations (Eskens et al., 2017). In practice, this right could result in an effective explicit personalization – especially related to political news.

On the other hand, a 'right to information explorability' to increase serendipitous encounters and to reduce potential filter bubbles and echo chambers (Reviglio, 2019). This could be achieved increasing information findability and discoverability, for instance with the possibility to choose the accuracy and diversity of information filtering. In practice, this might imply more interactive control of the algorithmic outputs and, as such, discoverability throughout sliders, topic categories, filters to navigate the information by source and keyword, or even algorithmic recommender personae as specific avatars to filter information (Harambam et al., 2018). As Hildebrandt (2019) advocates, societies must demand companies to explore and enable alternative ways of datafying and modelling the same person. Multiple and dynamic profiling could be paramount to this scope.

By doing this, we should in fact consider new approaches to data collection, such as those used in qualitative social science (Greene and Shmueli, 2019). By incorporating more diverse forms of personal data into personalized scores, we could be able to reduce the gap between one's identity as a person embedded in social and cultural space and as

a feature vector embedded in feature space. In fact, personalization processes will never be perfect.

4.3.2.4 Critical Issues

Certainly, the above outlined general preconditions need to be discussed further as they may have potentially unintended and undesirable consequences, impossible to systematically assess or even speculate in a short article such as this. Yet, we briefly identify some major weaknesses and critical issues.

If institutions empower users who can't exercise their algorithmic sovereignty, they may rely on default choices, and this may only leave aside the weaker users and enable other actors as well as mainstream social media. This scenario is not different from a free market where the players are not equal, and it might even legitimize current mainstream platform at the expense of potential emerging competitors. Now, if we imagine the "liberalization" of algorithms would happen tomorrow, it is not hard to believe that a technocratic group of skilled individuals would try to claim their algorithm to be the best. By leveraging existing conditions of influences, they might just reproduce the same form of algorithm oppression but with a variety of small actors in the market.

Also, this liberal market approach might lead to the misleading message that an algorithm can be better than another one. The perfect algorithm, in absolute terms, does not exist. Every one of us has different priorities, interests and time availability. The correct algorithm is the one that best fulfills the needs of the individual, and this might not be true anymore if one begins to change. The fitting algorithm thus cannot be permanent. Instead, the above mentioned natural metrics, dashboards and tools are fundamental, and should be imagined as the regulation imposed in the food industry (to declare the allergens, ingredients, kilocalories etc).

Because algorithm sovereignty is a political challenge, the solution cannot be only technical. Algorithmic sovereignty indeed implicitly calls for more responsibility over citizens which ought to be able to decide on their own instead of delegating to someone else. We believe that institutions like schools, academia and public service media ought to be proactively responsible for the most significant preconditions: digital, media and algorithmic literacy.

Altogether, the above projects and technical preconditions question – and to some extent disrupt – the current power asymmetry between users and the public towards mainstream social media platforms and their personalization algorithms. Yet, they could also renew trust between platforms and users. Ultimately, the aim is to discuss, identify and test a series of essential technical preconditions to ensure full autonomy, transparency and control of users and civil societies. An important player in this context, especially in EU, may be Public Service Media which might undertake a new critical role in the emerging digital landscape.

4.3.3 The Role of Public Service Media

In this Section we discuss the role of Public Service Media[39] – the digital version of Public Service Broadcasting (PSB) – and attempt to answer the following questions; what might be the role of PSM in this emerging personalized media landscape? How can it help to counteract the detrimental principles that shape social media today, such as high speed, superficiality, tabloidization, short-term pleasure, hedonism and homophily, and to sustain Public Service Broadcasting (PSB) traditional values, such as universality, publicness and quality? Can it represent an institutional mediator able to steer the development and governance of personalization and proactively help to solve the crisis of media and, more generally, reinforce democracy and human rights? Last but not least: how can PSB and PSM still survive and maintain a relevant role in shaping the media landscape?

The notion of public service in relationship to media was developed in the early 20th century under a specific set of political, technological and social conditions. Perspectives on this relationship changed across time and space. Yet there remained 'an overlapping consensus on certain core normative criteria' that can be categorized into three main principles, where the latter two are derived from the first (Picard and Pickard, 2017): (1) Enhancing, developing and serving social, political and cultural citizenship; (2) Universality; and (3) Quality of services and output. Van Cuilenburg and McQuail (2003) identified three phases in Western communication policy-making paradigms that show

---

[39] Splichal (2007) gives a concise definition of PSM: "In normative terms, public service media must be a service of the public, by the public, and for the public. It is a service of the public because it is financed by it and should be owned by it. It ought to be a service by the public – not only financed and controlled, but also produced by it. It must be a service for the public – but also for the government and other powers acting in the public sphere. In sum, public service media ought to become 'a cornerstone of democracy'" (p. 255).

the evolution of these principles: (1) emerging communications industry policy in the period before the Second World War (WWII), (2) public service media policy after the War until the 1980s, and (3) new communications policy since the end of PSB monopolies.

The first phase was characterized by public monopolies with the limited goals of protecting the interests of government and nation, promote innovation and provide an efficient public service, whereas the second phase was characterized more by normative and sociopolitical rather than technological and economic considerations. Media policy started to promote diversity of ownership and content, limit monopoly and deal more effectively with the press. Such policies went much further in Europe than in the US, in particular regarding PSB, despite significant differences from one country to another. In the last phase, the rise of commercial television, which, from the 1980s, led to a majority of private stations by the mid-1990s, gradually reversed the monopoly PBSs used to have. Even if it retains some legitimacy and popularity, the normative policy paradigm of the postwar era declined in authority and scope and changed in the means for achieving its goals. This was not much due to technological changes but mostly to economic and political forces legitimized by the ambitions of media companies and governments. As soon as PSB became the exception rather than the rule, its activities and funding, if not its very existence, began to be considered as a disturbance of market relations (Gulyás and Hammer, 2014). PBSs indeed changed their programs to better adapt to audiences and advertisers while taking over certain orientations of private competitors, such as competition and cost awareness. As a consequence, PSB core values began to be eroded (Van Cuilenburg and McQuail, 2003).

Despite differences that relate to the national context of PSBs, it is still widely believed that basic functions, such as a low-cost and universally available reliable provision of information, education and culture, and the catering for minority tastes and interests, cannot or will not be sufficiently served by the commercial market (Gulyás and Hammer, 2014). Yet, it is believed that PSB should not be restricted to remedying 'market failures' but to respond to the needs of citizens – which are indeed different from those of consumers. Thus, not only should it provide a counterweight to the commercial media but it is expected to set quality standards for the whole media landscape. Anyhow, fierce

competition and new consumption patterns in the television landscape made it increasingly difficult to attract and sustain audiences.

At the beginning of the digital revolution, the discussion about the future of PSB revolved around two competing visions: those who were critics of the further commercialization of PSB and those who simply wanted to save PSB by adapting it to the system (most of the PBSs themselves and policy-makers). In particular, the European Broadcasting Union (EBU)[40] followed the second path. The EBU Digital Strategy Group (EBU, 2002) considers PSB as an 'island of trust' amidst multimedia companies offering linear broadcasts and online programs. Public broadcasters have learned to adapt to predominantly commercial media markets, by choosing a middle way between popularization and purification in their program strategy.

As a consequence of the new emerging landscape, the transition from PSB into PSM was debated. In general, this implied the extension of public services beyond radio and television to encompass the full specter of the Internet. As a consequence, public broadcasters were to redefine their relation to the public. One of the main challenge remained the gradually diminishing reach of PSB among minorities and 'problematic groups' such as younger generations, migrants and the less educated (Burri, 2015). To overcome this, many PBSs have started experimenting with new ways of legitimization, accountability and transparency towards citizens and society. PSBs initially started using social media to reach and engage new audiences. Such turn to the digital environment also stimulated PSBs in Europe to experiment with a public version of "social TV". Online participation has been considered a key-strategy to regain position in national arenas. Such commitments, though, do not appear particularly effective in shaping information consumption, and therefore are often considered insufficient.

4.3.3.1 Towards a New Public Service Media?

Most scholars agree that PSB as well as PSM require new forms of justification, not simply to save its role as a public institution but to educate a generation grown up in a global multi-platform world. A new media policy paradigm is indeed needed, and the

---

[40] EBU is an alliance of public service media (PSM) organizations, established on 12 February 1950, made up of 72 members in 56 countries, yet unrelated to European Institutions. Its main objective is to assist its Members in this period of unprecedented technological changes.

three main paradigms discussed earlier highlight the deterministic influence not much of technology but of social values.

How PSM has reacted to the development of personalization so far? A majority of PSM in Europe are currently moving in the direction of digital and algorithmic personalization (Van den Bulck and Moe, 2017), assuming the role of a "Public Service Navigator", meaning a mechanism for influencing the conditions of access to content, particularly its visibility, discoverability, and usability (Burri, 2015). The encouragement of content discovery to disrupt filter bubbles has been already stressed by EBU (2017). Actual specifications for RSs are, however, still to be developed. Due to fears of filter bubbles and echo chambers, the attention of EBU in the attempt to steer such development, for instance, was focused exclusively on exposure to diversity—at the cost of considering the need for common arenas of discourse, thus the values of universality and publicness, and how broadcasting and interactivity/choice can reinforce each other. Moreover, many institutions with similar histories and comparable media system frameworks are taking up different positions. Some consider the possibility to reach the value of universality through personalization, while other consider personalization to work against it. Also, some privilege implicit personalization over an explicit one, with very different outcomes on PSM goals.

This may not be sufficient. Some scholars advocated alternative and more radical proposals such as a 'Public Service Internet' (Fuchs, 2017) or 'Public Service Algorithms' (Burri, 2015; Bennett, 2018). Burri (2015) also advocates the idea of a "Public Service Navigator". Yet, other scholars have been skeptical of the possibility to reform the public service system as such. As seen in Section 3.1, the debate has thus mainly revolved on platforms' regulation versus self-regulation (Napoli, 2015). In the European media policy it is in fact assumed that in most cases soft law promotes self-regulation, even if doubts concerning its effectiveness and the transparency of certain tools remain. A paradigmatic example is in fact illustrated by EBU which is currently seeking to support its members as they struggle with both the opportunities and challenges of Big Data through the *EBU Big Data initiative*, an interdisciplinary network launched in 2015 that aims to guide PSM in the implementation of data-driven strategies.

To legitimize any more radical policy intervention as well as to guarantee an online media environment aligned with the most profound individual and societal needs, PSM

should especially focus on the individual cognitive factors that challenge the experience of diversity online (Hoffmann et al., 2015). This means to employ a user-centric perspective and to especially focus on media literacy. To achieve this, we introduce two critical perspectives. Firstly, we argue the need to acknowledge the significance of power relations, that is, applying an agonistic approach to media pluralism. This could result in providing tools for contestation and for proactively seeking – but also be exposed to – challenging and minoritarian information. Secondly, we introduce an often underestimated element to improve users consumption: attention which can be translated into 'attention self-management' to actually counteract the risks of addictive and superficial usage of social media and, more generally, the Internet itself, so as to foster a slow and more constructive consumption of media (Rauch, 2018).

To begin with, media pluralism as a normative principle is vague and under-theorized, and it is not a reliable indicator of a society's level of freedom and it can also create only the illusion of content diversity (Karpinnen, 2008). Moreover, in the digital age it is becoming less clear in which sense it is meaningful to speak of media pluralism, if the consumption is potentially characterized by limitless choice. As such, Karpinnen (ibid) advocates the idea of *'agonistic pluralism'* to the context of media politics. The starting point is that media pluralism cannot be conceived only in terms of heterogeneity and a diversification of options but it needs to be analyzed in connection with the structural relations of power that define the criteria that guide information selection and limit available choices. Such perspective helps to defend concrete institutional arrangements in media policy. PSM can indeed be seen as a key tool in creating a plurality of power structures that is more open to democratic contestation, that resists the hegemonic tendencies of the market, and that reduces semantic inequalities. By analyzing digital tools that help to burst filter bubbles and weaken echo chambers, Bozdag and Van den Hoven (2015) concluded that not all democratic models are represented in these tools, and that agonistic elements should also be included so that the needs/voices of minorities can be heard.

Furthermore, to regain a relevant role in shaping users' information consumption and sustain knowledge acquisition, a special focus must be given to attention and its management. Attention management is indeed considered one of the most important skills for the 21st century, even replacing time management (Davenport and Beck, 2001).

The ability to control distractions and stay focused is indeed essential to produce higher quality results in all aspects of life, particularly in hyperconnected societies (Floridi, 2015). Yet, the Internet is nowadays mostly based on attention as a commodity. The attention of the users is, therefore, exploited, as it was – to a much more limited extent – in the pre-digital era. As such, users hold in a subordinate position. It is therefore imperative for PSM to counteract systems that are designed to capture users' attention and to provide tools that give them the autonomy to opt-out from certain design choices and better manage their own attention – also called attentive user interfaces. Psychological factors and media literacy are indeed paramount to overcome the challenges to experience diversity online and, more generally, for public discourse (Hoffmann et al., 2015; Carson, 2015).

As we argued in Reviglio (2019), PSM could provide affordances and tools for the Internet – and especially in social media – according to its traditional values, yet with updated goals. As such, the concept of meta-design has been preliminary framed as the re-design of a digital environment in order to change design choices as well as to inform and educate users; more broadly, as the act to reshape and enrich a website's or platform's information architecture and design throughout user-friendly tools such as aggregators and browser extensions so as to primarily influence and improve user experience as well as to sustain media literacy.

In such endeavor, a user-centric perspective is implicitly applied to reflect on ways to extract value from users' profiles meaningfully, to illustrate blind-spots, and to ultimately sustain the experience of diversity. Importantly, the paper also shows that most of these architectural modifications can be done without necessarily dealing with social media platforms. In fact, it is possible to help users to navigate the complexities of the new and fast-changing media landscape by exploiting the potential of 'meta-design'. This can be primarily enacted through APIs (application programme interfaces). These are in fact one of the most important sources of data regarding RSs through which outside researchers can download platform data in bulk (Leersen, 2020).

Let me give some paradigmatic examples; *Social Fixer* is an extension that improves user experience in Facebook and allows several design choices changes. Interestingly, it can set Facebook's NewsFeed to switch to "Recent Stories First" – the chronological feed, which is not to algorithmically curated. By default, in fact, Facebook only shows you the

top stories rather than every post made by your friends. Yet anytime one logs out of the website, the algorithmic curation reset by default the NewsFeed. Such simple design choice would permanently free users from the algorithmic influence of Facebook. Another example is the possibility to hide any metrics' posts in order not to get captured into so-called 'trendism', that is, be influenced by comparative quantifications (e.g. likes and shares). This could help users to self-control more effectively their attention. Yet, not only this project highlights the potential for re-shaping certain design choices but also the potential to communicate with messages and, thus, inform and educate users during their navigation.
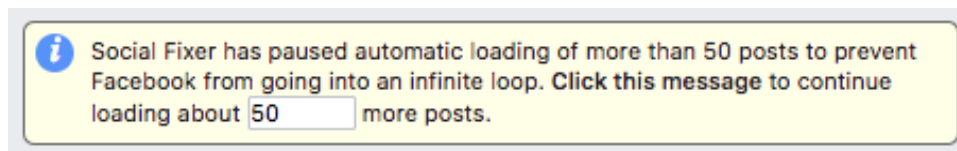

Figure 22. Limited Post Loading in Social Fixer Affords (screenshot, May, 2019)

Another example of an exploratory (and serendipitous) tool is provided by the MIT project *Gobo*, a social media news aggregator with sliders that users can control to filter information.[41] This project sheds light not only on the possibility to provide sliders to filter media content based on certain democratic principles but, more interestingly, the possibility to aggregate and re-filter the information a user might want to consume in social media. Hence, it shows the possibility to bypass the algorithmic curation of information intermediaries. This, however, applies only to public posts. Yet, there may be implemented specific privacy settings to permit them to opt-in so that their posts may be included in the re-filtered feeds. Similar aggregators can be designed also for PSBs' websites and video sharing platforms like Youtube in order to increase information discoverability and the experience of diversity.

---

[41] Other similar news aggregators are *Newsbird, Allside, News 360, Digg and Flipboard*.
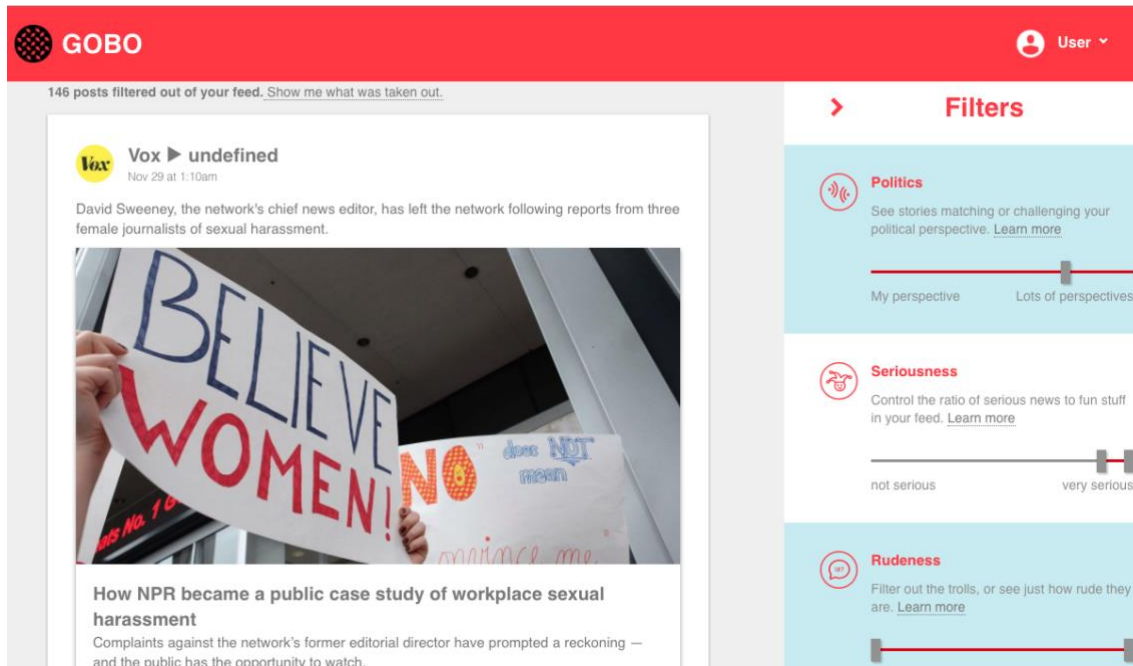
Figure 23. News Aggregator Gobo Made by MIT Social Media Lab (screenshot, May, 2019).

More generally, a significant mean to gain more influence over information consumption is by being aware of the production and consumption of information (*what* and *why* information is filtered and *how* one consumes it) (Thurman and Schifferes, 2012). For example, it is possible to encourage blind-spots explorations. In this respect, a notable example is the visualization tool developed by Nagulendra and Vassileva (2016) which displays to users their filter bubbles, showing them which categories and friends are in their bubble and which ones are not, allowing them to control the algorithm by manipulating the visualization to escape the bubble, by adding or removing friends of a certain category to the filters. To our knowledge, this is the only attempt to show filter bubbles to users. Yet the quality of the design can certainly be improved and its scope expanded. While data collection may prove to be limited, users may opt-in to provide their personal data. Equally important, notifications management could be similarly useful to avoid distraction. Eventually, so-called *information nutrition labels* (Gollub et al., 2018) can also help to manage your attention as well as increase awareness. For example, they can show the time reading, style and reliability of any article by text analysis.

The above browser extensions, visualization tools and aggregators provide insights on the potential role of PSM in improving users' experience according to its values. It is by

no means a ready-made framework any PSM could operationalize. Yet it highlights how certain extensions and tools are bringing innovations that PSM might replicate and eventually ameliorate. Despite technical challenges, meta-design can actually help PSM to collect data about users to better understand them, then inform, educate and nudge them towards its values. If such potential is employed, it might help to set the standards on a more mature information consumption, afford novel conditions to participate and benefit from social media usage and, eventually, contribute to solve the main challenges to experience diversity online (see Hoffmann et al., 2015). At the same time, it would sustain the four modalities of expressing voice[42] (Harambam et al., 2018). Ultimately, it aims to preserve not only media pluralism and users' experience of diversity but, above all, their awareness and autonomy of choice. In addition, it has the potential to help to overcome some of the major identified causes of the crisis of PBSs (such as outdated values, harsh competition and government control, see Gulyás and Hammer, 2014).

One may thus wonder why such tools have not been institutionally employed so far. To the knowledge of the author, in fact, there are in fact no noteworthy experiences of public efforts towards the creation of similar news aggregators, nudging by design or visualization tools. One explanation may be the widespread assumption that users are fully autonomous players able to manage the complexities of online information consumption.

API-based browser extensions have come under significant pressure over the past years (Leersen, 2020). Since the Cambridge Analytica scandal, in which academics helped to leak and abuse large sets of user data from Facebook, important APIs have incurred major restrictions in their functionality. This development, which Axel Bruns has described as the 'APIcalypse', has caused many widely-used research tools, both commercial and academic, to be shuttered. Of course, the quality of API access differs between platforms, and here we won't provide an analysis of these differences. Finally, another significant issue is that through technical interventions platforms are able to obstruct independent study of their RSs, leaving even the basic outputs unclear.

Furthermore, there is a lack of scientific evidence, both about the risks of Internet usage and the effectiveness of these tools. Political challenges may also be paramount. Citizens

---

[42] Building on the work of sociologist Hirshman, Harambam et al. (2018) intend 'voice' as the possibility to exert control over the data-driven processes that shape news provision.

may not want to use these tools, while platforms proved to be difficult stakeholders to engage with whenever their revenues are threatened. Somewhat paradoxically, another significant element is likely to be that such public intervention would more likely account for an irrelevant share of investments given to PBS and PSM. Clearly, these tools are extremely cheap compared to the content produced by PSB. Thus, economic incentives in this respect appear to be very weak for an institution that is especially struggling with funding. However, considering the rising awareness in the public opinion of the risks of social media and the educative potential of browser extensions, there is room to expect their employment in the near future.

Despite obvious political challenges, there are certainly several other limitations for a more proactive role of PSM. First of all, its effectiveness, in particular concerning personalization and the attempt to provide users with exposure to challenging information. Interactivity might indeed threaten the principles of publicness and universality and not fulfil the goal of maintaining a common sense and belonging. The political consequences of exposure to dissimilar views, then, have long been a subject of polarization research, but findings are largely mixed. For example, polarized individuals may polarize further while exposed to challenging information (so-called *backfire effect*). Another related-risk is that users do not exploit such tools – they do not always want to be in control (e.g. the *paradox of control*). As a consequence, inequality may be even strengthened. These issues certainly require more understanding of users' information behaviors as media landscape changes rapidly and steadily. It is not possible to make such harsh generalizations: the heterogeneity and nuances of users must be better understood as well as any context deserves specific analysis.

5. Conclusions

The use of ICTs, Internet and, more importantly, social media disrupted – and still are disrupting – various aspects of society. It is changing what we do, how we do it, how we think, how we socialize, how we believe and how we take decisions. As both individuals and societies, we still need to realize and properly grasp these changes. Above all, it is emerging a profound need to educate for a new epistemological approach. In particular, we should critically understand our everyday media consumption, its challenges and its opportunities. Then, it is equally important to develop an effective governance of information intermediaries so as to limit undesirable unintended consequences, for example of personalization algorithms. The focus of this research was in fact on personalization of social media content, its main challenges and opportunities from an interdisciplinary perspective. As such, a number of essential conclusions from the above analysis are drawn in order to realize the opportunities of personalization, and diminish possible negative consequences.

The first question of this research was widely descriptive: what are the challenges of social media's personalization for democratic societies? To answer this question, many other sub-questions had to be answered beforehand, such as: what is intended with personalization? How does it work? How it is employed by users and by mainstream social media platforms? Finally these answers are now discussed in a comprehensive way.

To begin with, there is still no accepted definition of online personalization. In fact, it remains an ambiguous and under-examined concept, lacking consensus on its essential characteristics. After all, a personalization system can be said to be composed of a recommender system which may also be complemented with certain design choices. Thus, it is not only algorithmically-based but there is also a significant influence of design choices that ultimately affect its outputs.

Personalization is not really 'personal'. It occurs largely beyond the control of users as it is mostly based on implicit data rather explicit ones. Thus, preferences are mostly inferred, not expressed. Data used are mostly derived from what is actually observed and are, of course, only a small subset of observed and observable behaviors. Also, data used come not only from the individual user, but also from other users. The main consequence is that current personalization – especially of mainstream social media – is highly

'behaviorist' its assumptions, taking into account only what can be observed, measured and recorded, often not even individually. This is an obvious reductionism.

As a matter of fact, profiling technologies that allow personalization cannot produce or detect a sense of self. They indeed create a kind of knowledge (i.e. profiling) that is inherently probabilistic. Yet, they can influence the individual's sense of self, especially in the long-run daily social media usage. Not only they threat the foundational microeconomic principle of preference formation but, more importantly, the individual "aspirational self" (or second-order preferences). The complexity and the 'multiplicity of the self' is likely not computable at all. In fact, not only no personal profile can ever entirely identify an individual, but there are always many—and sometimes radically different— ways of computing the same person. Personalization actually raises fundamental philosophical questions, such as: how to create meaningful digital versions of ourselves? Which information should constitute me as a person? And, ultimately, what is a person?

And yet, for mainstream social media personalization is merely an economic mean. In fact, we have seen how the media landscape is, in the first place, not much of an enriching space where public opinion is formed. Similarly to the other mass media, also social media are characterized by market concentration, infotainment and power inequalities. Networks effects have indeed resulted in an oligopoly of platforms that ultimately threatens innovation. Facebook, in particular, controls most of the social media's information flows (considering it also owns *Instagram* and *Whatsapp*). This gives to it ever more power. Indeed, those companies who collected since long huge data are also those who have more capacities to extract value from it and, thus, to provide (and predict) the best (or most persuasive) services, in a reinforcing feedback loop. After all, the primary aim of social media platforms and its stakeholders is not individual progress or societal cohesion but, more simply, the maximization of engagement in order to gain more computational power and market shares, as well as to sell ever more ads. The consequence is the widely critiqued social media landscape in which hate, "fake news" and addiction become prominent features. Admittedly, the critical role of mainstream social media is incresingly contested by institutions and citizens.

Social media users, however, are mostly unaware of personalization dynamics. Of course, attitudes towards personalization are very heterogeneous and they change rapidly,

thus they need to be better and consistently understood. In general, people across the world seem to embrace automatically generated personalized recommendations, often because they trust themselves and algorithms more than they trust journalists. For the majority of users, in fact, algorithms select neutrally and recommend impartially. Yet, research indicates that users are concerned about losing important and challenging information and expect to play a more active role in the interaction with personalization systems. There is still a long path towards digital, algorithmic and media literacy as well as to ensure accountability and to preserve autonomy but a fundamental issue is actually trust in media, and in the scientific community and democratic institutions.

  The potential challenges and negative consequences of personalization of social media content are serious and wide. On the one hand, there are evident challenges such as privacy, algorithmic bias, manipulation and reduced media diversity. While the formers represent broader challenges in the context of digital technologies and algorithms, the latter – media diversity – is the most specific to personalization systems and the one we focused the most in this research. On the other hand, reduced media diversity due to algorithms creates filter bubbles while strengthening echo chambers, and can eventually lead to potentially negative consequences both for individuals and societies. These could in fact fragment the audience, politically polarize and, at times, even radicalize. At the same time, they can help to spread misinformation and disinformation. Consequently, they could further reduce trust among individuals, groups and institutions, and, ultimately, weaken democratic societies.

  Thus, we analyzed whether it is possible to burst filter bubbles and weaken echo chambers. The answer is not clear-cut. Firstly, there is a lack of scientific evidences to act upon and prevent these ill-defined phenomena. As long as researchers and policy-makers cannot access real-time data of large populations of users, research will be extremely limited in unveiling these complex dynamics on a big, meaningful scale. Personalization is in fact, to some extent, unique to individuals. Thus, effects are also very heterogeneous. Secondly, there are human natural proclivities that trigger filter bubbles and echo chambers, such as homophily, selective exposure and confirmation bias. Therefore, they are primarily user-driven. As a matter of fact, there are several ways to increase own's own exposure to media diversity online. Yet, they remain phenomena that are socially and technologically co-produced. As always, education is paramount.

Most users still ignore basic personalization's dynamics and techniques. Also, re-imagine the design of social media platforms and their algorithms remains necessary. There is, however, no easy technological fix. More generally, it can be argued that more individual control over personalization, more control over the quality of information and more 'diversity by design' can represent fundamental step to mitigate the above risks.

In this regard, we discussed the value of diversity and its nuances, such as news diversity, media diversity and media pluralism. Diversity is indeed a fundamental value for democratic societies and an increasingly important concept to understand the contemporary news landscape. Of course, diversity is not an absolute value, but needs to be balanced with other public values, such as privacy, autonomy and non-discrimination, and faces significant trade-offs, notably society cohesion versus truth finding. Different democracy theories in fact value media diversity differently. It is, therefore, extremely challenging to eventually agree on how much exposure to how many different content, topics or sources can be considered 'sufficient' and, eventually, to operationalize that. Indeed, in the digital age has become less clear in which sense it is meaningful to speak of media diversity if the consumption is potentially characterized by limitless choice. There is an evident need for more detailed normative conceptualizations and empirical analysis. For example, standardized measures of core "sub-dimensions" of diversity (i.e. supply, content, exposure etc.) are highly needed. Also, different methodological approaches and measurement instruments need to be combined to better understand actual media consumption. As such, the issue of diversity cannot be 'solved' objectively or definitively, rather, throughout more interdisciplinary experimentations and collaborations it could be possible to approximate a social media content filtering and consumption that is, ultimately, 'sufficiently' diverse.

Much of this research has focused on the value of serendipity as a design principle. Even though we showed how serendipity is a polysemantic notion and polymorphic phenomenon – and thus we attempted to preliminary frame a taxonomy for its design in personalized news feeds – it underlies technical, metaphorical and educational values. Not only it is a natural by-product of digital environments but it can be intentionally cultivated. Essentially, researching and designing for digital serendipity can help to fruitfully understand how to improve users' openness to media diversity by design, to expand their information horizons and progressively make them discover new

information. Similarly, it helps to design and assess 'perspective widening' tools and techniques to balance individual media diet without resulting in "backfire effects". Designing for digital serendipity also acknowledges the need for some "bounded chaos" in our everyday life, as well as "reliable surprises" in media consumption. Ultimately, cultivating serendipity is individually desirable and socially beneficial. It could help to burst filter bubbles and weaken echo chambers by halting the risks of an overly personalized Internet experience. In theory, its design can sustain media diversity, help individuals to better realize and manage personalization practices and strengthen human rights, such as freedom of expression and the right to receive information.

Similarly to diversity, serendipity works well in theory while in practice shows many challenges and limitations. One of the main challenge is certainly its measurement. Serendipity is not a purely discrete concept. Rather, it swings in a spectrum and it is a really subjective experience. Yet, this might still be complemented with other measurements of diversity. Also, the beneficial effects of an architecture for serendipity remain uncertain; looking for serendipity may be time-consuming, risk-taking or distracting. However, this can be equally true for consonant and highly personalized information. Anyhow, being exposed to challenging information (what we termed 'political serendipity') may not necessarily result in experience of diversity, de-polarization and/or more tolerance. In many cases, the opposite may be true. Therefore, while from a theoretical perspective serendipity can be advocated as a desirable and ethical design principle, from an empirical perspective it is rather problematic. Not only there are still improvements in the understanding of serendipity from a theoretical perspective but there is also significant design, psychological and educational research gap.

To complete the answer on the first research question, to understand personalization of social media content and its challenges it is necessary, above all, a shared language and basic knowledge of these highly complex socio-technical systems. This is often not the case. There are in fact common misconceptions in the public, policy and even academic debate that should be acknowledged. Let me now briefly summarize some fundamental premises to discuss personalization more fruitfully.

Personalization debates should not focus exclusively on algorithms and recommender systems but take into account all the stakeholders, the business model, shareholders,

users, online newspapers, the broader media landscape essentially and, particularly, all the design process. In fact, not only design choices and affordances directly influence personalization, such as information findability and discoverability, but they ultimately affect public discourse more indirectly but substantially. Moreover, algorithms, and in particular "AI", have become a shorthand for an ever-changing suite of techniques, especially in policy discussions,. This reductionism need to be fully acknowledged. Policy makers should also resist simplistic narratives that blame for personalization risks and challenges exclusively with algorithms. Instead, policy makers should recognize the role of users and communities. Any solution to a problem is necessarily multifaceted and the responsibility is, to various extent, shared among several actors.

Similarly, the fundamental nature of recommender algorithms need to be better understood in public debates. Contrary to what many users still believe, there is no such thing as a "neutral" recommendation algorithm. Every recommender system embeds certain values, thus politics. Their content selections are dependent on a wide range of factors, including commercial and, in some cases, political considerations. Conversely, policy makers should avoid simplistic mandates for "neutrality" or "non-discrimination." Indeed, recommender systems are not mere code but complex and dynamic socio-technical systems; they are designed and deployed by specific people in specific contexts and for specific purposes. At the same time, they are highly context-dependent and often depend on changing parameters: dataset (what the space of options is), values, (what counts as a good recommendation) and assessment (how performance can be evaluated).

In the same vein, other fundamental metaphors and notions in the context of online personalization such as filter bubbles, echo chambers and diversity still need to be better understood and further discussed. They are in fact often ill-defined, sometimes overlapping and also have different meanings for different disciplines. The major risk, again, is to banalize, confuse and reduce the complexity of the challenges faced. Societies need to come together in communities comprised of multi-disciplinary researchers, innovators, policy-makers, citizens, developers and designers to foster the development of common knowledge and understanding. In this sense, it worth to repeat that researchers need to access more data. There is, in fact, a crisis with regard to the study of algorithms. They are indeed "black-boxed" which means their functioning is opaque and their interpretability may not even be clear to engineers.

The second major research question was about regulation, specifically how social media's personalization is regulated and whether the European legislation is able to prevent the risks arising from personalization. These issues require very complex answers that will be now summarized.

Personalization algorithms of mainstream social media are rather novel and complex entities and, therefore, they clearly require novel solutions and public discussions around their development and regulation. Regulators are faced with two major interrelated questions in this context: who should regulate platforms and how they should be regulated. Relatedly, a significant issue is which jurisdiction and law are more suited for regulation and, also, whether there is a need for modification of existing legal frameworks or to create new legislation. So, let me introduce these issues.

Platforms are already self-regulating entities. They determine the terms and conditions of their intermediary function and define online and offline standards of behaviour. Clearly, this paradigm is radically disputed in terms of protecting privacy and properly inform users. There is overwhelming evidence that most users neither they (can) read nor they understand online privacy policies. According to behavioral sciences as well, existing notice and consent model cannot be relied upon to protect the right to privacy. More generally, the extensive uncertainty and context dependence imply that people cannot be counted on to navigate the complex trade-offs involving terms of services and privacy self-management. So, it is this contractual paradigm in the first place that allows platforms to base their business model to engage users with personalization.

And yet, in the past years European policymakers have undertaken several different initiatives to regulate social media, especially their recommender systems. On the one hand, this has included horizontal instruments, such as competition and data protection law (E-Commerce Directive, E-privacy Directive, GDPR), which are not tailored to social media's personalization in particular but may still have some spillover benefits for its purposes.

On the other hand, as a more apt legal instrument for personalization there is media law. Nevertheless, as platforms used to argue that they are technology companies rather than media companies, so far they avoided media regulation and editorial responsibilities. Also, considering the special position of media and the preference for self-regulation following from media freedom, for some time it has been considered unlikely that the

normative principles in the media context would have to directly translate into legal obligations. Such context has enabled social media mainstream platforms to actually privatize personalization governance.

Governments for a long time have been indeed reluctant to take a strong lead. In the last years, the EU has adopted some provisions that give consumers the power to manage their personal data and not to be subject to automated decision-making as in personalization practices. These regulatory interventions – notably e-Privacy and GDPR – bring to the fore a reshaping of the traditional landscape of the consumer protection rules providing a more comprehensive vision of "data consumer law". They are expected to re-balance the relationship between data subjects and data controllers and to encourage competition between companies. Overall, the European legislation provides several safeguards but it still has blind-spots and a limited authority over personalization systems. Also, it represents, however, an excessively individualistic paradigm of social media governance, in which platforms are required to provide an enabling environment in which individual responsibility and autonomy can be realized in relation to the production, dissemination, and consumption of news and information.

The individualistic approach of user empowerment is insufficient. As we have already highlighted, there are three fundamental 'paradoxes' that highlight the complexities to autonomously manage personalization; the 'personalization paradox', that is, a trade-off between privacy and personalization accuracy; the 'privacy paradox', that is, the infamous users' inconsistent will to protect their own privacy; and the 'paradox of choice', that is, the more choices users have, the more easily they rely on simple-to-use personalized tools. In other words, in order for personalization systems to provide a 'better service', users are surveilled and datafied. Even if they would disagree – as they more often do – they usually do not proactively react. Even in cases where the users are provided with more agency, they are unlikely to take advantage. Therefore, another approach that takes more seriously into account individuals' cognitive limitations and information asymmetries is advocated.

Nonetheless, the common heritage of European fundamental rights and values can serve as an anchor point for regulatory discussions. In the context of personalization, the European Court of Human Rights provides several meaningful values such as human dignity, freedom of thought, privacy and, particularly, the right to receive information

(Article 10 ECHR) that proves an important point of departure to realize democratic values in the personalized media landscape. It is indeed an under-theorized right, lacking a framework to fully understand the rights of news consumers or the obligations of states regarding news recipients. There is a need to further discuss what the right to receive information should mean nowadays, how it relates to data protection, and to empirically study how people's information seeking strategies and privacy attitudes influence the exercise of this right. Information consumption is indeed deeply changed and needs to be reconceived. Given the vast amount of information produced and consumed, to some extent users are inevitably passive actors which have to delegate information filtering to algorithms and, therefore, to platforms. Regulation thus needs to catch up with newly social habits.

Finally, it was asked whether social media's personalization need to be further regulated and, if yes, how. The answer is affirmative. Yet, there are several nuances and subtleties that clearly deserve further discussion.

No policy intervention is without externalities. Above all, regulating recommender systems raise free speech risks comparable to outright removal of content and should be treated with comparable safeguards. Imposing any kind of information prioritization may become a matter of censorship. Content regulation is particularly problematic due to the lack of transparency and the potential for unintended consequences for online discourse. Therefore, any information block must occur in only exceptional circumstances laid down by international law on human rights. The full guarantee of freedom of expression must be the norm in the deployment of personalized recommender systems, next to the right to privacy and the prohibition of discrimination. Yet, the right to express does not mean a right to reach an audience. In this respect, it is important to eventually reach a consensus over which information (e.g. misinformation and disinformation) does not deserve a wide reach, as it was actually the case in the last decade. It is indeed necessary to keep to a minimum of regulation.

Self-regulation did not prove to be a reliable and effective approach. This is normally non-binding and the rules can be interpreted rather freely by social media companies. More and more critics in academia and civil society are arguing that such forms of regulation do not provide sufficient incentives. In any case, both the research community

and regulators need to collaborate with industry to test the long-term impact of personalization.

The question of regulating personalization is part of a complex regulatory system involving many different legislative acts. The regulatory space has also increased dramatically due to the borderless nature of the digital world, as has the technical expertise needed to create effective, appropriate and enforceable rules. Government authorities are often capacity-constrained, particularly with regard to the technical expertise required, for example, to perform complex algorithmic auditing. Thus, instead of self-regulation and regulation, it is co-regulation of social media's platforms the likely most effective approach to the governance of personalization systems and practices. In the European context, this would lead to a situation in which institutions define legislative standards that are subsequently implemented by platforms. There is the need to move away from homogenous top-down model towards a decentralized, reflexive, truly effective and collaborative framework that is 'polycentric'. New governance is predicated upon a dispersal and fragmentation of authority, and rests upon dynamic systems of power sharing. Social media policy is moving towards a cooperative responsibility, meaning to collectively define the essential public values at play and to develop a multi-stakeholder process of public deliberation and exchange and, eventually, to translate the outcome of public deliberation and agreements into regulations, codes of conduct, terms of use and technologies. In this respect, we also argued how the European Public Service Media might acquire a more proactive and innovative role.

The role of ethics has been widely discussed too. Despite its inevitable instrumentalization by technological companies, ethics can shape personalization regulation and governance through the relation of moral evaluation. On the one hand, hard ethics makes or shapes the law as it concerns the discussion of values, rights, duties and responsibilities in the course of formulating new regulations or challenging existing ones. In this respect, relatively few attempts to frame ethical personalization have been done so far. On the other hand, soft ethics concerns what ought and ought not to be done over and above the existing regulation. As such, there is a need for the development of professional algorithmic ethics. Also, a unified commitment to ethical standards in the form of a code of conduct between the stakeholders involved in personalization systems need to be established, safeguarding reasonable and responsible use of data access. In

particular, since academic research for the public good in this field depends to a large extent on data from the industry, there is a need to find a way to exchange data and keep the use of data conforming with legal regulations, especially GDPR.

In any case, enabling the dual advantage of 'ethical personalization'— so that the opportunities are seized, while the harms are foreseen and minimized or prevented — requires asking difficult questions about design, development, deployment, practices, uses and users, as well as the data that fuel the whole process. The hope is that ethical principles can act as normative constraints. Here it lies the distinction between the 'what' (ethical principles) and the 'how' (technical requirements). Complexity, variability, subjectivity, and lack of standardization, including variable interpretation of the 'components' of each of the ethical principles make this endeavor challenging. To make a relevant example, the principle of transparency, while a laudable goal in many respects, is too vague and runs up against insurmountable structural limitations within the political economy of the social media industry. This policy approach is subsumed by a discourse of consumer empowerment that has been rendered meaningless in the contemporary environment of pervasive commercial surveillance. Comprehensive transparency is effectively impossible to implement because privacy asymmetry is a cornerstone of the mainstream social media business model. Transparency might become a mere shield to deflect the threat of government regulation.

European Union, in particular, can become a global leader in developing and using AI – including recommender systems – for good and promoting a human-centric approach and ethics-by-design principles. Several guidelines, reports and drafts have indeed been discussed in the last years. Many of them attempted to discuss the fundamental principles that need to be applied in this context. In particular, we discussed how the principle of explicability complements the other four major identified principles, namely beneficence, non-maleficence, autonomy and justice. Explicability means enabling the other principles through intelligibility and accountability. It is in fact that principle that should help to explain algorithmic logics. Especially in the context of personalization it is fundamental. More research to realize such ethical principles – especially with an intercultural information ethics perspective – is thus highly needed.

A fundamental issue that personalization raises is the ethics of nudging and the risk of manipulation. In fact, individual and mass behavioral change and manipulation are

becoming ever more easier thanks to new techniques and types of data such as affective computing which help to deliver highly personalized messages. More banally, these techniques have been already deployed in the 2016 US elections (notably micro-targeting). It is therefore paramount to popularize 'persuasive knowledge' and to avoid addictive patterns of design and the pollution of the public debate, especially under elections. Among the proposed solutions, platforms should routinely disclose to their users and the public any experiment that the users were subjected to. Independent boards could review and approve experiments in advance. And, as for political microtargeting, platforms should clearly identify political advertisements and who paid for them and offer a transparent repository of all the material used in the campaign provided by candidates and political parties.

In this respect, we also introduced newly discussed human rights such as the 'right to not be measured, analyzed or coached', a 'right not to be deceived' or a 'right to cognitive sovereignty' that, more generally, ought to protect individuals' right to mental self-determination. Yet, as the Italian political philosopher Norberto Bobbio argued in an essay concerning the present and the future of human rights (1988), the serious problem of our times is not to create human rights but to protect them and to prevent, despite the solemn declarations, that they are continuously violated. More simply, regulatory frameworks need to consider the way in which the different more traditional human rights interlink.

Still, personalization algorithms, particularly of mainstream social media, have a collective value that deserve further discussion. In light of all the consequences of personalization of social media content, it is fundamental to recognize that these algorithms are common goods. Personalization algorithms shape individual identities, public opinion and, eventually, political action. Thus, politics in the digital age need to consider social media users also as political subjects. To make a substantial example, unions should be extended to users. The exploitation of digital labour is in fact connected to the commodification of the commons, including personalization algorithms.

Many of the "public interest considerations", however, remain relatively undertheorized and underdeveloped, in terms of providing guidance that can be operationalized and evaluated in algorithmic systems. So far, regulatory standards seem immature, and implementing them would be both technically infeasible and politically

controversial. Particularly with such vague and subjective rules, government action, as said above, also raises questions about freedom of expression and the rule of law. This is why we introduced and preliminary framed the notion of personalization algorithmic sovereignty, as the moral right of a person to be the exclusive controller of one's own algorithmic life and, more generally, the right and capacity by citizens as well as democratic institutions to make self-determined choices on personalization systems.

Similarly to other related concepts that have been introduced and discussed in academia and media such as data sovereignty, digital sovereignty and technological sovereignty, this notion has the ultimate goal to assert some form of individual and collective control empowerment over digital technologies. As such we analyzed experiences and experiments and framed some fundamental technical and political preconditions, such as education and media diet awareness, neutrality and transparency, design requirements and information discovery. In fact, these could renew trust between platforms and users, which is a fundamental element for social media to blossom. We should indeed be aware of the value that these platforms already have for societies; this raise other questions such as: what is the value of mainstream social media platform? Are global social media beneficial or even necessary? What if the fragmentation of social media would translate in an ever more fragmented public sphere?

This research certainly faced several limitations. To begin with, it is a qualitative research with no empirical data. As such, it is a research that is grounded on various theories and concepts and attempts a discursive and critical analysis of the subject of personalization. Clearly, pros and cons of an interdisciplinary research need to be considered too. Furthermore, no specific social media platform has been analyzed. Facebook received more focus but no methodological approach was applied. After all, how it is possible to comprehensively grasp, let alone effectively make regulatory recommendations on, a rather new subject which is not well-defined, that is employed by different companies, that is not accountable and cannot be properly researched, and that, at the same time, is global and culturally-dependent? Afterwards, we realize how the investigation of personalization requires an inter- and even trans-disciplinary approach, where computer scientists are prepared to discuss social and legal theories, social scientists show an interest in technical and legal regulatory mechanisms, and legal scholars engage with technical and social mechanisms. It is plenty of technologists in

civil society that are providing innovative solutions to solve many of the problems current mainstream social media are facing. Also, there have been several academic and European projects that aim to realize what we defined as personalization algorithmic sovereignty. There is indeed hope for substantial improvements.

Certainly, more than new regulation media reforms are needed. Social media oligopolies should be tackled more effectively; workable data portability and interoperability are preconditional to a healthier social media landscape; to sustain at least a fair competition, taxing large corporations and channelling the resulting income into alternative non-commercial social media could be fruitful too; in any case, the foundational mechanism of surveillance capitalism need to be counteracted; a collective discussion over alternative data ownership paradigms may be necessary; eventually, the ad-based model business ought to be re-discussed too; new institutions to deal with the challenges that online personalization entails are ultimately necessary. During the above analysis some proposals have been introduced, notably an agency for auditing RSs algorithms, an independent review board to evaluate potential experiments through RSs, and a 'public disclosure' institution providing transparent public records of RSs. In other words, experimental solutions and, eventually, updated regulations and new institutions may be needed.

Still, attempting to foresee the future social media landscape is merely a naïve exercise. Consider some of the changes occurred from the beginning of this research; many social and technical changes have occurred and new social media entered the arena, made famous predominantly by younger generations, such as *Snapchat* and *Tiktok*. The latter is, at the time of writing, used the same average amount of time of Youtube. In the United States it also became the favored app to share videos documenting the 'black lives matter' movement. This reminds us that technologies are not pre-determined and people make often unexpected use of these. Yet, it is also important to consider that the dominant positions of certain platforms are still untouched so far and that equally naïve is the idea that they will be, one day, necessarily undermined by an invisible hand.

To conclude, personalization systems are technologies that inscribe values, thus visions, ideas, and beliefs to satisfy individual's needs and desires, but also to facilitate social relations and to ultimately govern the public debate as information gatekeepers. Their role is fundamental in information societies. In the last chapter of "The Origins of

Totalitarianism", Hannah Arendt argues that loneliness is the defining condition of totalitarianism and the common ground of all terror. Loneliness is the inability to act all together, either with others or alone. This, in turn, is linked to the loss of what she calls "common sense" – the shared reality that allows us to know ourselves, to know where we end and the world begins, and how we are connected to others. Personalization algorithms can actually isolate individuals in filter bubbles and groups in echo chambers by limiting the breadth of information while threatening our common sense. It is now time to reimagine and redefine social media and its personalization systems to preserve human dignity, social trust and, ultimately, democratic institutions.

Bibliography

Abbott, A. (2008). The traditional future: A computational theory of library research. *College & Research Libraries*, 69(6), 524–545.

Abdollahpouri, H., Burke, R., & Mobasher, B. (2017, July). Recommender systems as multistakeholder environments. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization* (pp. 347-348).

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 509-514.

Albanie, S., Shakespeare, H., & Gunter, T. (2017). Unknowable manipulators: Social network curator algorithms. *arXiv preprint arXiv:1701.04895.*

Algo:aware (2018). Raising awareness on algorithms. Procured by the European Commission's Directorate-General for Communications Networks, Content and Technology.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society*, 20(3), 973-989.

Arendt, H. (1973). *The origins of totalitarianism* (Vol. 244). Houghton Mifflin Harcourt.

Bakshy, E., Messing, S., Adamic, L.A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science* 348(6239), 1130–1132

Bannerman, S. (2019). Relational privacy and the networked governance of the self. *Information, Communication & Society*, 22(14), 2187-2202.

Baudrillard, J. (1994). *Simulacra and simulation*. University of Michigan press.

Bauman, Z., & Lyon, D. (2013). *Liquid surveillance: A conversation*. John Wiley & Sons.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

Bennett, J. (2018). *Public service algorithms*. In: Freedman, D., Goblet, V. (eds.) A Future for Public Service TV, pp. 112–120. University of Goldsmith Press, London.

Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Science vs conspiracy: Collective narratives in the age of misinformation. *PloS one*, 10(2), e0118093.

Björneborn, L. (2017). Three key affordances for serendipity: Toward a framework connecting environmental and personal factors inserendipitous encounters. *Journal of Documentation*, 73(5),1053–1081.

Bloustein, E. J. (1978). *Individual and group privacy*. New Brunswick, N.J.: Transaction Publishers.

Bobbio, Norberto. (1988). *Presente e avvenire dei diritti dell'uomo*.

Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, 65(4), 619–638.

Bodó, B., Helberger, N., Eskens, S., & Möller, J. (2019). Interested in Diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization. *Digital Journalism*, 7(2), 206-229.

Bogers, T., & Björneborn, L. (2013). Micro-serendipity: Meaningful coincidences in everyday life shared on Twitter. *iConference*, 2013, 196–208.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, 13(1), 210-230.

Bozdag, E., & Timmermans, E. (2011). Values in the filter bubble Ethics of Personalization Algorithms in Cloud Computing. In *Proceedings, 1st International Workshop on Values in Design–Building Bridges between RE, HCI and Ethics*, Lisbon.

Bozdag, E., & van den Hoven, J. (2015). Breaking the filter bubble: Democracy and design. *Ethics and Information Technology*,17(4), 249–265.

Bozdag, E., & van de Poel, I. (2013). Designing for diversity in online news recommenders. In 2013 *Proceedings of PICMET'13: Technology Management in the IT-Driven Services (PICMET)* (pp. 1101-1106). IEEE.

Bruns, A. (2019). *It's not the technology, stupid: How the 'Echo Chamber' and 'Filter Bubble' metaphors have failed us.* Paper presented at the IAMCR, Madrid, Spain.

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication &Society*, 20(1), 30–44.

Büchi, M., Fosch-Villaronga, E., Lutz, C., Tamò-Larrieux, A., Velidi, S., & Viljoen, S. (2019). The chilling effects of algorithmic profiling: Mapping the issues. *Computer Law & Security Review*, 105367.

Burrell, J. (2016). How the machine 'thinks': understanding opacity in machine learning algorithms. *Big Data Soc*. 3(1), 1–12.

Burri, M. (2015). *Public Service Broadcasting 3.0: Legal Design for the Digital Present*. Routledge, London.

Burri M. (2019). Discoverability of Local, National and Regional Content Online: Mapping Access barriers and Contemplating New Orientation Tools, discussion paper.

Bygrave, Lee A. (2001). The place of privacy in data protection law. *UNSWLJ*, 24, 277.

Campanario, J. M. (1996). Using citation classics to study the incidenceof serendipity in scientific discovery. *Scieontometrics*, 37(1), 3–24.

Campos, J., & Figueiredo, A. D. (2002). Programming for serendipity. In *Proceedings of the AAAI fall symposium on chance discovery—The discovery and management of chance events*.

Capurro, R. (2005). Privacy. An intercultural perspective. *Ethics and Information Technology*, 7(1), 37-47.

Carr, N. (2016). *Utopia is creepy: And other provocations*. New York:W W Norton & Co Inc.

Carr, P. L. (2015). Serendipity in the stacks: Libraries, informationarchitecture, and the problems of accidental discovery. *College & Research Libraries*, 76, 831–842.

Carson, A. B. (2015). Public discourse in the age of personalization: Psychological explanations and political implications of search engine bias and the filter bubble. *Journal of Science Policy & Governance*, 7(1).

Castells, M. (2011). *The rise of the network society* (Vol. 12). John wiley & sons.

Cobbe, J., & Singh, J. (2019). Regulating Recommending: Motivations, Considerations, and Principles. Considerations, and Principles (April 15, 2019).

Colangelo, G., Maggiolino, M. (2019). From fragile to smart consumers: shifting paradigm for the digital era. *Comput. Law Secur. Rev*. 35(2), 173–181.

Darbellay, F., Moody, Z., Sedooka, A., & Steffen, G. (2014). Interdisciplinary research boosted by serendipity. *Creativity Research Journal*, 26 (1), 1-10.

Davenport, T.H., Beck, J.C. (2001). *The Attention Economy: Understanding the New Currency of Business.* Harvard Business Press, Boston.

Delacroix, S. (2018). Taking turing by surprise? Designing digitalcomputers for morally-loaded contexts. *arXiv preprintarXiv:1803.04548.*

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Echo chambers: Emotional contagion and group polarization on facebook. *Scientific reports*, 6, 37825.

De Filippi, P., & McCarthy, S. (2012). Cloud computing: Centralization and data sovereignty. *European Journal of Law and Technology*, 3(2).

Delacroix, S., & Veale, M. (2019). Smart Technologies and Our Sense of Self: Going Beyond Epistemic Counter-Profiling. *Law and Life in the Era of Data-Driven Agency, O'Hara & Hildebrandt (eds.).*

Deleuze, G. (1992). Postscript on the Societies of Control. *October*, 59, 3-7.

Deibert, R.J. (2019). The road to digital unfreedom: three painful truths about social media. *J. Democr*. 30(1), 25–39.

de Melo, R. M. C. (2018). *On serendipity in the digital medium: Towards a framework for valuable unpredictability in interaction Design*. Ph.D. Thesis, Universidade do Porto, Porto, Portugal, 2018.

DeNardis, L., & Hackl, A. M. (2015). Internet governance by social media platforms. *Telecommunications Policy*, 39(9), 761-770.

De Rond, M. (2014). The structure of serendipity. *Culture and Organization*, 20(5), 342–358.

DeVito, M. A. (2017). From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism*, 5(6), 753–773.

DIEM25. (2019). Progressive Agenda for Europe Technological Sovereignty: *Democratising Technology and Innovation Green Paper No. 3 2019.*

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

Dylko, I., Dolgov, I., Hoffman, W., Eckhart, N., Molina, M., & Aaziz, O. (2018). Impact of customizability technology on political polarization. *Journal of Information Technology & Politics*, 15(1), 19–33.

Edward Foster, A., & Ellis, D. (2014). Serendipity and its study. *Journal of Documentation*, 70(6), 1015–1038.

Edwards, L., Veale, M. (2017). Slave to the algorithm: why a right to an explanation is probably not the remedy you are looking for. *Duke L. Tech. Rev*. 16, 18.

Erdelez, S. (1997). Information encountering: a conceptual frameworkfor accidental information discovery. In *Proceedings of an international conference on Information seeking in context*(pp. 412–421). Taylor Graham Publishing, London.

Erdelez, S. (2004). Investigation of information encountering in thecontrolled research environment. *Information Processing & Management*, 40(6), 1013–1025.

Erdelez, S., & Jahnke, I. (2018). Personalized systems and illusion of serendipity: A sociotechnical lens. In *Workshop of WEPIR 2018*.

Esau, K., Friess, D., & Eilders, C. (2017). Design matters! An empirical analysis of online deliberation on different news platforms. *Policy & Internet*, 9(3), 321-342.

Eskens, S., Helberger, N., & Moeller, J. (2017). Challenged by news personalisation: five perspectives on the right to receive information. *Journal of Media Law*, 9(2), 259-284.

Fabris, A. (2018). *Etica per le tecnologie dell'informazione e della comunicazione*. Carocci Editore.

Festinger, L. (1957). *A theory of cognitive dissonance* (Vol. 2). Stanford university press.

Fletcher, R., Nielsen, R.K. (2017). Are news audiences increasingly fragmented? a cross-national comparative analysis of cross-platform news audience fragmentation and duplication. *J. Commun*.

Finck, M. (2018). Digital co-regulation: designing a supranational legal framework for the platform economy. *European Law Review*.

Fletcher, R., Nielsen, R.K. (2018). Are people incidentally exposed to news on social media? A comparative analysis. *New Media Soc*. 20(7), 2450–2468.

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality.* OUP Oxford, Oxford

Floridi, L. (2011). The informational nature of personal identity. *Minds and Machines*, 21(4), 549.

Floridi, L. (2015a). The politics of uncertainty. *Philosophy & Technology*, 28(1), 1–4.

Floridi, L. (2015b). *The onlife manifesto*. Cham: Springer.

Floridi, L. (2016a). Mature information societies—A matter of expectations. , 29(1), 1–4.

Floridi, L. (2016b). Tolerant paternalism: Pro-ethical design as a resolutionof the dilemma of toleration. *Science and Engineering Ethics*, 22(6), 1669–1688.

Floridi, L. (2018a). Soft ethics: its application to the general data protection regulation and its dual advantage. *Philosophy & Technology*, 31(2), 163-167.

Floridi, L. (2018b). Soft ethics and the governance of the digital. *Philosophy & Technology*, 31(1), 1-8.

Fogg, B. J., Lee, E., & Marshall, J. (2002). *Interactive technology and persuasion. The Handbook of Persuasion: Theory and Practice.* Thousand Oaks: Sage.

Francis, P. (2015). Laudato si: On care for our common home. *Our Sunday Visitor*.

Fuchs, C. (2013). *Internet and surveillance: The challenges of Web 2.0 and social media* (Vol. 16). Routledge.

Fuchs, C., & Marisol, S. (2015). The political economy of capitalist and alternative social media. *The Routledge Companion to Alternative and Community Media*.

Fuchs, C. (2017). Towards the Public Service Internet as Alternative to the Commercial Internet. *ORF Texte*, 20, 43-50.

Fuster, G. G. (2014). The emergence of personal data protection as a fundamental right of the EU (Vol. 16). *Springer Science & Business*.

Gabielkov, M., Ramachandran, A., Chaintreau, A., & Legout, A. (2016). Social clicks: What and who gets read on Twitter?. In *Proceedings of the 2016 ACM SIGMETRICS international conference on measurement and modeling of computer science* (pp. 179-192).

Gal, M. S. (2017). Algorithmic challenges to autonomous choice. *Michigan Telecommunications and Technology Law Review*, 2017.

Galimberti, U. (2010). *L'ospite inquietante (Vol. 1).* Feltrinelli Editore.

Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010). Beyond accuracy: Evaluating recommender systems by coverage and serendipity.In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 257–260). ACM, New York.

Gibson, J. J. (2014). *The ecological approach to visual perception: Classic edition.* Hove: Psychology Press.

Goodman, B., Flaxman, S. (2016). European union regulations on algorithmic decision-making and a "right to explanation". *arXiv preprint arXiv:1606.08813*.

Gottfried, J., & Shearer, E. (2016). News Use Across Social Medial Platforms 2016. *Pew Research Center*.

Granovetter, M. S. (1977). The strength of weak ties. In *Social networks* (pp. 347-367). Academic Press.

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (patterns) side of UX design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 534). ACM.

Greene, T., & Shmueli, G. (2019). How Personal is Machine Learning Personalization?. *arXiv preprint arXiv:1912.07938.*

Gollub, T., Potthast, M., Stein, B. (2018). Shaping the information nutrition label. In: *Proceedings of the NewsIR 2018 Workshop at ECIR*.

Gup, T. (1997). Technology and the end of serendipity. *The Chronicle of Higher Education*, 44(21), A52.

Gulyás, Á., Hammer, F. (eds.). (2014). *Public Service Media in the Digital Age: International Perspectives*. Cambridge Scholars Publishing, Cambridge.

Habermas, J. (1996). *Between facts and norms: contributions to a discourse theory of law and democracy*. Polity Press.

Habermas, J., Lennox, S., & Lennox, F. (1974). The public sphere: An encyclopedia article (1964). *New German Critique*, (3), 49-55.

Han, B. C. (2016). *Psicopolitica*. Nottempo, Roma.

Hakala, J., Rytkönen, E., Snellman, K., & Särkkä, N. (2014). Ethical Personalisation Act 2025.

Hausman, D. M., & Welch, B. (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy*, 18(1), 123-136.

Harambam, J., Helberger, N., & van Hoboken, J. (2018). Democratizing algorithmic news recommenders: How to materialize voicein a technologically saturated media ecosystem. *Philosophical Transactions A*, 376(2133), 20180088.

Hasselbalch, G. (2019). Making sense of data ethics. The powers behind the data ethics debate in European policymaking. *Internet Policy Review*, 8(2), 1-19.

Hayes, B. (2012). *The surveillance-industrial complex*. Routledge handbook of surveillance studies, 167-175.

Helberger, N. (2011). Diversity by design. *Journal of Information Policy*, 1, 441-469.

Helberger, N. (2015). Public service media| merely facilitating or actively stimulating diverse media choices? public service media at the crossroad. *International Journal of Communication*, 9, 17.

Helberger, N., Kleinen-von Königslöw, K., & van der Noll, R. (2015). Regulating the new information intermediaries as gatekeepers of information diversity. *info*, 17(6), 50-71.

Helberger, N., Borgesius, F.Z., Reyna, A. (2017). The perfect match? A closer look at the relationship between EU consumer law and data protection law. *Common Market Law Rev.* 54(5), 1427–1465

Helberger, N., Karppinen, K., & D'acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, 21(2), 191-207.

Helberger, N., Pierson, J., Poell, T. (2018b). Governing online platforms: from contested to cooperative responsibility. *Inf. Soc*. 34(1), 1–14.

Helberger, N., Eskens, S. J., van Drunen, M. Z., Bastian, M. B., & Möller, J. E. (2019). Implications of AI-driven tools in the media for freedom of expression.

Hildebrandt, M. (2009). *Profiling and AmI. In The future of identity in the information society* (pp. 273–310). Berlin: Springer.

Hildebrandt, M. (2019). Privacy as protection of the incomputable self: From agnostic to agonistic machine learning. *Theoretical Inquiries in Law*, 20(1), 83-121.

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.

Hirsch, D. D. (2010). The law and policy of online privacy: Regulation, self-regulation, or co-regulation. *Seattle UL Rev.*, 34, 439.

Hoffmann, C. P., Lutz, C., Meckel, M., & Ranzini, G. (2015). Diversity by choice: Applying a social cognitive perspective to the role of public service media in the digital age. *International Journal of Communication*, 9(1), 1360–1381.

Hoven, J. van den, Miller, S., & Pogge, T. (Eds.). (2017). *Designing in ethics*. Cambridge: Cambridge University Press.

Howell L. (2013). Digital wildfires in a hyperconnected world. In *WEF Report 2013*. World Economic Forum, 2013.

Jenkins, H. (2007). *Cultura convergente*. Maggioli Editore.

Johansen, A. H. (2019). *Moving fast and breaking things? Facebook's Lobbying of the European Union's General Data Protection Directive* (Master's thesis).

Ippolita, (2011). *Luci e Ombre di Google, Futuro e Passato dell'Industria dei Metadata*. Feltrinelli Editore. Milano.

Ippolita. (2016). *Anime elettriche*, Jaca book, Milano.

Kaltheuner, F., & Bietti, E. (2018). Data is power: Towards additional guidance on profiling and automated decision-making in the GDPR. *Journal of Information Rights, Policy and Practice*, 2(2).

Karppinen, K. (2008). Media and the paradoxes of pluralism. *The Media and Social Theory*, 27–42.

Keymolen, E. (2016). Trust on the line: a philosophycal exploration of trust in the networked era.

Kidron, B., Evans, A., Afia, J., Adler, J. R., Bowden-Jones, H., Hackett, L., & Scot, Y. (2018). *Disrupted childhood: The cost of persuasive design.* 5Rights.

Kohl, U., Davey, J., & Eisler, J. (2019). Data-driven personalisation and the law-a primer: collective interests engaged by personalisation in markets, politics and law.

Kop, R. (2012). The unexpected connection: Serendipity and human mediation in networked learning. *Journal of Educational Technology & Society*, 15(2), 2–11.

Kotkov, D., Wang, S., & Veijalainen, J. (2016). A survey of serendipity in recommender systems. *Knowledge-Based Systems*,111, 180–192.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790.

Kroes, P., & van de Poel, I. (2015). *Design for values and the definition, specification, and operationalization of values.* Handbook of Ethics, Values, and Technological Design: Sources, Theory,Values and Application Domains, 151–178.

Kuhn, T. S. (1962). The structure of scientific revolutions. *The Un. of Chicago Press*, 2, 90.

Lanzing, M. (2018). "Strongly recommended" revisiting decisional privacy to judge hypernudging in self-tracking technologies. *Philos. Technol*. 32, 549–568.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Schudson, M. (2018). The science of fake news. *Science*, 359(6380), 1094-1096.

Lepri, B., Staiano, J., Sangokoya, D., Letouzé, E., & Oliver, N. (2017). The tyranny of data? the bright and dark sides of data-driven decision-making for social good. In *Transparent data mining for big and small data* (pp. 3-24). Springer, Cham.

Lynch, M. P. (2016). *The Internet of us: Knowing more and understanding less in the age of big data*. New York: WW Norton & Company.

Löcherbach, F., & Trilling, D. (2018). 3bij3: A framework for testing effects of recommender systems on news exposure. In *2018 IEEE 14th International Conference on e-Science (e-Science),* (pp. 350-351). IEEE.

Lupo, L. (2012). *Filosofia della serendipity* (Vol. 73). Napoli: Guida Editori

Lynch, M. P. (2016). *The Internet of us: Knowing more and understanding less in the age of big data*. New York: WW Norton& Company.

Lynskey, O. (2019). Grappling with "data power": normative nudges from data protection and privacy. *Theor. Inq. Law* 20(1), 189–220.

Lovink, G., & Rasch, M. (2013). Unlike us reader: Social media monopolies and their alternatives. *INC reader*, (8).

Makri, S. (2014). Serendipity is not Bullshit. Paper presented at the EuroHCIR 2014, The 4th European Symposium on Human-Computer Interaction and Information Retrieval, 13 Sep 2014, London, UK.

Makri, S., & Blandford, A. (2012). Coming across information serendipitously—Part 1, p. A process model. *Journal of Documentation*, 68(5), 684–705.

Makri, S., Blandford, A., Woods, M., Sharples, S., & Maxwell, D. (2014). "Making my own luck": Serendipity strategies and howto support them in digital information environments. *Journal of the Association for Information Science and Technology*, 65(11), 2179–2194.

Maloney, A., & Conrad, L. Y. (2016). *Expecting the unexpected: Serendipity, discovery, and the scholarly research process* (whitepaper), Thousand Oaks: SAGE.

Marcus, G. E. (2010). *Sentimental citizen: Emotion in democratic politics*. University Park: Penn State Press.

Marx, Karl. (1842). Freedom in General. Retrieved from:
http://www.marxists.org/archive/marx/works/1842/freepress/ch06.htm

Marwick, A. E., & Boyd, D. (2014). Networked privacy: How teenagers negotiate context in social media. *New media & society*, 16(7), 1051-1067.

Matt, C., Benlian, A., Hess, T., & Weiß, C. (2014). Escaping from the Filter Bubble? The Effects of Novelty and Serendipity on Users' Evaluations of Online Recommendations. In *Proceedings of the 35th International Conference on Information Systems (ICIS2014)*, Auckland, New Zealand.

McCay-Peet, L., & Toms, E. G. (2013). Proposed facets of a serendipitous digital environment. In *Teoksessa iConference 2013 Proceedings*, ss. 688–691.

McCay-Peet, L., & Toms, E. G. (2017). *Researching serendipity indigital information environments*. Synthesis Lectures on Information Concepts, Retrieval, and Services, 9(6), i–i91.

Meckel, M. (2011). "Sos—save our serendipity", Personal Blog. https://www.miriammeck el.de/2011/10/11/sos-save-our-serendipity/.

Merton, R. K., & Barber, E. (2006). *The travels and adventures ofserendipity: A study in sociological semantics and the sociologyof science*. Princeton: Princeton University Press.

Metcalf, J., & Moss, E. (2019). Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics. *Social Research: An International Quarterly*, 86(2), 449-476.

Mills, C. Wright. (1956). *The Power Elite*. Oxford: Oxford University Press.

Mittelstadt, B. (2016). Automation, algorithms, and politics| auditing for transparency in content personalization systems. *Int. J. Commun*. 10, 12.

Moeller, J., & Helberger, N. (2018). Beyond the filter bubble: Concepts, myths, evidence and issues for future debates.

Moen, G. M., Ravna, A. K., & Myrstad, F. (2018). *Deceived by design*. Forbrukerrådet.

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), 959-977.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2019). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*, 1-28.

Nagulendra, S., & Vassileva, J. (2016). Providing awareness, explanation and control of personalized filtering in a social networking site. *Information Systems Frontiers*, 18(1), 145–158.

Napoli, P.M. (2015). Social media and the public interest: governance of news platforms in the realm of individual and algorithmic gatekeepers. *Telecommun. Policy*, 39(9), 751–760.

Napoli, P., & Caplan, R. (2017). Why media companies insist they're not media companies, why they're wrong, and why it matters. *First Monday*.

Nenadić, I. (2019). Unpacking the "European approach" to tackling challenges of disinformation and political manipulation. *Internet policy review*, 8(4), 1-22.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A., & Nielsen, R. K. (2016). Digital news report 2016. *Reuters Institute for the Study of Journalism*.

Nikolov, D., Oliveira, D. F., Flammini, A., & Menczer, F. (2015). Measuring online social bubbles. *PeerJ Computer Science*, 1, e38.

Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus*, 140(4), 32-48.

Noelle-Neumann, E. (1984). *The spiral of silence. A theory of public opinion*, Chicago: University of Chicago Press .

O'Hara, K., Stevens, D. (2015). Echo chambers and online radicalism: assessing the internet's complicity in violent extremism. *Policy Internet* 7(4), 401–422.

O'Connor, B. (1988). Fostering creativity: Enhancing the browsing environment. *International Journal of Information* Management, 8(3), 203–210.

Olma, S. (2016). *In Defence of Serendipity*. Watkins Media Limited, 2016.

Ostrom, E. (2015). *Governing the commons*. Cambridge university press.

Papacharissi, Z. (2008). The virtual sphere 2.0: The Internet, the public sphere, and beyond. In *Routledge handbook of Internet politics* (pp. 246-261). Routledge.

Paraschakis, D. (2017). Towards an ethical recommendation framework. In *2017 11th International Conference on Research Challenges in Information Science (RCIS)* (pp. 211-220). IEEE.

Pariser, E. (2011). *The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think.* Penguin, Westminster.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard: Harvard University Press.

Pentina, I., Tarafdar, M. (2014). From "information" to "knowing": exploring the role of social media in contemporary news consumption. *Comput. Hum. Behav*. 35, 211–223.

Pentland, A. (2015). *Social physics: How social networks can make us smarter*. New York: Penguin.

Peirce, C. S. (1992). *The essential Peirce: Selected philosophical writings (Vol. 2).* Indiana: Indiana University Press.

Picard, G., Pickard, V. (2017). *Essential Principles for Contemporary Media and Communications Policymaking*. Reuters Institute for the Study of Journalism. University of Oxford.

Powers, E. (2017). My news feed is filtered? Awareness of news personalization among college students. *Digital Journalism*, 5(10), 1315–1335.

Purtova, N. (2018). The law of everything. Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology*, 10(1), 40-81.a

Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echo chambers on Facebook. Available at *SSRN: https ://ssrn.com/abstr act=27951 10.*

Race, T., & Makri, S. (2016). *Accidental information discovery: Cultivating serendipity in the digital age*. Cambridge: ChandosPublishing.

Rannenberg, K., Royer, D., & Deuker, A. (Eds.). (2009). *The future of identity in the information society: Challenges and opportunities.* Springer Science & Business Media.

Rauch, J. (2018). *Slow Media: Why Slow is Satisfying, Sustainable, and Smart*. Oxford University Press, Oxford.

Reviglio, U. (2017). Serendipity by design? How to turn from diversity exposure to diversity experience to face filter bubbles in social media. In International Conference on Internet Science (pp. 281-300). Springer, Cham.

Reviglio, U. (2019). Serendipity as an emerging design principle of the infosphere: challenges and opportunities. *Ethics and Information Technology*, 21(2), 151-166.

Reviglio, U. (2019b). Towards a Right not to Be Deceived? An Interdisciplinary Analysis of Media Personalization in the Light of the GDPR. In *Conference on e-Business, e-Services and e-Society* (pp. 47-59). Springer, Cham.

Reviglio, U. (2019c). Improving User Experience by Browser Extensions: A New Role of Public Service Media?. In *International Conference on Internet Science* (pp. 257-271). Springer, Cham.

Reviglio U. (2019d). Towards a Taxonomy to Design Serendipity in Personalized News Feeds. In *Information Research - An International Electronic Journal*.

Reviglio, U., & Agosti, C. (2020). Thinking Outside the Black-Box: The Case for "Algorithmic Sovereignty" in *Social Media. Social Media+ Society*, 6(2), 2056305120915613.

Reviglio U., Alunge R. (2020). "I Am Datafied Because We Are Datafied": An Ubuntu Perspective on (Relational) Privacy. In *Philosophy & Technology*. Springer.

Rheingold, H. (2012). *Net smart: How to thrive online.* Mit Press.

Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: introduction and challenges. In *Recommender systems handbook* (pp. 1-34). Springer, Boston, MA.

Rice, J. (1988). Serendipity and holism: The beauty of OPACs. *Library Journal*, 113(3), 138-41.

Roio, D. (2018). Algorithmic Sovereignty. Doctoral dissertation, University of Plymouth.

Rubin, V. L., Burkell, J., & Quan-Haase, A. (2011). Facets of serendipity in everyday chance encounters: A grounded theory approachto blog analysis. *Information Research*, 16(3), 27

Semaan, B.C., Robertson, S.P., Douglas, S., Maruyama, M. (2014). Social media supporting political deliberation across multiple public spheres: towards depolarization. In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1409–1421. ACM. 21.

Schmidt, E. (2006). How we're doing and where we're going. Google Inc. *Press Day* 2006.

Scholz, T. (Ed.). (2012). *Digital labor: The Internet as playground and factory*. Routledge.

Schönbach, K. (2007). 'The own in the foreign': Reliable surpriseanimportant function of the media? *Media, Culture & Society*,29(2), 344–353.

Sen, A. (1990). Justice: Means versus freedoms. *Philosophy & Public Affairs*, 19(2), 111–121.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.

Shearer, E., & Gottfried, J. (2017). News use across social media platforms 2017. *Pew Research Center*, 7(9), 2017.

Simon H. (1971). Designing organizations for an information-rich world, In: *Greenberger M, editor. Computers, Communications, and the Public Interest*. Baltimore, MD: The Johns Hopkins Press. p. 37–72.

Skinner, B. F. (1972). *Beyond Freedom and Dignity*. New York: Bantam.

Smith, A. (2018). Many Facebook users don't understand how the site's news feed works. *Pew Research Center*, Journalism and Media.

Solove, D. J. (2002). Conceptualizing privacy. *Calif. L. Rev.*, *90*, 1087.

Solove, D. J. (2007). I've got nothing to hide and other misunderstandings of privacy. *San Diego L. Rev.*, 44, 745.

Splichal, S. (2007). Does history matter? Grasping the idea of public service at its roots. In: Lowe, G.F., Bardoel, J. (eds.) From Public Service Broadcasting to Public Service Media. *RIPE@2007*, pp. 237–256. Nordicom, Gothenburg.

Stiegler, B. (2014). *Il Chiaroscuro della Rete*. Youcanprint.

Stiegler, B. (2017). The new conflict of the faculties and functions:Quasi-causality and serendipity in the anthropocene. *Qui Parle*, 26(1), 79–99.

Sunstein, C. R. (2017b). Default rules are better than active choosing(Often). *Trends in Cognitive Sciences*, 21(8), 600–606.

Sunstein, C.R. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press, Princeton. 4.

Suzor, N. (2018). 'Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms', *Social Media + Society*.

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder (Vol. 3).* New York: Random House.

Tang, T. Y., & Winoto, P. (2016). I should not recommend it to you even if you will like it: the ethics of recommender systems. *New Review of Hypermedia and Multimedia*, 22(1-2), 111-138.

Thaler, R.H., Sunstein, C.R. (2009). Nudge: *Improving Decisions About Health, Wealth, and Happiness*. Penguin, New York.

Thurman, N. (2011). Making 'The Daily Me': Technology, economics and habit in the mainstream assimilation of personalized news. *Journalism*, 12(4), 395–415

Thurman, N., & Schifferes, S. (2012). The future of personalization at news websites: Lessons from a longitudinal study. *Journalism Studies*, 13(5-6), 775-790.

Thurman, N. (2019). Personalization of news. *The International Encyclopedia of Journalism Studies*, 1-6.

Thurman, N., & Schifferes, S. (2012). The paradox of personalization: The social and reflexive turn of adaptive news. *The handbook of global online journalism*, 373-391.

Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital Journalism*, 1-23.

Tucker, J., et al. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature.

Turel, O., He, Q., Xue, G., Xiao, L., & Bechara, A. (2014). Examination of neural systems sub-serving Facebook "addiction". *Psychological reports*, 115(3), 675-695.

Turkle, S. (2017). *Alone together: Why we expect more from technology and less from each other*. Hachette UK.

Turing, A. M. (1950). 'Computing machinery and intelligence'. *Mind*, 59, 433–460.

Turilli, M., Floridi, L. (2009). The ethics of information transparency. *Ethics Inf. Technol*. 11(2), 105–112.

Tutt, A. (2016). An FDA for algorithms. *Social Science Research Network*. 22.

van Andel, P. (1994). Anatomy of the unsought finding. serendipity:Origin, history, domains, traditions, appearances, patterns and programmability. *The British Journal for the Philosophy of Science*, 45(2), 631–648.

Van Cuilenburg, J., McQuail, D. (2003). Media policy paradigm shifts: towards a new communications policy paradigm. *Eur. J. Commun*. 18(2), 181–207.

Van den Bulck, H., Moe, H. (2017). Public service media, universality and personalisation through algorithms: mapping strategies and exploring dilemmas. *Media Cult. Soc*. 40(6), 875–892.

van den Hoven, J., & Rooksby, E. (2008). *Distributive justice and the value of information: A (broadly) Rawlsian approach. Information Technology and Moral Philosophy*, p. 376. Cambridge: Cambridge University Press.

van Es, K.F. (2017). An Impending Crisis of Imagination: Data-Driven Personalization in Public Service Broadcasters. *Media@ LSE Working Paper Series* (43), pp. 1–18.

van Est, Q. C., Gerritsen, J., & Kool, L. (2017). *Human rights in the robot age: challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality*, The Hague, The Rathenau Institute.

Vercellone, C., Bria, F., Fumagalli, A., Gentilucci, E., Giuliani, A., Griziotti, G., & Vattimo, P. (2015). Managing the commons in the knowledge economy. *European Union's Seventh Framework Programme*, Brussels, Belgium.

Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review*, 11(1), 274-302.

Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Quattrociocchi, W. (2016). The spreading of misinformation online. *Proc. Natl. Acad. Sci*. 113(3), 554–559.

Yadamsuren, B., & Erdelez, S. (2016). Incidental exposure to online news. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 8(5), i–i73.

Yeung, K. (2017). 'Hypernudge': big data as a mode of regulation by design. *Inf. Commun. Soc*. 20 (1), 118–136.

Youyu, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 1036–1040.

Wachter, S., Mittelstadt, B., Floridi, L. (2017). Why a right to explanation of automated decisionmaking does not exist in the general data protection regulation. *Int. Data Priv. Law* 7(2), 76– 99.

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report, 27.

Winner, L. (1980). Do artifacts have politics?. *Daedalus*, 121-136.

Zarsky, T.Z. (2019). Privacy and manipulation in the digital age. *Theor. Inq. Law* 20(1), 157–188.

Zipf, G. K. (1949). Human behavior and the principle of least effort. Addison-Wisley. Reading, MA.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.

Zuckerman, E. (2008). Serendipity, echo chambers, and the front page. *Nieman Reports,* 62 (4), 16.

Zuckerman, E. (2013). *Rewire: Digital cosmopolitans in the age of connection*. New York: W. W. Norton & Company.

Zuiderveen Borgesius, F. J. (2015). Personal data processing for behavioural targeting: which legal basis?. *International Data Privacy Law*, 5(3), 163-176.

Zuiderveen Borgesius, F., Trilling, D., Möller, J., Bodó, B., De Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles?. Internet Policy Review. *Journal on Internet Regulation*, 5(1).

Legislation, Policy Documents and Expert Advice:

Council of Europe (2018), Recommendation CM/Rec(2018)1 of the Committee of Ministers to member States on media pluralism and transparency of media ownership: Guidelines on media pluralism and transparency of media ownership, Article 2.5.

Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes, 2019, Decl(13/02/2019)1 (declaration on manipulative capabilities);

Draft Recommendation of the Committee of Ministers to member States on human rights impacts of algorithmic systems, 2018, MSI-AUT(2018)06 (draft recommendation on algorithmic systems).

EBU Digital Strategy Group: Media with a Purpose: Public Service Broadcasting in the Digital Age. EBU, Geneva (2002)

EBU: Big Data Initiative: Time to Invest. Le Grand-Saconnex (2017)

European Commission. Directorate-General for Communication Networks, Content and Technology. (2018). A multi-dimensional approach to disinformation: Report of the independent high level group on fake news and online disinformation. Publications office of the European Union.

European Commission. (2018). EU Code of Practice on Disinformation.

Guidelines to Convention 108 on Artificial Intelligence and data protection, 2019, T-PD(2019)01 (AI guidelines to Convention 108);

Guidelines to Convention 108 on the protection of individuals with regard to the processing of personal data in a world of Big Data, 2017, T-PD(2017)01(Big Data guidelines to Convention 108);

Harris, T. (2019). U.S. Senate. Committee on Commerce, Science & Transportation. Hearing before the Subcommittee on Communications, Technology, Innovation, and the Internet on: Optimizing for Engagement: Understanding the Use of Persuasive Technology on Internet Platforms. Retrieved from https://www.commerce.senate.gov/public/_cache/files/96e3a739-dc8d-45f1-87d7-ec70a368371d/35DE175BA77F95FB2EFC5E94018ABF1A.06-25-19harris-testimony.pdf

Recommendation of the Committee of Ministers to member States on Big Data for culture, literacy and democracy, 2017, CM/Rec(2017)8 (recommendation on cultural big data).

Recommendation of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries, 2018, CM/Rec(2018)2

Newspapers Articles:

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16(7), 16-07. https://www.wired.com/2008/06/pb-theory/

Derakhshan, H. (2016). "Social Media Is Killing Discourse Because It's Too Much Like TV," *MIT Technology Review*, Nov 29. https://www.technologyreview.com/s/602981/social-media-is-killing-discourse-because-its-too-much-like-tv/.

Krotoski, A. (2011). "Digital serendipity: Be careful what you don't wish for", *The Guardian*, Aug 21. from https://www.theguardian.com/technology /2011/aug/21/google-serendipity-profiling-aleks-krotoski.

Ha, A. (2014). "Edward Snowden's Privacy Tips: "Get Rid Of Dropbox," Avoid Facebook And Google", June 15, 2014. https://techcrunch.com/2014/10/11/edward-snowden-new-yorker-festival/

Oksman, O. (2016). "Conspiracy craze: why 12 million Americans believe alien lizards rule us", The Guardian, April 07. https://www.theguardian.com/lifeandstyle/2016/apr/07/conspiracy-theory-paranoia-aliens-illuminati-beyonce-vaccines-cliven-bundy-jfk

Thi Nguyen, C. (2018). "Escape the echo chamber". Aeon, April 09. https://aeon.co/essays/why-its-as-hard-to-escape-an-echo-chamber-as-it-is-to-flee-a-cult

Nicoletti,G. (2015). Umberto Eco: "Con i social parola a legioni di imbecilli", *La Stampa*, June 11. https://www.lastampa.it/cultura/2015/06/11/news/umberto-eco-con-i-social-parola-a-legioni-di-imbecilli-1.35250428

Morozov, E. (2017). "Moral panic over fake news hides the real enemy – the digital giants", *The Guardian*, Jan 08. https://www.theguardian.com/commentisfree/2017/jan/08/blaming-fake-news-not-the-answer-democracy-crisis

Sunstein, C. R. (2017c). "In praise of serendipity", *The Economist*, Mar 09. https://www.economist.com/news/books-and-arts/21718464-social-media-should-encourage-chance-encounters-not-customised-experiences-praise.

List of Figures

Figure 1. The Pyramid of Knowledge.

Figure 2. Basic User profile construction for personalization (from Bozdag and van de Poel, 2013).

Figure 3. A General Taxonomy of the Main Algorithms Employed in Online Recommendations (from Lepri et al., 2017).

Figure 4. From humanistic concepts of the person to personalized scores: ML operationalization requires narrowing a person embedded in social and cultural space to a feature vector embedded in feature space (from Greene and Shmueli, 2019).

Figure 5. Semantic network analysis of the hashtags utilized on Instagram before 2016 UK referendum.

Figure 6. Political books relations for Amazon's recommender system (from The Economist, 2017).

Figure 7. Models of Democracy and Design Criteria (from Bozdag and van den Hoven, 2015).

Figure 8. Newsfeed management (screenshot, May 2017).

Figure 9. Proportion Who are Concerned About Each Potential Consequence of Personalization (Selected Countries) (from Newman et al., 2016).

Figure 10. Proportion Who Think That Each Selection Method is a Good Way to Get the News (from Newman et al., 2016).

Figure 11. The Percentage of Use of the Word 'Serendipity' in books (Google Books Ngram, 2019).

Figure 12. Cybernetic Serendipity Art Exhibition (1968).

Figure 13. Key affordances and sub-affordances for serendipity with coupled personal factors and sub-factors (from Björneborn, 2017).

Figure 14. Summary of the main opposing phenomena and antithetical values that the design of serendipity tends to favour and disfavour (Reviglio, 2019a).

Figure 15. A dashboard that displays users their filter bubbles (from Nagulendra and Vassileva, 2016)

Figure 16. Summary of the Serendipity Categories' Features.

Figure 17. Summary of the Relationships among the Categories.

Figure 18. The relationship between digital ethics, digital regulation and digital governance (frm Floridi, 2018).

Figure 19. Youtube's most recommended videos on the website Algortransparency.org. (screenshot, May, 2019).

Figure 20. An ethical framework of the five overarching principles for AI (from Floridi and Cowls, 2019).

Figure 21. Scoopinion, a browser add-on that displays user's news consumption habits. Larger circles are news outlets that the user consumed the most items (screenshot, May, 2019).

Figure 22. Limited Post Loading in Social Fixer Affords (screenshot, May, 2019).

Figure 23. News Aggregator Gobo Made by MIT Social Media Lab (screenshot, May, 2019).