Volume 11 issue 3 December 2018

ELR

Erasmus Law Review

eleven
international publishing

Contents

Mission Statement
The *Erasmus Law Review* seeks to foster independent critical scholarship as relevant to the discipline of law.
The Board of Editors encourages the submission of legally relevant manuscripts by legal scholars and practitioners as well as those versed in other disciplines relevant to law, such as criminology, sociology, political science and economics.
The *Erasmus Law Review* usually commissions articles around specific themes, although 'calls for papers' on specific topics might be issued occasionally and will be published on the Review's website. All prospective articles are submitted to double-blind peer review (two reviews per article), and final publication is dependent on the outcome of these reviews.

ISSN: 2210-2671

Nederlands uitgeversverbond
Groep uitgevers voor vak en wetenschap

The Conduit between Technological Change and Regulation

Marta Katarzyna Kolacz & Alberto Quintavalla*

Abstract

This article discusses how the law has approached disparate socio-technological innovations over the centuries. Precisely, the primary concern of this paper is to investigate the timing of regulatory intervention. To do so, the article makes a selection of particular innovations connected with money, windmills and data storage devices, and analyses them from a historical perspective. The individual insights from the selected innovations should yield a more systematic view on regulation and technological innovations. The result is that technological changes may be less momentous, from a regulatory standpoint, than social changes.

1 Introduction

The capacity of regulation to respond to the legal issues presented by new technologies is not an unknown topic. While socio-technological innovations tend to open new possibilities once introduced, they might also challenge pre-existing regulatory paradigms. Throughout history, questions concerning the design of optimal regulation have repeatedly emerged in reaction to a radical transformation in society, which may be due to multiple factors such as morality and technology. The discussion on *whether* and *how* the law[1] shall reflect these changes dates back over 2,000 years.

This introductory article to the present special issue of *Erasmus Law Review*[2] intends to discuss *how* the law has approached disparate socio-technological innovations over the centuries. The primary concern is to investigate the timing of regulatory responses. By doing so, we enter in the realm of regulation and technology, thus setting the conceptual framework for the other articles in the issue.

Legislatures and courts usually require a certain amount of time to handle the various challenges brought about by technology. This time period is necessary to acquire any relevant information about the legal issues posed by the new innovations.[3] The length of time needed for this operation should depend on the risks and complexity of innovation. Yet, it seems that other factors are deemed more influential: it is commonly argued that the law responded in the past more slowly than it does at present. The printing press may serve as an illustrative example. It was invented in Europe around 1439. It allowed printed books to move across borders and started the era of mass communication. But despite its disruptive potential, it took a long time before responses to legal issues began to emerge.[4] This was partly due to the slow pace of distribution and the difficulty of monetising the product.

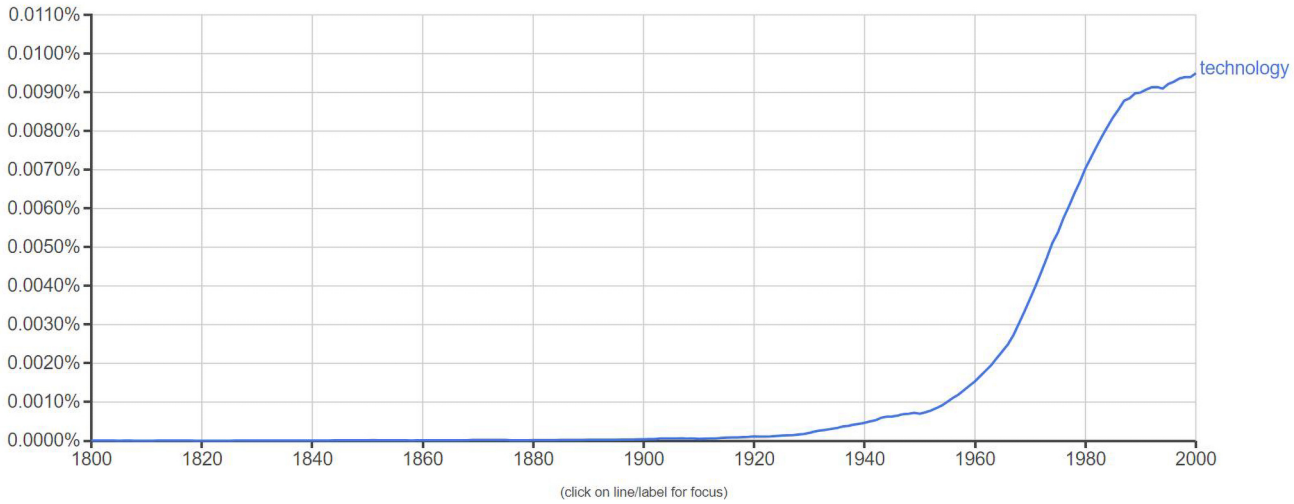
In the twenty-first century, however, innovation and technological changes move at a much more rapid pace. Significant and impactful advances are secured almost daily as a consequence of digitalisation. In today's globalised world, innovations appear to follow each other not only in quick succession, but also on a larger scale than ever before. For example, WhatsApp killed the SMS revenues of the telecom sector within a single quarter. SMS itself had been a novel technology only a couple of years before its demise. Similarly, technologies such as blockchain, currently still in their infancy, are widely expected to disrupt long-established markets. Globalisation and digitalisation, in combination with technology, have created a new socio-technological context. The emergence of new technologies often launches

* Marta Katarzyna Kołacz, Ph.D. Candidate in the Department of Private Law, Erasmus School of Law, Erasmus University Rotterdam, The Netherlands. Alberto Quintavalla, Ph.D. Candidate in the Rotterdam Institute of Law and Economics, Erasmus School of Law, Erasmus University Rotterdam, The Netherlands. The authors would like to thank Orlin Yalnazov and the two anonymous reviewers for their valuable comments as well as Luuk Hoogenboom for his excellent research assistance. The usual disclaimer applies.

1. With the term 'law', we refer to the law that exists at a specific point in time and that is implemented as such by courts or enforcement bodies.
2. The current special issue builds on the Erasmus Early-Career Scholars Conference, which was held from 11 April 2018 to 13 April 2018 at the Erasmus University, Rotterdam. Financial support received from the Erasmus Initiative 'Dynamics of Inclusive Prosperity', the Erasmus Trustfonds and the Erasmus Graduate School of Law is gratefully acknowledged.
3. The article looks at the regulatory responses to technological change after the fact. The *ex post* view therefore coincides with the stages of innovation and diffusion. Academic research usually divides technological innovation into three stages: invention, innovation and diffusion. *See e.g.* J.A. Schumpeter, *Theory of Economic Development* (1934). Yet, legal scholars tend to perceive the stage of diffusion as different from technological innovation *per se*. *See* on this point *e.g.* N.A. Ashford, C. Ayres & R.F. Stone, 'Using Regulation to Change the Market for Innovation', 9(2) *Harvard Environmental Law Review* 419 (1985). *See infra* Section 2.
4. One may think of the late establishment of copyright laws. With the exception of a (crude form of) copyright legislation by the Venetian State in Renaissance Italy, we had to wait until the enactment of the Statute of Anne in XVIII century. *See* B. W. Bugbee, *Genesis of American Patent and Copyright Law* (New York: Public Affairs Press) (1967), at 43-38; A. B. Birrell, *Seven Lectures on the Law and History of Copyright in Books* (London: Cassell) (1899), at 51-54; B. Kaplan, *An Unhurried View of Copyright* (New York: Columbia University Press) (1967).

doi: 10.5553/ELR.000112 - ELR december 2018 | No. 3

Figure 1 *Google Ngram Viewer for the word 'technology'*



discussions on emerging legal (and moral) issues. For instance, the creation of new tools for Internet users, such as social networks, has brought out digitally expressed 'hate speech' and 'fake news' as well as many other collateral problems. Legislatures and courts are therefore called to tackle the legal issues at stake in a quick and orderly fashion. Besides, the cross-border aspect of current technological changes may exacerbate this problem. This has also been the case for Google and its search support when a U.S. District Court annulled (within the U.S. soil) a Canadian court's judgment that had directed the tech giant to stop displaying certain references to pirated products.[5]

Against this background, our contribution attempts to answer if law has approached socio-technological changes in a uniform manner. Put slightly differently, we consider if different types of socio-technological change may entail a different rate of regulatory intervention once the technology starts yielding negative externalities. To do so, the article circles on particular innovations connected with money, windmills and data storage devices and analyses them from a historical perspective. The individual insights from the selected innovations should yield a more systematic view on regulation and technological innovations.

The article is structured as follows. Section 2 advances the backbone proposition and theoretical approach of this article. We present technological changes as a part of social change, which has some distinctive problems. Section 3 analyses the development of regulations on particular technologies from a historical perspective. While Section 4 develops some concluding observations in the context of regulating technologies, the last part (Section 5) provides an account of the (other) articles making up this special issue.

2 Technology as a Facilitator of Social Change

The term 'technology' has elicited a great deal of interest among scholars from disparate disciplines, such as philosophy, sociology and law. 'Technology', which is still a fuzzy concept,[6] emerged not so long ago. Although the word entered the English language in the seventeenth century, its use became frequent and regular only in the early decades of the twentieth century.[7] As proof to this, Figure 1 graphically displays how use of the word 'technology' only increased significantly in the 1930s. Similarly, the same word started appearing regularly in the EU parliamentary debates only in the last five years.[8]

The enmeshment of technology and law is thus quite recent.[9] Debates in the sphere are commonly framed as some variant of the question 'how to regulate technolo-

5. *See e.g. Google* v. *Equustek Solutions Inc.*, 2017 SCC 34, [2017] 1 S.C.R. 824. In such a legal challenge, the technology company Equustek Solutions filed a lawsuit against Datalink Technology Gateways. The accusation was to sell products belonging to Equustek, thus misappropriating *inter alia* its trade secrets. Google was therefore required to remove Datalink websites from its search results, both in Canada and globally. Yet, the U.S. District Court for Northern California granted Google an injunction to prevent enforcement of the Canadian Court order in the U.S. territory. For the U.S. Direct Court's judgment *see Google LLC* v. *Equustek Solutions Inc.*, 2017 WL 5000834 (N.D. Cal. Nov. 2, 2017).

6. There is no single definition of technology. For the purpose of this article and in line with previous literature, we employ the definition provided by Schon: 'any tool or technique, any product of process, any physical equipment or method of doing or making, by which human capability is extended'. *See* D. Schon, *Technology and Change: The New Heraclitus* (New York: Delacorte Press) (1967), at 1.

7. E. Schatzberg, 'Technik Comes to America. Changing Meanings of Technology before 1930', 47(3) *Technology and Culture* 486 (2006); L. Marx, 'Technology. The Emergence of a Hazardous Concept', 51(3) *Technology and Culture* 561 (2010).

8. The information is retrieved from www.europarl.europa.eu/plenary/en/minutes.html#sidesForm (last visited 28 January 2019). The search option allows looking for any word in the minutes of each plenary sitting of the European Parliament.

9. One may note that before 'technology' gained popularity, these discussions were couched in different terms – *e.g.* manufacturing, useful arts and invention. *See* Schatzberg, above n. 7.

gy?'.[10] Technology is regarded, in other words, as a rationale for regulation. As soon as a technological innovation takes place, it is expected that regulators should intervene to regulate it. Such a view may, however, fail to fully capture the meaning of technological change.

Technological changes enable people to broaden their usual field of action and, as such, may have different consequences for law and the organisation of society.[11] For example, the invention of e-mail and the Internet offered the opportunity to communicate with other individuals over long distances and in real time. Yet, for regulation to be necessary, the use of e-mails must raise legal issues – salient for certain individuals in the society, which cannot be solved by established legal frameworks.[12] From an *ex post* view, technology becomes a rationale for regulation only once it involves a societal disturbance.[13]

While technology offers individuals enhanced technical capabilities, it cannot determine historical outcomes by itself. Taking that view on technology leads us to perceive technological change as "one type of social change".[14] Social change generally refers to the idea of a society moving forward by evolutionary means to secure people's interests in a multiplicity of forms.[15] Social change can be driven by a wide array of forces, including *inter alia* behavioural changes or shifts in cultural beliefs. The Industrial Revolution and the feminist movement both exemplify this tendency.

It follows that technological change has some features that are distinct from those of social change. Technological change influences the course of social development. However, taken in isolation, it is not a reason to change the law.[16] The social change brought about by technological developments might require a modification of the pre-existing legal framework.[17] Technological changes are therefore less momentous, from a regulatory standpoint, than are social changes. Nevertheless, it seems that legislatures are sometimes urged to intervene solely because of the occurrence of technological change. For example, autonomous vehicles are not yet widespread but there are several attempts to regulate them.[18] Legal scholars have pinpointed a couple of reasons behind this tendency. First, technological change may occur faster than social change. The variation in rates of technological and social changes may generate a sense of unfamiliarity with the new technology, ultimately putting greater pressure for legal intervention.[19] Secondly, the issues raised by technological changes are perceived as more objective – not tainted by any *a priori* ideological vision – and accordingly easier to regulate.[20]

However, the quest for a speedy regulatory response often results in disenchantment: it seems that law fails to keep pace with rapidly evolving technology.[21] This narrative puts the time factor in the spotlight. Law, the argument runs, should be more effective and responsive in handling the challenges posed by technological innovations in anticipation of social change. In order to delve into this inquiry, one may ask whether law has approached various types of socio-technological changes in the same manner over the years. The *ex post* view we adopt forces us to consider the time period that coincides with the stages of innovation and diffusion of the technological change, thus excluding the *ex-ante* fear-driven legislation.[22] If this analysis shows a heterogeneity in the rate of regulatory responses, it becomes necessary to identify what particular socio-technological changes should be addressed first. Based on these assumptions, the following section intends to examine the regulatory responses to selected innovations.

10. R. Brownsword and K. Yeung, *Regulating Technologies: Legal Futures Regulatory Frames and Technological Fixes* (Opole: Hart) (2008).
11. Please *see* the definition provided at n. 6.
12. Some legal issues that may arise vary from managing certain risks to protecting individuals' rights. Besides, we are not considering the 'regulatory capture' option.
13. W.E. Bijker and J. Law, *Shaping Technology/Building Society: Studies in Sociotechnical Change* (Cambridge: MIT Press) (1992), at 20-22. A societal disturbance can also result from the identification of potential undesirable consequences triggered by the technological change. That is to say, for a societal disturbance to exist, it is not necessary to have the actual occurrence of negative consequences.
14. L.B. Moses, 'Why Have a Theory of Law and Technological Change?', 8 *Minnesota Journal of Law, Science & Technology* 589, at 598 (2007).
15. C.F. Sabel and J. Zeitlin, 'Historical Alternatives to Mass Production: Politics, Markets and Technology in 19th Century Industrialization', 108(1) *Past and Present* 133 (1985).
16. *See* Moses, above n. 14.
17. L.B. Moses, 'Regulating in the Face of Sociotechnical Change', in R. Brownsword, E. Scotford & K. Yeung (eds.), *The Oxford Handbook of Law, Regulation, and Technology* (OUP 2017), 573.
18. For example on the civil liability for damages caused by robots, including autonomous cars, *see e.g.* Report with recommendations to the Commission on Civil Law Rules on Robotics, European Parliament Committee on Legal Affairs (2017), 6-8, 12, 16-18, available at: www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A8-2017-0005+0+DOC+PDF+V0//EN (last visited 28 January 2019); on the product liability regarding the vehicles and its software, *see e.g.* State of Michigan Bill Number SB 663 (2013).
19. *See* Moses, above n. 14, at 600.
20. L. Lessing, 'Understanding Changed Readings: Fidelity and Theory', 47 *Stanford Law Review* 395, at 400 (1995); M.E. Price and J.F. Duffy, 'Technological Change and Doctrinal Persistence: Telecommunications Reform in Congress and the Court', 97 *Columbia Law Review* 976, at 1008-1009 (1997); M.E. Price, 'The Newness of New Technology', 22 *Cardozo Law Review* 1885 (2001).
21. L.B. Moses, 'Recurring Dilemmas: the Law's Race to Keep Up with Technological Change', 8(2) *University of Illinois Journal of Law, Technology and Policy* 239, at 247 (2007); G.E. Marchant, 'The Growing Gap between Emerging Technologies and the Law', in G.E. Marchant, B.R. Allenby & J.R. Herkert (eds.), *The Growing Gap between Emerging Technologies and Legal-Ethical Oversight. The Pacing Problem* (Berlin: Springer) (2011) 19, at 20.
22. The fear-driven legislation develops particularly in the first stage of a technological change. *See* on this point *e.g.* P. Kleve, 'Technology Law: Symbolic Solutions to Problems, or Solutions to Symbolic Problems?', in P. Kleve and C. van Noortwijk (eds.), *Something Bigger Than Yourself – Essays in Honour of Richard De Mulder* (Rotterdam: Erasmus School of Law) (2011), at 131-5.

3 Historical Instances of Particular Innovations

3.1 Preliminary Remarks

There is a virtually endless list of historical instances when socio-technological change has prompted regulatory responses. Selecting representative responses is a tall order. To begin with, we do need particular innovations that have triggered regulatory responses by the legislatures and courts, both in the past and in the present. For the purpose of this article, we focus on money, windmills and data storage devices.

These rubrics were selected for several reasons. Technology is an integral part of all of them: windmills and data storage devices are technological innovations *per se*, whereas money has been significantly affected by developing technologies over time. Contemporary electronic money can even be considered a pure technology, in the same mould as windmills and data storage devices. In addition, the selection of the three subjects allows us to consider regulatory responses from a fairly wide spectrum of legal fields, ranging from private law to administrative law. Specifically, monetary technology triggered the development of commercial laws (as well as laws of financial systems). Windmills prompted changes in administrative and environmental law. Lastly, data storage devices touch upon civil and consumer law.[23]

We do not, however, aim to provide an exhaustive list of all the regulatory responses that have occurred within the three rubrics. Our examples instead show facets of legal responses resulting from or triggered by technological developments. Since law responds to socio-technological change in a way that impinges upon disparate interests, it is important to confine ourselves to a fairly limited set of regulatory patterns. The selected examples in which regulation approaches the legal issues posed by new technologies will serve as a point of reference for further research.

3.2 Money

Before proceeding to the analysis of its specific regulatory responses, a definition of money shall be provided. According to Ferguson, money is

> a medium of exchange which has the advantage of eliminating inefficiencies of barter, a unit of account, which facilitates valuation and calculation; and a store of value, which allows economic transactions to be concluded over long periods as well as geographical distances.[24]

The physical object that symbolises money has changed over the centuries. Coins circulated in the Ancient Mediterranean world.[25] However, coins cannot be considered the only precursors of today's money. While clay tokens were popular a long time ago in ancient Mesopotamia, banknotes have been in circulation since the seventh century.[26] The twentieth century triggered the development of an electronic store of monetary value, known as e-money. More recently still, cryptocurrencies such as Bitcoins entered the 'market'.[27] These developments triggered regulatory responses, and it is on those facets of regulation that we focus here.

Regulation has traditionally focused on remedying asymmetries of information that pertain to standards of value. Such an approach was common since the early medieval times. For instance, several penal laws from that time advert to compensation payments in precious metals for the commission of various felonies.[28] In addition, regulations about coins often included technical requirements – that is, type, shape and weight – as well as the methods of production. The 1580 Mint Ordinance of the Polish King Stefan Batory is a striking example. This authoritative decree specified all the necessary requirements for the production of coins, as well as the type, stamp, weight of metals, ranks of craftsmen and systems of contracting between the Crown and local mints.[29] Setting these technical requirements can be seen as one response to emerging legal problems, such as the unification of governance systems on the Polish and Lithuanian lands as well as tax payments. It also facilitated local and international trade because the standardised monetary value increased certainty in transactions and prevented the activities of profiteers working against the interests of local traders.

Similarly, the modern coinage system operates by designating specific objects as money. As a consequence, those objects acquire a specific value.[30] The designation process thus happens through regulation. Regulation identifies certain means of payment that can serve as money. Bringing as an example the current monetary system of Poland, the relevant regulation states that *zloty* and *grosz*, which are operating in coins and banknotes, are the currency signs of the Republic of Poland.[31] The National Bank of Poland, on the basis of further implementing provisions, issues banknotes and coins according to certain technical requirements. Similarly, Poland regulates electronic money, which can be

23. One may also note that some of the selected rubrics, and namely money and data storage devices, relate to the other articles in this issue. The importance of historical framing helps to observe and understand the techno-legal debates of the past and present, their developments and directions.
24. N. Ferguson, *The Ascent of Money. A Financial History of the World* (London: Penguin Press) (2009), at 24.
25. G. Davies, *History of Money: From Ancient Times to the Present Day* (Cardiff: University of Wales Press) (2002), at 74-78.
26. Ferguson, above n. 24, at 28.
27. Cryptocurrencies reflect an encrypted value, existing not as a paper money or coins but as strings of digital code. For more *see* A.M. Antonopoulos, *Mastering Bitcoin: Unlocking Digital Cryptocurrencies* (Newton: O'Reilly) (2004).
28. P. Grierson, *The Origins of Money* (London: Athlone Press) (1977), at 12-19.
29. S. Tymieniecki, *Zarysy do dziejów mennic koronnych Zygmunta III. W szesnastym wieku* (Drukarnia Czas) (1917), at 3-10.
30. L. Kurke, *Coins, Bodies, Games, and Gold. The Politics of Meaning in Archaic Greece* (Princeton: Princeton University Press) (1999), at 305.
31. Art. 31 Ustawa z dnia 29 sierpnia 1997 r. o Narodowym Banku Polskim Dz.U.2017.0.1373 t.j.

considered an equivalent of money.[32] In parallel to the traditional means of payment, cryptocurrencies began to grow in popularity in 2009. The major issues posed by cryptocurrencies are the effective functioning (and strength) of centralised currencies and the illegitimate activities surrounding decentralised currency, such as money laundering, terrorist financing and tax evasion.[33] States are therefore urged to address these raising legal issues, which became popular because of the widespread use of cryptocurrencies in the society. In line with this approach, the Polish government has, for example, adopted regulations whereby a virtual currency is deemed a digital representation of value and not legal tender.[34]

Closely linked to money as a means of exchange is the regulation of standards for securities preventing forgeries. From a historical perspective, these standards already developed in ancient Rome where, in the third century BCE, coins were produced with serrated edges. For the same purpose, China introduced a concession for brass and established state printing-houses using specific colour printing, rich designs and official stamps in the Middle Ages.[35] Similar solutions can be found in Europe: under the aforementioned Polish Mint Ordinance, coins were exclusively produced in contracted mints. Each mint had its assayer responsible for the supervision of a production process and the maintenance of the quality of coins. Assayers were personally liable for any failure to observe the regulations.[36]

In modern times, banknote security printing started when, in the end of the eighteenth century, the Netherlands began using in-house designs and engravings for music types.[37] This and further technological developments such as colour printing, 3D devices, watermarks and holograms have led regulators to introduce technological standards. In the Eurozone, the European Central Bank and National Central Banks are the only agencies responsible for the issue of currencies. More specifically, they are entitled to introduce new series of euro banknotes with standardised security features, thus benefiting from advances in banknote technology.[38] The intertwinement of fiat currencies, technology and law facilitates the safety of the transactions by providing a regulatory response to the developments occurred in technology.

It can be seen from this sketch that monetary regulation has grown hand in hand with technological *and* social changes, both in the past and in the present. Regulation – which can encompass both the development of new rules and the adaptation of the existing ones – comes into play as soon as new objects or means of payment are widely accepted into circulation. In this way, the law recognises the growing usage of different means of payment, institutionalising them by setting regulatory frames such as the ones on issuance and turnover. In that sense, regulation seems an outcome based on actual results (effective social change) rather than a simple forecast (the occurrence of a technological change).

3.3 Windmills

The harnessing of wind power is a technology that has started developing in eastern Persia thousands of years ago.[39] These primitive wind devices were then followed by (vertical) windmills on the Dutch and Mediterranean territories in the fourteenth century.[40] At that time, the primary function of these windmills was to pump water, mill grain, and drain land.[41] In the nineteenth century, the high rate of technological progress spurred the development of new turbines, a new type of windmills. Both windmills and wind turbines have been rather popular in certain areas of the globe. Nevertheless, their use caused environmental disturbances that have required a regulatory response. It seems therefore appropriate to look a little more closely at the development of the corresponding legal frameworks.

Windmills are inherently embedded in the landscape of the Netherlands.[42] Windmills equipped with water-lifting technology have been integrated in the Dutch drainage system since the fifteenth century.[43] In other words, windmills were one of the effective tools to combat against demanding environmental conditions and continuous threat of floods. Thus, although the construction of these drainage windmills was rather costly for an ordinary farmer,[44] windmills spread around all the areas affected by poor drainage.

32. Ustawa z dnia 19 sierpnia 2011 r. o usługach płatniczych Dz.U. 2011 Nr 199 poz. 1175 t.j.
33. For more *see* R. Houben and A. Snyers, 'Cryptocurrencies and Blockchain: Legal Context and Implications for Financial Crime, Money Laundering and Tax Evasion' – European Parliament Study Requested by the TAX3 committee (2018), available at: www.europarl.europa.eu/cmsdata/150761/TAX3%20Study%20on%20cryptocurrencies%20and%20blockchain.pdf (last visited 6 December 2018).
34. Art. 1(26) Ustawa z dnia 1 marca 2018 r. o przeciwdziałaniu praniu pieniędzy oraz finansowaniu terroryzmu Dz.U. 2018 poz. 723 t.j.
35. M. Zajęcki, 'Regulacje prawne dotyczące monet i pieniądza papierowego w dawnych Chinach', in P. Wilinski, O. Krajniak & B. Guzik (eds.), *Prawo wobec wyzwań współczesności. Tom IV* (Poznan) (2007), at 233.
36. Tymieniecki, above n. 29, at 52-54.
37. K.J. Schell, 'History of Document Security', in K. De Leeuw and J. Bergstra (eds.), *The History of Information Security* (London: Elsevier) (2007), at 203-4.
38. Decision of the European Central Bank of 19 April 2013 on the denominations, specifications, reproduction, exchange and withdrawal of euro banknotes (recast) (ECB/2013/10) (2013/211/EU) L 118/37.
39. R.W. Righter, *Wind Energy in America: A History* (Norman: University of Oklahoma Press) (1996), at 7. Here, we refer to horizontal windmills. For more about the construction of the first mill devices *see* R.L. Hills, *Power from Wind* (Cambridge: Cambridge University Press) (1994), at 11-17.
40. J.K. Kaldellis and D. Zafirakis, 'The Wind Energy (R)evolution: A Short Review of a Long History', 26 *Renewable Energy* 1887, at 1887 (2011).
41. For more *see* Hills, n. 39, at 115-236.
42. M. Reuss, 'Learning from the Dutch: Technology, Management, and Water Resources Development', 43(3) *Technology and Culture* 465, at 466 (2002).
43. A. Kaijser, 'System Building from Below: Institutional Change in Dutch Water Control Systems', 43(3) *Technology and Culture* 521, at 530 (2002). Besides, consider that a large part of the Dutch territory is potentially threatened by flooding.
44. J. de Vries, *The Dutch Rural Economy in the Golden Age, 1500-1700* (New Haven: Yale University Press) (1974), at 198. In order to facilitate the construction of windmills, a framework for financing, building and operating windmills was also devised. For more *see* Kaijser, above n. 43, at 536.

However, the large number of windmills built had a detrimental impact on the water balance at a regional level.[45] More specifically, windmills were lifting an excessive amount of water into the so-called *boezem*.[46] By doing so, there was an actual danger that the surrounding farmlands would be flooded. A 1444 decree of the water authority of Delfland exacerbated this problem because it stipulated that drainage windmills would operate whenever there was sufficient wind.[47] This situation led to legal disputes concerning the appropriate water level in the *boezem*.[48] The debate was particularly lively between 'highlanders' and 'lowlanders' due to their differing interests vis-à-vis the water levels.[49] As a result, the regional water authorities decided to intervene from both technical and legal standpoints. With regard to the former, they increased the capacity of sluices.[50] As to the latter, the regional water authorities started issuing windmill permits, thus assuming more power and responsibilities.[51] In line with this new approach, a 1562 decree of the Delfland water authority set a fixed water level in the *boezem*.[52]

The modern usage of windmills' descendants – wind turbines – is aimed at the production and supply of energy. Popularised in the nineteenth century, wind turbines have become a common mean of producing energy in the twentieth century.[53] Wind turbines, like drainage windmills, can influence the neighbourhood both positively and negatively. In fact, these turbines are not only an energy source but also a cause of potential disturbances.

The Netherlands launched a large-scale programme for the development of wind turbines in the 1970s.[54] Although these policies were also promoted to enhance renewable energy deployment, local planning for wind farms[55] revealed to be problematic.[56] Specifically, locals tend to view wind farms with hostility due to environmental concerns, especially noise. Accordingly, it appeared that certain (social) standards had to be set in order to ensure acceptable noise levels. Dutch authorities began adopting environmental regulations for wind turbines a few years after the inception of the programme. The most recent standards indicate that the noise caused by wind turbines should be restricted to a maximum of 47 dB Lden and 41 dB Lnight at any noise-sensitive location.[57]

In both the Middle Ages and the 1970s, regulatory responses were not contemporaneous to technological change. More precisely, regulation aimed at responding to environmental disturbances of windmills and wind turbines resulted only after the use of the technological development became widespread. The technology *per se* was insufficient to trigger regulatory intervention. As in the case of money, regulation emerged because of issues raised by the widespread use of windmills (social change) rather than by the creation itself (technological change).

3.4 Data Storage Devices

The collection and aggregation of information has always driven improvements in social welfare. Collecting data has yielded evidence of historical events, as well as the discovery of the origins of certain customs and practices. Data were originally passed on through storytelling, songs and dances, which were also testimonies of local culture and belief. As time passed, writing and storage technologies have vastly expanded our society's ability to store and disseminate information. This has been recognised to serve various state's and societal needs, especially in the era of digitisation. Nevertheless, the related risks have not escaped regulators' attention.

Collecting and storing data for public purposes has been common practice for centuries. Public registers, in particular, have been an integral part of state organisation.[58] Public registers served the political system. Cadasters were kept for taxation purposes. One of the oldest examples dates to ancient Rome. Registers with data

45. Kaijser, above n. 43, at 536.
46. The *boezem* is an area in which excess water can be stored before it is permanently discharged onto a river that brings the water to the sea.
47. Het hoogheemraadschap van Delfland, *Het oudste keurboekje*, at 55.
48. P.J.E.M. van Dam, *Vissen in veenmeren: De aalvisserij bij de sluizen tussen Haarlem en Amsterdam en de ecologische transformatie in Rijnland 1440-1530* (Hilversum: Uitgeverij Verloren) (1998), at 82-86.
49. C. Postma, *Het hoogheemradschap van Delfland in de middeleeuwen, 1289-1589* (Hilversum: Uitgeverij Verloren) (1989), at 372-5; D. van Doorn, *Gedenkschrift uitgegeven ter gelegenheid van het 700-jarige bestaan van het Hoogheemraadschap van Schieland* (Uitgever: De Boer-Cuperus) (1994), at 62-65.
50. For the readers who are not very familiar with hydraulic engineering and water management, a sluice is a passage for water usually controlled by a gate. For more about the improvements of sluices *see* P.J.E.M. van Dam, 'Ecological Challenges, Technological Innovations. The Modernization of Sluice Building in Holland, 1300-1600', 43(3) *Technology and Culture* 500 (2002).
51. Kaijser, above n. 43, at 538.
52. Postma, above n. 49, at 378-83. A similar approach can be found in other regions of Europe such as some territories currently belonging to Poland. That is because of the Mennonites who were prosecuted and forced to leave their home territories. In the sixteenth century, Mennonites settled in the region called Żuławy Wiślane – the delta area of Vistula River. There, they implemented irrigation systems including polders and windmills. Operation and maintenance of polders was within the competence of the so-called embankment unions. These were established to protect the region against floods. Their growing importance as far as flood protection was concerned resulted in several decrees institutionalising their operations. For example, the King of Prussia Wilhem II issued a decree giving a statue to the Embankment Union of Vistula and Nogat ('Związek Wałowy Wisły i Nogatu') in 1889. It contained detailed regulations, such as technical maintenance parameters regarding water level in the Vistula River. For more *see* K. Cebulak, *Detla Wisły powyżej i poniżej poziomu morza* (Nowy Dwor Gdanski: Stowarzyszenie Żuławy i Lokalna Grupa Działania Żuławy i Mierzeja) (2010).
53. S. Mathew, *Wind Energy Fundamentals, Resource Analysis and Economics* (Berlin: Springer) (2006), at 4-6.
54. L.M. Kamp, R.E.H.M. Smits and C.D. Andriesse, 'Notions on Learning Applied to Wind Turbine Development in the Netherlands and Denmark', 32 *Energy Policy* 1625, at 1628 (2004).
55. Wind farm consists of an area with a group of wind turbines.
56. S. Breukers and M. Wolsink, 'Wind Energy Policies in the Netherlands: Institutional Capacity-building for Ecological Modernisation', 16(1) *Environmental Politics* 92, at 101-102 (2007).
57. Besluit wijziging milieuregels windturbines (14 oktober 2010).
58. Registers used for collections of taxes were already known to the civilizations of Mesopotamia, Assyria, Babylon and Egypt. For more *see* A. Hopfer and W. Wilkowski, 'Kataster nieruchomości w Polsce – jest czy go nie ma?', 79(1) *Przegląd Geodezyjny* 6, at 6 (2007).

– which were collected manually – were used to produce an inventory of lands and people. Accordingly, the population was classified into different social classes depending on income level.[59] In August's period, all the citizens were required to declare size and types of crops, as well as property income. The unified *capitastrum* (known then as *catastrum*) became the basis for taxation.[60]

The use of inventories for public purposes continued into the Middle Ages. An efficient collection of public receivables required the use of increasingly formalised registers. These registers reflected the various fiscal burdens on citizens. The use of registers made it possible to prevent fraud and enhance enforcement. In Poland, registers indicated tax obligations imposed by the King on particular states.[61] The owners were obliged to pay levies, both regular and extra regular.[62] In the fifteenth and sixteenth century, the extraordinary land tax (*poradlne*) was calculated on the basis of the register from 1578.[63] Registers were carried out also when the Crown was acquiring new territories.[64] For example, in 1650, a special register (*abiurata*) was issued. It indicated the number of declared land possessions belonging to the population of Smolensk, which had been annexed from the Russian Empire in the Time of Troubles.

A special regulation concerning registers *per se* started only in the eighteenth and nineteenth centuries. Poland, after the collapse of the Polish-Lithuanian Commonwealth, was divided between three countries, these being Habsburg Austria, the Kingdom of Prussia, and the Russian Empire. Each of these countries has started to implement their administration on the occupied territories. For example, the Kingdom of Prussia established a fixed register of land and real estate taxes in an 1867 act. This was followed by a land register ordinance aiming at the further standardisation of registers in 1872.[65]

In the past, the main purpose of registers had been to itemise lands and people in order to bring benefits to the state. Modern registers serve different functions. Past research considers digitisation as the main driver of change.[66] The state has recognised the benefits of technology by explicitly regulating the various electronic procedures that may be relevant for its citizens. An example may be Article 61(3a) of the Polish Code of Administrative Procedure.[67] Similarly, Article 14 indicates that public authorities in charge of public registers that use ICT systems must meet the minimum criteria established for any ICT system.[68] It is therefore possible to observe that the use of regulation is primarily aimed at meeting the needs of private citizens.

Data storage devices have developed considerably in the last centuries. However, regulation was only introduced when social conditions began to change. Again, technology *per se* was not sufficient to trigger any regulatory action. Conversely, the combination of technological change and social factors contributed to the development of new data storage regulation.

4 Concluding Remarks

The previous section considered historical responses of legal systems to changes in technology and, ultimately, society. It showed that competent institutions have employed different regulatory means for dealing with technological changes. However, a common pattern can be identified: legal intervention often follows social change. It seems that legislatures will not offer regulatory responses in every instance of technological change after the fact. For the expense of regulation to be justified, it is necessary for technological change to trigger social change, and further that the pre-existing legal framework cannot accommodate the social change.

There is thus no correspondence between the rate of technological change and the intensity of regulatory responses. On one hand, it is possible for new innovations to be distributed around the globe in hasty fashion due to globalisation and the advantages of modern-day life. On the other hand, governments and courts often struggle to provide speedy legal responses. Adapting old legal structures to new situations is sometimes insufficient. For regulation to be effective, some time must lapse between the innovation and the resultant change in social organisation. Then regulators have two main choices. Firstly, they can steer the evolution of rules

59. For more *see* W. Suder, *Census populi. Demografia starożytnego Rzymu* (Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego) (2003).
60. A. Zachariasz, 'Odczytywanie historii zapisanej w krajobrazie', 5(8) *Roczniki Geomatyki* 45 (2007).
61. Similarly as the case of Dutch water management that was presented in the preceding section, registers and special maps were prepared for a better administration of dikes and polders. Registers and maps were the basis for taxation assessment of people who were supposed to bear the costs of their maintenance. For the better preparation of maps, different regions started to put more formal obligations on surveyors. They were obliged to increase the level of competence and achieve a status of a sworn surveyor. For more *see* R.J.P. Kain and E. Baigent, *The Cadastral Map in the Service of the State: A History of Property Mapping* (Chicago: University of Chicago Press) (1992), at 9-24.
62. The most famous ones were: the land tax imposed in the fourteenth century by the King Kazimierz Wielki ('*poradlne*' – as of the sixteenth century change for '*łanowe*') or the seventeenth century family tax ('*podymne*').
63. A. Gomułowicz and J. Małecki, *Podatki i prawo podatkowe* (New York: Lexis Nexis) (2006), Chapter XVIII section 1.2.
64. A. Rachuba (ed.), *Metryka Litewska Rejestry podymnego Wielkiego Księstwa Litweskiego. Województwo Smoleńskie 1650 r.* (Warsaw: DiG Instytut Historii PAN) (2009).
65. M. Mika, 'Historia Katastru Polskiego', 6 *Infrastruktura i Ekologia Systemów Wiejskich* 75, at 78-80 (2010).
66. As described by Fred Cate, digital information is easier to generate, manipulate, transmit and store. Costs connected with these operations are lowered. Additionally, generation or storing of information triggers generation of additional digital information because of operating parameters of computer systems (*e.g.* through back-up copies). *See* F.H. Cate, *Privacy in the Information Age* (Washington: Brookings Institution Press) (1997), at 14-15.
67. Ustawa z dnia 14 czerwca 1960 r. Kodeks postępowania administracyjnego Dz.U. 1960 nr 30 poz. 168 t.j. The article was included in the novelization of the code in 2010. This article indicates that the date of initiation of proceedings at the request of the party, which is brought electronically, coincides with the day when the request is entered into the ICT system of the public administration authority.
68. Ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne Dz.U. 2005 nr 64 poz. 565 t.j.

alongside the development of technologies. This would allow the adaptation of the existing legal rules to new legal problems – this being the case of adapting the rules to changing features of money and public registers. Secondly, the regulator can devise new rules responding to new characteristics of technologies and related legal questions – this being the case of administrative rules regarding the usage of early windmills.

5 The Special Issue

This introductory article has started discussing how regulatory responses may not immediately follow the technological change after the fact. We did not aim to provide any definitive answer to that question. The intention was to instead present technological change as a part of social change. The current article has not, however, discussed any substantive regulatory efforts. That is the task that each of the articles in this issue takes up. More precisely, these articles will isolate specific issues raised by technology and compare them vis-à-vis existing regulatory frameworks. This type of operation requires a keen eye as well as employing, if needed, innovative approaches. In fact, regulatory adaptation may also necessitate from traditional forms of regulation.[69] The further articles of this special issue purport to do so.

The special issue consists of (another) four articles discussing legal approaches to socio-technological changes. These socio-technological changes are broadly connected with digitisation and the operation of the Internet. Some of the phenomena that are discussed in those articles are not new. However, digitisation has caused them to acquire new meanings and cause new problems. In all the remaining four contributions, the authors consider how law could or should approach socio-technological changes.

The article of *Katharina Kaesling* discusses enforcement mechanisms in social networks. The author tackles the well-known problem of hate speech and defamation and presents it in a new context involving a technological change. A technological change refers here to online social networks (*e.g.* Facebook) where hate speech or defamation can 'go viral'. As a result, the uncontrollable distribution goes beyond the control of the statement creator. Kaesling notices that this also goes beyond the capabilities of public policy makers. Accordingly, they need to rely on private entities.

Staying in the field of humans' online outputs, but turning more to the previously discussed data storage, the articles of *Alessandro El Khoury* and *Joanna Mazur* bring the problem of personal data and the right to information regarding automated decision-making solutions using personal data. Both articles contribute to describing social changes connected with people moving in the online reality and thus losing their anonymity. In this regard, the articles are based on the analysis of certain aspects of General Data Protection Regulation (GDPR). El Khoury discusses the binary notion of personal data and highlights its limitations in the GDPR. Mazur, on the other hand, brings limitations of GDPR by focusing on privacy protection in regard to the right to explanation. Like Kaesling, both authors highlight the contingent inability of public policy makers to draft timely, effective legal responses to socio-technological changes.

While the previous articles aim at analysing the situation of an individual in digital reality, *Morshed Mannan* brings in some aspects of worker cooperatives becoming a part of the digitised world. Mannan explores how organisational innovations can draw from blockchain projects and potentially facilitate the growth of worker cooperatives. The article of Mannan, similarly to the previous three, indicates the necessity of a continuous assessment of innovations, which cannot be detached from the context in which they occur. In other words, a proper understanding of the new technology would allow to better address the emerging legal issues.

69. Marchant, above n. 21.

Privatising Law Enforcement in Social Networks: A Comparative Model Analysis

Katharina Kaesling*

Abstract

These days, it appears to be common ground that what is illegal and punishable offline must also be treated as such in online formats. However, the enforcement of laws in the field of hate speech and fake news in social networks faces a number of challenges. Public policy makers increasingly rely on the regu-lation of user generated online content through private entities, i.e. through social networks as intermediaries. With this privat-ization of law enforcement, state actors hand the delicate bal-ancing of (fundamental) rights concerned off to private entities. Different strategies complementing traditional law enforcement mechanisms in Europe will be juxtaposed and analysed with particular regard to their respective incentive structures and conse-quential dangers for the exercise of fundamental rights. Propositions for a recommendable model honouring both pri-vate and public responsibilities will be presented.

1 Introduction: Fake News and Hate Speech on Social Networks

The Internet provides platforms for many forms of speech, with social networks emphasising user-generated content (UGC) like tweets, Facebook posts and Instagram pictures and videos. Digitally expressed 'hate speech' and 'fake news' on social networks have been the topic of public debate worldwide. The term 'fake news' has only recently entered colloquial language. While it is applied in different contexts to characterise political sentiments, manipulation and propaganda, use is made of the term here to describe deliberately false factual claims, i.e. disinformation with no viable basis. False claims are susceptible to be proven either wrong or false, which distinguishes them from opinions.

In that sense, fake news, much like hate speech and defamation, are not new phenomena. However, the particularities of the Internet add a new dimension to them.[1] The Web 2.0, i.e. websites designed to allow easy content creation by end users,[2] facilitates the dissemination of defamatory material. The reach of statements made online in social networks is increased by social media functions like sharing and liking posts. Due to these mechanisms, statements can 'go viral', i.e. trigger a snowball effect. They lead to a quick and global spread at no extra cost for the source. These effects largely lie beyond the control of the statement's creator, though they can be wilfully enhanced by different means including bots.

Hate speech is a political term rather than a legal one. It is not a clear-cut concept; it can encompass incivilities as well as insults and defamation. The specific danger of hate speech lies within the disparagement of a particular group of people. Traditionally, the term 'hate speech' refers to expressions inciting hatred, mainly racial, national or religious in nature.[3] Individuals are offended as members of a group, for example by reason of nationality, gender, race, ethnicity, religion or sexual tendencies. Hate speech has been found particularly worrisome by policy makers as it can stimulate further hatred against these groups. It can greatly influence recipients of such messages depending on the speaker's influence, the message's dissemination and the social and historical context and can be understood as call for action against the targeted groups. While hate can be planted both by illegal and undesirable content, the regulation of UGC, however, has to respect the boundaries of the law. These boundaries define the degree to which the exercise of individual fundamental rights such as free speech is limited in order to safeguard other rights such as the general right of personality.

In recent years, the question has shifted from *whether* to regulate online activities to *how* to do it. While John Perry Barlow proclaimed the independence of cyberspace in his 1996 declaration of the same name,[4] the current prevailing opinion is that illegality offline equals illegality online.[5] Substantive law standards are thus also

* The author is research coordinator at the Center for Advanced Study 'Law as Culture', University of Bonn.

1. D. Cucereanu, *Aspects of Regulating Freedom of Expression on the Internet* (2008), at 7.
2. T. O'Reilly, *What Is Web 2.0 – Design Patterns and Business Models for the Next Generation of Software* (2005), available at https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html.
3. *See* H. Darbishire, Hate Speech: New European Perspectives, *Roma Rights*, No. 4 (1999), available at www.errc.org/article/hate-speech-new-european-perspective/1129; F.M. Lawrence, 'Resolving the Hate Crimes/Hate Speech Paradox: Punishing Bias Crimes and Protecting Racist Speech', 68 *Notre Dame Law Review* 673 (1993).
4. J. P. Barlow, *A Declaration of the Independence of Cyberspace*, Davos, Switzerland, 8 February 1996, available at https://www.eff.org/de/cyberspace-independence.
5. *See* UK House of Commons, 'Hate Crime: Abuse, Hate and Extremism Online', *Fourteenth Report of Session* 2016-17, HC 609, at 11 (2017);

doi: 10.5553/ELR.000115 - ELR december 2018 | No. 3

applicable to online contexts. Nonetheless, an online–offline divide cannot be denied when it comes to the enforcement of substantive law, namely, criminal law provisions, in social networks. The special environments of social networks and the often-invoked borderless nature of the Internet pose massive challenges for an effective law enforcement. Particularities of these environments, principally the relative anonymity of users, the fast dissemination of large volumes of UGC across borders and the global activity of platform operators set significant hurdles.

Social networks were initially rather seen as merely opening new means of communication for users without triggering a responsibility for UGC.[6] Faced with the particularities of the Internet, state actors have increasingly opted to assign responsibility to social networks as intermediaries. Sweden already passed a law to that effect in 1998,[7] while there were no 'precise ideas' on the enforcement of ICT law in Germany, France, the United Kingdom and the United States in 2000.[8] The debates on fake news and hate speech emerged later on and recently invited a number of state interventions worldwide.

In Germany, the 'Act to Improve the Enforcement of Rights on Social Networks' was adopted in 2017.[9] It has gained international attention, as it threatens large fines on social networks that systematically breach their obligations regarding the timely removal of illegal UGC. In the United Kingdom and the Russian Federation, the German law has been cited as model for respective legislative projects. The UK Home Affairs Committee of the House of Commons recommended 'that the Government consult on a system of escalating sanctions to include meaningful fines for social media companies which fail to remove illegal content within a strict timeframe'.[10] The Russian Duma advanced a bill considered 'copy-and-paste of Germany's hate speech law' shortly after its adoption.[11]

In Europe and elsewhere, traditional law enforcement mechanisms are considered inadequate to implement legal provisions in the field of online hate speech and fake news. More and more public policy makers in Europe and elsewhere are contemplating and adopting various additional mechanisms to put the respective laws into effect.

In that context, the German venture appears to show model character, but is it really a good policy example? How does it hold up in comparison with other systems in the European Union? A comparative model analysis will reveal advantages and dangers so as to contribute to the shaping of a superior model for law enforcement in social networks.

Different laws and policy approaches currently in effect in the Europe will be described (2) before turning to the underlying question of delimitating the roles of public and private actors (3). Against that background, three models will be distinguished and evaluated with particular regard to dangers for the exercise of free speech (4). Finally, conclusions and propositions for a recommendable model for law enforcement in social networks honouring both private and public responsibilities will be presented (5).

2 Law Enforcement Strategies in Social Networks

Law enforcement has a servicing function in relation to the substantive law. Traditional law enforcement mechanisms are put into place by the state. More and more, alternatives are considered by policy makers in numerous fields of law.[12] With regard to illegal UGC on social networks, legal norms have been created and policy initiatives launched to complement criminal prosecution and civil law actions. Balkin characterised these informal control measures as new-school speech regulation rather than old-school speech regulation like penalties and injunctions directed at speakers and publishers.[13]

Following a short overview of the legal provisions to be enforced in the context of hate speech and fake news (2.1), the traditional law enforcement strategies of criminal prosecution and civil law actions will be scanned with particular regard to mechanisms to overcome online anonymity (2.2). These laws are complemented by EU law and policy. The elemental legal source within the European Union is the E-Commerce Directive of 2000 (2.3). More recently, the EU Commission has, however, favoured voluntary commitments by social networks (2.4). On a national level, the German and the Swedish regulation will be described (2.5 and 2.6) before briefly summarising the findings (2.7).

2.1 Enforceable Legal Provisions

Online content is illegal when it is contrary to the applicable legal order. In the context of fake news and hate speech, relevant legal provisions are mainly national criminal and civil law affording protection of honour and rights of personality. In addition to criminal prosecution, unlawful statements touching a person's honour, reputation or personality rights generally also trigger the civil liability of the infringer.

B.-J. Koops, 'Cybercrime Legislation in the Netherlands', in P.C. Reich (ed.), *Cybercrime and Security* (2005) 1, at 6.

6. D.M. Boyd and N.B. Ellison, 'Social Network Sites: Definition, History, and Scholarship', 13 *Journal of Computer-Mediated Communication* 210 (2007).
7. *See* 2.6.
8. B.-J. Koops, J. E. J. Prins & H. Hijmans, *ICT Law and Internationalisation* (2000), at 129.
9. *See* 2.5.
10. House of Commons, above n. 5, at 14.
11. Reporters Without Borders, 'Russian Bill is Copy-And-Paste of Germany's Hate Speech Law', published 19 July 2017, available at https://rsf.org/en/news/russian-bill-copy-and-paste-germanys-hate-speech-law.
12. *See* for competition law and ADR, J. Basedow, ‚Rechtsdurchsetzung und Streitbeilegung – Die Vielfalt von Durchsetzungsformen im Lichte von Zielkonflikten', *JZ* 1, at 5 ff. (2018).
13. J.M. Balkin, 'Old-School/New-School Speech Regulation', 127 *Harvard Law Review* 2296, at 2298 (2014).

Despite certain efforts,[14] fake news is not as such illegal in most countries. Regarding both hate speech and fake news, defamation and insult laws are relevant. A number of legal orders foresee a specific criminal provision for cases in which the fact supported by the speaker is false.[15] Prohibited behaviours in the context of hate speech vary widely, also among the Member States of the European Union.[16] International instruments such as the *EU Council Framework Decision on combating certain forms and expressions of racism and xenophobia by means of criminal law*,[17] the *UN Convention on the elimination of all forms of racial discrimination of 21 December 1965* and the *Council of Europe Additional Protocol to the Convention on cybercrime concerning the criminalization of acts of a racist and xenophobic nature committed through computer systems of 28 January 2003* have only had a very limited harmonising effect.

Under the aforementioned EU Framework Decision, hate speech is to be considered a criminal offence when it publicly encourages violence or hatred against a person or group of people because of race, colour, religion, descent or national or ethnic origin. Even so, public incitement to violence is only criminalised in some Member States when its manner is likely to disturb public order or public peace.[18] In addition, varying defamation and insult laws play a considerable role in the fight against hate speech when penalising collective defamation and insults of groups.

In some countries, mainly Common Law countries, a demise of criminal defamation and insult law could be observed.[19] In the context of fake news and hate speech, however, these provisions have (re-)gained importance. Online communication, especially on social networks, has made defamation and insult laws very topical.[20] While rules on the illegality of statements differ, the problem of how to put existing rules to effect in social networks contexts occurs in all legal orders.

2.2 Traditional Public Law Enforcement Mechanisms and Its Limits

Traditional public law enforcement encompasses the criminal prosecution of perpetrators (2.2.1) as well as civil legal protection afforded (2.2.2). In online contexts, their effectiveness is largely called into question by the relative anonymity provided to social network users so that tools helping to overcome that online anonymity are specifically taken into account (2.2.3).

2.2.1 Criminal Prosecution

Criminal prosecution presupposes not only personal jurisdiction over the accused, but generally also his presence at trial, which might prove difficult in international contexts with extradition treaties being limited. Criminal provisions are generally enforced by instituting proceedings in the proper court on behalf of the public. In that case, the public prosecutor somehow learns of potential illegal online activity, investigates *ex officio* and then brings charges. Especially concerning general defamation and insult laws, prosecution presupposes the active involvement of the affected individual. In numerous legal orders, such charges cannot be brought without the victim's consent.[21] Alternatively, victims can act as a private prosecutors, file a criminal suit and prove the relevant facts of the case without the public prosecutor's participation.[22]

The enforcement of general defamation and insult laws is consequently already limited as it largely depends on the victim's authorisation or even legal action. Insofar, law enforcement is left to the victim's discretion. Victims also have the option of choosing civil over criminal action, which might be preferable due to the lighter burden of proof in civil cases.[23] Criminal cases can also be combined with the corresponding civil ones in many legal orders.[24]

2.2.2 Civil Legal Protection

UGC on social networks can also trigger the civil liability of the infringer. In the civil law context, sanctions generally include injunctive relief and damages. Victims of untrue rumours disseminated on social networks, for example, have the demand injunctive relief and revocation from the infringer.[25] This right can be secured by means of interim injunctions. In a social media context, the concerned can thus demand the deletion of tweets, media or short postings. The further dissemination of false information can be prevented by an order to rectify false statements made. In some cases and countries, the victim also has general civil law claims against the platform operator, i.e. the social network provider. For example, under German law, the affected individual can request that the platform operator (temporarily) blocks the account of the infringer in exceptional cases.[26]

Civil (interim) legal protection generally depends on the active intervention of the victims. They have to issue takedown notices or institute civil legal proceedings.

14. *E.g.* U.S. *Honest Ads Bill* of 2017, 115th Congress, 1st session, S. 1989.
15. *E.g.* Germany, Greece and Switzerland.
16. Mandola Intermediate Report – Monitoring and Detecting Online Hate Speech in the Framework of Rights, Equality and Citizenship Programme of the EU Commission of 20 July 2016, at 9, available at http://mandola-project.eu/m/filer_public/7b/8f/7b8f3f88-2270-47ed-8791-8fbfb320b755/mandola-d21.pdf.
17. 2008/913/JHA of November 2008; follow-up to Joint Action 96/443/JHA of 15 July 1996.
18. Mandola Intermediate Report, above n. 16, at 10.
19. *See e.g.* UK Defamation Act 2013 (c 26); for an overview of U.S. States; *see* L.Y. Garfield, 'The Death of Slander', 17 *Columbia Journal of Law & the Arts* 17, at 53-54.
20. *See* for the U.S.A. A. J. Wagner and A. L. Fargo, 'Criminal Libel in the Land of the First Amendment', *Special Report for the International Press Institute*, at 27-28 (2015).
21. *See* S. Griffen, 'Defamation and Insult Laws in the OSCE Region: A Comparative Study', at 10 (2017), available at https://www.osce.org/fom/303181?download=true.
22. *E.g.* Russian Criminal Code Art. 128.1(1); German Criminal Procedural Code Section 374 para. 1, No. 2.
23. Griffen, above n. 21, at 11.
24. *Ibid*. at 10.
25. *E.g.,* German civil code Section 823 para. 1 Civil Code in conjunction with Art. 1 para. 1 and Art. 2 para. 1 Basic Law and Section 1004 Civil Code; Civil Code Section 823 para. 2 in conjunction with criminal law.
26. C. M. Giebel, Zivilrechtlicher Rechtsschutz gegen Cybermobbing in sozialen Netzwerken, *NJW* 977, 980 (2017).

Judicial legal protection can be costly, particularly if multiple jurisdictions are involved as it is likely regarding online UGC.

Victims usually also have a claim for damages if their personality rights were infringed. Damages are supposed to compensate the victim for any harm to his or her reputation or emotional well-being. Their amount differs considerably from legal order to legal order; the incentives for the victim to pursue such a civil legal action vary accordingly.

2.2.3 Mechanisms to Overcome Online Anonymity

The identification of the infringer as potential perpetrator and defendant is crucial for both criminal prosecution and civil legal protection.[27] Social media, however, offers a relative anonymity to its users. Commonly, identity verifications are not required. E-mail addresses are generally needed to register, but can in turn be easily created using false information. IP addresses associated with illegal postings can sometimes, but not always, be traced back to the actual user at the time in question. The anonymity provided is not absolute, as the infringer's identity can also be revealed in the course of investigations going off his social media contacts and information. In many cases, effective legal protection will, however, hinge on mechanisms to overcome that anonymity.

Insofar, the protection of personality rights lags considerably behind intellectual property law. The identification of the infringer can be a question of the applicable substantive or procedural law. By now, a number of legal orders know mechanisms to identify online users hiding behind a pseudonym or commenting anonymously. For example, in Germany, platform operators are now allowed to disclose details about users in cases of insult, defamation, incitement to violence and similar instances.[28] In contrast to copyright law[29] and despite proposals to that effect,[30] there is, however, no specific claim to information in that context.[31] If the applicable substantive law does not provide for a claim for information, there might be procedural court orders available to that end. In the famous UK Internet libel case *Motley Fool*, the service provider was ordered to reveal details about the user posting under a pseudonym under Section 10 of the Contempt of Court Act.[32]

The need for identification of the infringer also affects the ability to quickly move forward with the initiation of judicial protection measures, above all interim legal protection. Its effectiveness is correspondingly tied to the processing time at the competent court, with time being of the essence with the risks of quick uncontrolled proliferation of the personality right violations in online contexts.

2.3 The EU E-Commerce Directive

The basic EU rules on duties of social networks regarding illegal UGC on their platforms were already included in the E-Commerce Directive of 2000 (ECD).[33] The ECD aims to establish a coherent legal framework for the development of electronic commerce within the Single Market.[34] The ECD does not pertain to social networks specifically and concerns all types of illegal content.

Primarily, it regulates the role of information society service providers (ISPs) such as social networks. The ECD distinguishes between three types of services depending on the ISP's activities, i.e. mere conduit (Article 12 ECD), caching (Article 13 ECD) and hosting (Article 14 ECD). Social networks fall under the third category of hosting services, i.e. ISPs that store information by a recipient of the service. These ISPs are not liable for information stored at the request of a recipient on two conditions. Firstly, the ISP may not have actual knowledge of the illegal activity and secondly, the ISP has to act expeditiously to remove or to disable access to the information (Article 14, Recital 46 ECD).

According to Article 15 ECD, Member States shall not impose any obligation to monitor the information that they transmit or store or a general obligation to actively seek facts or circumstances indicating illegal activity on any type of ISP.[35] Member States may, however, establish specific requirements that must be fulfilled expeditiously prior to the removal or disabling of information (Recital 46 ECD) and monitoring obligations in specific cases (Recital 47 ECD). They may require hosting services to apply duties of care that can reasonably be expected from them in order to detect and prevent certain types of illegal activities (Recital 48 ECD). In summary, the ECD only prohibits a general obligation to monitor, while more specific monitoring obligations under national law are permissible.[36] Distinctive features of these two categories remain to be developed.[37]

Legal uncertainty exists regarding the delimitation of the types of ISPs and as to the definition of the relevant terms, such as 'expeditiously', which does not give any specification of a particular time frame in question. Recital 42 ECD clarifies that the exemptions from liability only extend to 'cases where the activity of the information society service provider is limited to the technical process of operating and giving access to a communication network'. It further specifies that this activity is of a mere technical, automatic and passive

27. Cf. R. Perry, and T. Zarsky, Who Should Be Liable for Online Anonymous Defamation?, *University of Chicago Law Review Dialogue* (2015) 162.
28. Section 14 para. 3-5 in conjunction with Section 15 German Teleservices Act and Section 1 III NetzDG.
29. German Copyright Law Section 101.
30. Statement of the German Federal Assembly on the 2nd amending law of the German Teleservices Act of 6 November 2015, BT-Drs. 18/6745.
31. G. Spindler, 'Rechtsdurchsetzung von Persönlichkeitsrechten', *GRUR* 365, at 372 (2018).
32. *Totalise Plc* v. *The Motley Fool Ltd. Anor* [2001] EWHC 706 (QB).
33. Directive 2000/31/EC.
34. EU Commission Press Release, Electronic commerce: Commission proposes legal framework, IP/98/999, Brussels, 18 November 1998.
35. *See also* Recital 47 ECD.
36. *Ibid*.
37. P. Van Eecke, 'Online Service Providers and Liability: A Plea for a Balanced Approach'48 *CMLR* 1455, at 1486-1487 (2011).

nature, thus implying that the ISP has neither knowledge of nor control over the information that is transmitted or stored.[38] Recital 46 spells out that the expeditious removal or disabling of access is in fact a precondition for the limitation of liability. Failing to comply with that obligation, ISPs are not in the safe harbour. The ECD has therefore led to the institution of takedown procedures for social networks.

According to Recital 49 ECD, Member States and the Commission are to encourage the drawing-up of voluntary codes of conduct. In line with this, the Commission has recently presented more targeted approaches aimed at hate speech and fake news.

2.4 EU Hate Speech Code of Conduct and Fake News Initiative

Both with regard to hate speech and to fake news, the EU Commission now works with the biggest social networks towards voluntary commitments without sanctions for non-compliance.

2.4.1 Hate Speech Code of Conduct

In order to combat illegal online hate speech, the European Commission and significant IT companies announced the *Code of Conduct on countering illegal hate speech online* in 2016. This code of conduct was agreed upon by Facebook, Microsoft, Twitter and YouTube. In 2018, Instagram, Google+ and Snapchat also publicly committed to it.[39]

The Hate Speech Code of Conduct relies on the signatory private companies to take the lead, as emphasised by the EU Commission.[40] It does not primarily aim at ensuring compliance with national laws. Social networks firstly test the content against their individual 'Rules or Community guidelines', which have to clarify that the promotion of incitement to violence and hateful conduct is prohibited.[41]

The review of UGC by the participating IT companies is limited to notified posts. Posts can be notified by other users, special 'trusted flaggers' that can use specific channels to alert the social networks and national law enforcement authorities that learned about that content. Upon notification, they examine the request for removal against their rules and community guidelines and where necessary national laws on hate speech transposing the Framework Decision 2008/913/JHA. To that purpose, they set up 'dedicated teams'.[42] The social networks pledged to assess 'the majority of valid notifications' in less than twenty-four hours after notification and remove or disable access to such content, if necessary.[43] Notification of law enforcement authorities and 'trusted flaggers' should be addressed more quickly than others.[44]

In March 2018, the Commission has published an additional *Recommendation on measures to effectively tackle illegal content online*.[45] It reiterates the importance of cooperation of social networks with state actors and further specifies them. Service providers are encouraged to take voluntary proactive measures beyond the notice-and-action mechanisms, including automated means.[46]

2.4.2 Fake News Initiative

In light of the fake information spread on social media in the run-up to the 2016 US presidential election, the European Parliament and Commission are particularly worried about fake news ahead of the 2019 EU election.[47] So far, it has tackled the problem by setting the Fake News Initiative into motion and threatening legislation if social network self-regulation does not prove sufficient. In April 2018, the European Commission gave online platforms the assignment to develop a common Code of Practice on Disinformation by July 2018.[48] This instrument of voluntary public commitment shall be prepared by a multi-stakeholder forum representing not only online platforms, but also the advertising industry and major advertisers. The Commission also urged social networks to promote voluntary online identification systems. A Commission report on the progress made shall be published by December 2018. It will include an evaluation as to whether further (legislative) action is warranted.[49]

The Commission has stressed that proactive measures taken by social networks – as they are encouraged by its fake news initiative – are without prejudice to Article 15 (1) ECD.[50] This also includes 'using automated means in certain cases',[51] which appears to refer to a voluntary monitoring with the help of available filtering and/or research software. According to the Commission, hosting service providers therefore do not risk losing their liability exemption under Article 14 ECD.[52]

2.5 The German Act to Improve the Enforcement of Rights on Social Networks

The recently adopted German Netzwerkdurchsetzungsgesetz (NetzDG) aims to raise the level of protection on social media.[53] The German legislator introduced this Act in 2017 specifically as action against hate speech and fake news following reports about the latter in the

38. Cases C-236/08 – 238/08, *Google France and others* v. *Louis Vuitton and others* [2010] ECR I-02417, Rec. 120.
39. European Commission, Daily News 7 May 2018, MEX/18/3723.
40. Hate Speech Code of Conduct at 2, available at http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=54300; Press release 'European Commission and IT Companies announce Code of Conduct on illegal online hate speech Brussels' of 31 May 2016, IP/16/1937.
41. *Ibid*.
42. *Ibid*.
43. *Ibid*.; European Commission Communication 'Tackling Illegal Content Online. Towards an enhanced responsibility of online platforms', COM (2017) 555 Final, 28 September 2017, para. 13.
44. European Commission, *ibid*.
45. Commission Recommendation of 1 March 2018 on measures to effectively tackle illegal content online (C [2018] 1177 final).
46. *Ibid*, at Rec. (24).
47. EPRS, Online disinformation and the EU's response, PE 620.230 – May 2018.
48. EU Commission Press Release, Tackling online disinformation: Commission proposes an EU-wide Code of Practice, IP/18/3370, Brussels, 26 April 2018.
49. *Ibid*.
50. Above *see* n. 45, Rec. 24.
51. *Ibid.*
52. European Commission, above n. 43, at para. 3.3.
53. R. Schütz, 'Regulierung in der digitalen Medienwelt', *MMR* 36 (2018).

course of the last U.S. Presidential Election. Its name – 'Act to Improve the Enforcement of Rights on Social Networks' – highlights the difficulty the German legislator perceived regarding law enforcement in online contexts and against globally active platform operators that do not have a bricks-and-mortar presence in the state's controlled territory. The NetzDG therefore creates a link to that territory the legislator can control by requiring every social media network to designate a domestic agent as point of contact for public authorities.

The Act did not introduce any new enforceable legal provisions. Instead, illegality within the meaning of the NetzDG is defined by referring to more than twenty criminal law provisions, including defamation and insult, public incitement to crime and hatred as well as propaganda and use of symbols of unconstitutional organisations. In principle, the NetzDG ascertains existing obligations in the framework of the notice-and-takedown procedures as instituted following the ECD. However, it adds further specifications regarding the self-control procedures of social networks and provides for sanctions in case of non-compliance.

The Act sets standards for the social network's complaint mechanism and decision-making. Under the NetzDG, social networks are obligated to institute a procedure for complaints regarding illegal content that allows for a timely deletion. The deadlines for removal depend on the obviousness of the content's illegality. Content that is 'clearly illegal' has to be blocked within twenty-four hours after receiving a complaint. If the illegality is less obvious, the social network has seven days to investigate and delete, with the deadline being extended in case of participation of an 'agency of regulated self-regulation'. These agencies are private outside institutions that were recognised by the Ministry of Justice according to guidelines set out in the NetzDG. Above all, its examiners have to be independent and possess the necessary expertise. Moreover, the agency of regulated self-regulation has to guarantee an examination within seven days and foresee rules of procedure and a complaint mechanism. In case of organisational and systematic failure to comply, social media networks may be fined up to fifty million EUR by the competent public authority. This includes a systematically false decision-making practice, but not a single failure to remove notified illegal UGC. Social networks receiving more than hundred complaints about illegal content in a calendar year are also obliged to publish biannual reports on these complaint procedures.

It is unclear how the NetzDG fits with the ECD.[54] In light of the number of issues, the German legislator at least risked a potential violation of ECD and other EU law principles, most notably the country-of-origin principle as mirrored in Article 3 ECD.[55] The German legislator applied a public policy derogation as criminal offences needed to be respected and the fight against hate speech made regulation necessary.[56] It can also be argued that the NetzDG imposes considerably higher standards on social networks than foreseen by the ECD.[57] While the NetzDG maintains the ECD's general liability and notification system, it sets rather precise deadlines for the deletion of illegal content, which begin with the receipt of the respective complaint.[58] In that regard, the German legislation could possibly exceed the Member States' margin of discretion. Especially in light of these EU law concerns, the NetzDG demonstrates the legislator's determination to combat illegal content like hate speech and fake news more efficiently. The means of choice for the German legislator is – not unlike the EU Commission's more recent approaches – imposing more responsibility on social networks.

2.6 The Swedish Act on Responsibility for Electronic Bulletin Boards

Sweden already regulated illegal content management on 'electronic bulletin boards' in 1998 with the Act on Responsibility for Electronic Bulletin Boards (EBB).[59] According to its Section 1, electronic bulletin boards are services for mediation of electronic messages, i.e. platforms where users can upload data, read news and exchange messages with other users.[60] The Act aims at establishing the provider's responsibility to remove messages that clearly constitute incitement, hate speech, child pornography, unlawful depiction of violence or messages where the posting user manifestly infringes on copyright.[61] The ECD was incorporated by the Act on Electronic Commerce and Information Society Services of 2000.[62]

Under the Swedish regime, owners and providers of Internet-based information services are responsible for illegal content on their systems.[63] UGC considered illegal under the EBB has to be removed by the service provider. According to Section 4 EBB, the service provider has to supervise the service to an extent that is reasonable considering the extent and objective of the service in order to fulfil its obligations to remove or block illegal content under Section 5 EBB. Service providers like social networks thus generally have an obligation to

54. Cf. W. Schulz, 'Regulating Intermediaries to Protect Privacy Online – The Case of the German NetzDG', in M. Albers and I. Sarlet (ed.), *Personality and Data Protection Rights on the Internet* (2018) 1, at 6 et seq., available at https://ssrn.com/abstract=3216572.
55. In support of a violation M. Liesching, 'Die Durchsetzung von Verfassungs- und Europarecht gegen das NetzDG', *MMR* 26, at 29 (2018); G. Spindler, 'Der Regierungsentwurf zum Netzwerkdurchsetzungsgesetz – europarechtswidrig?', *ZUM* 474, at 477 (2017).
56. Art. 3 lit a, no. i. ECD, Bundestag printed matter 18/12356, at 13-4.
57. Spindler, above n. 55, at 478; Liesching, above n. 55, at 29.
58. Section 2 para. 2, No. 2 and 3 NetzDG.
59. Swedish Code of Statutes 1998:112.
60. C. Kirchberger, *Cyber Law in Sweden* (2011), at 35.
61. S. Larsson, 'Metaphors, Law and Digital Phenomena: The Swedish Pirate Bay Court Case', *International Journal of Law and Information Technology* 370 (2013); B.-J. Koops, J. E. J. Prins & H. Hijmans, above n. 8, at 164.
62. Swedish Code of Statutes 2002:562.
63. G. Antonsson and A. Fernlund: Franchising: E-Commerce and Data Protection Issues in Sweden, 4 *Int'l J. Franchising L.* 26, at 26-7 (2006); M. Klang, *The APC European Internet Rights Project, Country Report — Sweden*, available at http://europe.rights.apc.org/c_rpt/sweden.html.

monitor its platforms.[64] Social networks do not fall under the explicit exemptions, as they were introduced to implement the ECD categories of mere conduit and caching.[65]

Removal obligations are limited to specific matters. Relevant illegality under Swedish law is defined in Section 5 with regard to Swedish criminal law provisions on the incitement of rebellion, agitation against a national ethnic group, child pornography crime, and unlawful depiction of violence as well as the infringement of copyrights. An intentional or negligent violation of this obligation is a criminal offence.[66]

Limitations to the general obligation to monitor are set by the law itself, as it stipulates that this obligation is limited to a reasonable extent. Consequently, not all UGC has to be checked under all circumstances. Periodical controls can be sufficient.[67] Service providers like social networks can also make use of notification procedures like user reporting functions and abuse boards, to which users can complain about illegal messages.[68] It is however not sufficient to generally limit the social network's activity to reaction to complaints.[69] How often the provider has to go through the content of the electronic bulletin board depends on the content of the service.[70] In particular, commercial services must check more regularly than private services.[71] For areas where illegal content is common, the provider of the area must check regularly and remove illegal content.[72] Hence, social network providers must maintain a (more) regular control if they learn of illegal UGC.[73]

2.7 Summary

In summary, traditional public law enforcement is increasingly complemented by additional mechanisms largely depending on social networks as intermediaries. These mechanisms range from voluntary self-commitment, code of conducts to negligence liability systems with or without fines to strict liability approaches with an obligation to monitor.

Law enforcement is traditionally seen as state function, albeit relying on the active intervention of the entitled parties. Illegal content is created and disseminated in multilateral constellations involving the infringer and perpetrator, the victim(s), social networks as intermediaries and other users that come into contact with prohibited forms of hate speech and fake news. Within this multi-player context, public and private responsibilities of the actors involved are to be marked down.

64. *See* T. Verbiest, G. Spindler and G.M. Riccio, Study on the Liability of Internet Intermediaries (November 12, 2007), available at http://dx.doi.org/10.2139/ssrn.2575069, p. 109; Klang, above n. 63.
65. Antonsson and Fernlund, above n. 63, at 27.
66. Verbiest et al., above n. 64.
67. *Ibid.*
68. J. Palme, *English Translation of the Swedish Law on Responsibilities for Internet Information Providers*, 3 June 1998, available at https://people.dsv.su.se/~jpalme/society/swedish-bbs-act.html.
69. Klang, above n. 63; Antonsson and Fernlund, above n. 63, at 27.
70. Palme, above n. 68, Comment to Art. 4.
71. *Ibid.*
72. *Ibid.*
73. Klang, above n. 63.

3 Private and Public Responsibilities

The Internet is governed by multiple, overlapping modalities including social norms, code, market and the law. Social media companies serve as intermediaries, who supply the environment enabling users to create and access UGC. Naturally, they are not public utilities, but private entities carrying out a business endeavour. While they are thus prone to implement market-oriented business strategies, it is the public policy makers' task to adequately safeguard the exercise of fundamental rights. At the same time, the individual social media user voluntarily joins and frequents social networks according to his habits. The task of preventing and combating hate speech and fake news could be attributed to all three groups of actors – social media users, social networks and public policy makers.[74]

Could social media users not simply be trusted to make their own choices, thus making any intervention from the other two actors expendable (3.1)? Why should law enforcement not be largely delegated to social network providers (3.2) and what are public non-disposable core responsibilities (3.3)? These questions will be answered in order to pave the way for a comparative model analysis against that background (4).

3.1 User Self-Censorship

It has been argued that commercially available filtering software can be applied by users to block sites on the basis of content, thus making (additional) governmental regulation unnecessary.[75] Individual users can customise these filters in accordance with their moral and social attitudes and by this means control their receptions.[76] Rather than a censorship by the state, users only censor themselves. Technological tools that allow the blocking of sites on the basis of content were especially developed to shield children from inappropriate content.[77] Shortcomings of these tools have however also been identified.[78] Like all technological tools, further development can certainly improve the overall software quality.

Even with enhanced technological tools, factual limits of hate speech would, however, be placed in hands of commercial interests.[79] Moreover, with the referral to commercially available filtering devices, hate speech remains accessible to all those that did not install adequate filtering software. The socially destabilising force of hate

74. Cf. for a new structure of speech regulation J.M. Balkin, 'Free Speech is a Triangle', Colum. L. Rev. 1, at 4 et seq (forthcoming 2018), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3186205.
75. R. Weintraub-Reiter, 'Hate Speech Over the Internet: A Traditional Constitutional Analysis or a New Cyber Constitution?', 8 *Boston University Public Interest Law Journal* 145, at 173 (1998).
76. *Ibid*.
77. *E.g.* CyberPatrol, NetNanny, SurfWatch, HateFilter.
78. M. Krantz, 'Censor's Sensibility: Are Web Filters Valuable Watchdogs or Just New Online Thought Police?', *Time Magazine*, 11 August 1997, 48.
79. *See* A. Tsesis, 'Hate in Cyberspace: Regulating Hate Speech on the Internet', 38 *San Diego Law Review* 817, at 867 (2001).

speech is therefore not fended off.[80] Compliance with laws lies in the discretion of users, thereby circumventing the ratio of hate speech laws. Particularly with regard to content whose illegality stems from incitement to violence, the danger lies primarily in reaching out to those recipients who might not be interested in blocking that very illegal content. Self-censorship by users is therefore clearly insufficient for protecting societal welfare and the individual rights at stake.[81]

3.2 Law Enforcement in the Hands of Social Networks – Why Not?

Social networks operate platforms for social traffic online. By creating these environments, they do not only render communication between users possible, but also shape it according to the platform design. Unlike telephone landlines, they do not only make a means of communication between a small number of communicators possible for a monetary consideration.[82] The success of business models of Facebook, Twitter, Instagram and the like is based on the creation of UGC in large volumes and at fast publishing rates. Social media has a magnifying effect for all ideas and opinions expressed, while at the same time offering a (relative) anonymity to the user creating content. It also favours the creation and organisation of groups on national and international levels, including extreme movements prone to generate illegal content.[83] Hence, the facilitation of the spread of illegal content is provoked by the business model itself. Why then not simply give the responsibility for the lawfulness of content to the social media operators?

As platform operators, these social networks like Twitter, Facebook and Instagram create the environment for user statements and naturally govern social interaction on their platforms. They certainly are in a good position to carry out control[84] – arguably in a better position than state regulators.[85] Social networks therefore seem to be the point of least cost at first glance.[86] Unlike national state governments, social networks are able to set rules on all markets they are active on (3.2.1). In addition, they can make more effective use of technology (3.2.2).[87] However, both of these apparent advantages are limited by practical considerations. Moreover, the application and interpretation of relevant (criminal) provisions by social networks hold considerable dangers for fundamental rights and thus (democratic) societies (3.2.3).

3.2.1 Social Media Policies and Terms of Use

Social media operators have long had policies against the use of hate speech as part of their corporate responsibility,[88] i.e. by reserving themselves the right to revoke accounts that are against their hate speech policy. They have contracts with their users and can unilaterally impose terms of use for their worldwide operations. These contracts generally contain provisions prohibiting users from creating content in violation of the law, especially defamatory, harassing, hateful, or racially or ethnically offensive content. For example, Facebook's terms of use require the users not to bully, intimidate or harass any user and not to post content that is hate speech, threatening, or pornographic; incites violence; or contains graphic or gratuitous violence.[89]

Such terms of use are, however, not effective, not even for the purposes of deterrence. Social media terms of use, much like any small print, are hardly actually read by the end users who manifest their consent merely by clicking a button in a pop-up window or a dialogue box. It thus cannot even be assumed that the terms of use create awareness amongst the users.[90] Besides, users willing to violate criminal law are likely to be willing to violate terms of use as well.

3.2.2 Use of Technology

Social networks could implement technological tools for the detection and blocking of hate speech and fake news more effectively than external state actors.[91] However, not only state actors, but also social networks are confronted with the large volume and high rate of publication of UGC. This renders thorough monitoring of content fairly difficult for those intermediaries as well.[92]

With regard to copyright infringements, filtering mechanisms employing digital fingerprinting, i.e. matching uploaded and protected works, have been successfully employed on a voluntary basis for years. The software 'Content ID' has been used by YouTube and Facebook to filter illegal extremist content.[93] Only after clearly extremist content has been identified, can a hash be created in order to compare this content via digital fingerprinting. While other filtering devices are tested, there is currently no appropriate technology that allows for an effective monitoring for illegal hate speech and fake news. Striking a fair balance between fundamental rights affected in specific cases at hand is not easily programmed.[94] Filtering tools would also have to take into account the specificities of the jurisdictions concerned.

80. Tsesis, above n. 79, at 866.
81. *Ibid*.
82. Cf. *See* T. Gillespie, 'Regulation of and by Platforms,' in J. Burgess, A. Marwick, T. Poell (eds.), *The Sage Handbook of Social Media* (2017), at 257-8; T. Gillespie, 'Platforms are Not Intermediaries', 2 *Georgetown Law Technology Review* 198 (2018).
83. *See* B. Perry and P. Olsson, 'Cyberhate: The Globalization of Hate', 18 *Information and Communications Technology Law* 185 (2009).
84. C.E. George and J. Scerri, 'Web 2.0 and User-Generated Content: Legal Challenges in the New Frontier', 2 *Journal of Information, Law and Technology* 1, at 10 (2007).
85. *Ibid*. at 18.
86. *See* Rec. 59 Copyright in the Information Society Directive.
87. George and Scerri, above n. 84.
88. K. Klonick, 'The New Governors: The People, Rules and Processes Governing Online Speech', 131 *Harvard Law Review* 1598, at 1626 (2018).
89. Facebook Terms of Use U.S., retrieved from Germany, available at https://www.facebook.com/terms.php (last visited 18 June 2018), Section 3.
90. George and Scerri, above n. 84, at 12.
91. *Ibid*., at 18.
92. *Ibid*., at 10.
93. O. Solon, 'Facebook, Twitter, Google and Microsoft Team up to Tackle Extremist Content', *The Guardian*, 6 December 2016.
94. *See* D. Burk and J. Cohen, 'Fair Use Infrastructure for Copyright Management Systems', *Georgetown Public Law Research Paper* (2000) 239731/2000 for 'fair use' in copyright law.

3.2.3 *Dangers of the Application and Interpretation of the Law by Private Entities*

Assigning responsibility for the lawfulness of UGC to social networks as intermediaries involves the application and interpretation of the relevant, mainly criminal, provisions. It is then left up to social media operators as private entities to draw the oftentimes thin line between legitimate exercise of the right to free speech and criminal conduct. This is namely – but not exclusively – due to the underlying importance of constitutional law. The interpretation of criminal legal norms safeguarding personal honour and dignity against factual claims, opinions and incitement to hatred and violence as well as their application to the individual case is strongly shaped by fundamental right considerations. The application and interpretation of the provisions of criminal law are to be carried out in light of the affected fundamental rights,[95] such as the protection of personal honour as part of the general right of personality, freedom of speech and expression and potentially artistic freedom.

For example, in German law, there is no general precedence of one over the other, which makes the determination of a statement's legality – both online and offline – particularly challenging. The German Federal Constitutional Court has underlined the general principle that certain contents of statements, especially regarding political views, shall not be sanctioned.[96] Nonetheless, there are limits to freedom of speech and freedom of the media, such as restrictions inherent in other fundamental rights, especially human dignity.[97] The German constitutional jurisprudence on that matter shows that sufficient consideration of freedom of speech and expression has proven consistently difficult even for judges of the ordinary jurisdiction with the federal constitutional court repeatedly overturning judgements.[98] The ECHR has also consistently stressed the overriding and essential nature of freedom of expression in a democratic society, while at the same time accepting and setting limits in case of incitement to hatred, discrimination and violence.[99] If the application and interpretation of relevant provisions is carried out by the competent state courts, a constitutional review by the competent constitutional authorities is secured. If these tasks are handed off to social networks, the participation of the concerned before the decision on the removal depends on the social networks' good will.

3.2.4 *Social Networks as Rational Market Players*

'Services have a moral duty to fight illegal behaviour online', David Cameron is quoted as stating in the context of child pornography.[100] Surely, the question of moral responsibility of social networks becomes more and more pressing in light of their developing role in society. It is linked to a number of ethical issues regarding both users and network administrators.[101] Notwithstanding that worthy discussion, social networks as private commercial entities do not serve public policy purposes or other altruistic interests. They are not directly bound by fundamental rights and by no means guardians of their protection. Reliance on private entities 'relegates governmental duties to private prejudices, incentives and priorities'.[102]

When evaluating law enforcement strategies, social networks have to be seen as rational market players acting in accordance with their interests. For example, obscene and violent material can negatively impact advertising revenue.[103] The platform operators' principal aim as businesses is economic gain. Hence, the incentive structure created for these economic social networks has to be analysed in order to determine the consequences for fundamental right protection. The premise in this context is that social networks will act to prevent the dissemination of illegal, but not legal content, if this outcome is in line with its own interests like the maximisation of profits and the reduction of risks. This is especially true as most cases of illegal content are not as easily identifiable and not as severe as child pornography. In less severe cases, both the moral scruples and the public relations issues are weaker and so are the incentives to combat such illegal content.

3.2.5 *Conclusions*

With all models delegating the responsibility of legal tests to the private entities that social networks are, the application and interpretation of legal norms is left to them and their agenda, even though this is an essential state task. As has been shown for Germany in an exemplary manner, this application and interpretation in consideration of the fundamental rights at stake is a rather complicated task that regularly leads to the repeal of judgements by constitutional bodies. As the applicable legal sources vary, the decentralised character of worldwide social networks is no advantage.

Social networks certainly have the potential to be technical chokepoints in the fight against hate speech and fake news. Social media policies and terms of use are, however, not effective tools to ensure the legality of UGC. While networks are well placed to implement detection technology and filtering devices, no such software currently exists with regard to hate speech and fake news. The successful technique of digital footprinting can only be used with regard to certain, severe cases. Generally, the determination as illegal presupposes the consideration of the context in the individual case and the affected fundamental rights. The application and interpretation of relevant provisions by social networks therefore hold risks for the exercise of fundamental

95. For Germany *see* Federal Constitutional Court, *NJW* 2943(1994); *NJW* 3303 (1995); D. Grimm, ‚Die Meinungsfreiheit in der Rechtsprechung des – Bundesverfassungsgerichts', *NJW* 1697, at 1701-02 (1995).
96. Federal Constitutional Court, *NJW* 257, 258 f. (1958).
97. Federal Constitutional Court, *NJW* 1303 (2003).
98. Federal Constitutional Court, *NJW* 2022 (2015); *NJW* 2643 (2016); *NJW* 1092 (2017).
99. *Belkacem* v. *Belgium*, Application no. 34367/14, ECHR, 27 June 2017.
100. R. Watts, 'David Cameron: Web Firms Have a "Moral Duty" to Wipe Out Indecent Images', *The Telegraph*, 20 July 2013.
101. M. Turculeţ, 'Ethical Issues Concerning Online Social Networks', 149 *Procedia, Social and Behavioral Sciences* 967 (2014).
102. Tsesis, above n. 79, at 868.
103. Klonick, above n. 88, at 1627.

rights. Their realisation depends on the particular law enforcement model in place and will thus be further examined below with regard to the models compared.[104] The protection of fundamental rights is not a task of private economic entities, but one of the public core responsibilities.

3.3 Public Core Responsibilities

State actors are the guardians of fundamental rights. They are bound by law to respect and safeguard fundamental rights. Hence, they cannot comprehensively delegate this underlying responsibility to private actors, as the enforcement of existing law by the state is the necessary counterpart to the state monopoly on the legitimate use of force. The restriction of the rights of the individuals depends on the state's empowerment with enforcement rights. When the state entrusts private actors with the enforcement of the law, its delegating mechanisms are to be analysed with regard to the fundamental rights ramifications and the effectiveness of enforcement. This is also true when responsibility is assigned to social networks and public policy is implemented by shaping their incentive structure.

4 Comparative Model Analysis against That Background

The spectre of possibilities to safeguard social networks against hate speech and fake news in addition to the traditional law enforcement mechanisms covers many different approaches. It ranges from a laissez-faire approach with user self-censorship all the way to active monitoring obligations of social networks. Naturally, there is a continuum between these different strategies; legal regimes like the ones referenced above[105] can fall anywhere along that continuum. Given the private and public core responsibilities identified above,[106] different models will be juxtaposed and evaluated.

Self-censorship has already been dismissed, as it does not effectively serve the purpose of fighting the dissemination of hate speech and fake news.[107] Keeping in mind the reservations regarding the delegation of law enforcement to social networks,[108] different schemes of network responsibility will be examined. For that purpose, three basic models shall be distinguished, namely a strict liability approach with an obligation to monitor (4.1), a negligence-based liability system with a notice-and-takedown mechanism (4.2) and voluntary commitments of social networks to code of conducts and the like (4.3).

4.1 Obligation to Monitor and Strict Liability

Proactive monitoring obligations are generally and increasingly used to impose a strict liability standard on Internet intermediaries such as social networks.[109] With a strict liability approach, social networks are held responsible for illegal content on their platforms even if they did not have any knowledge of the content concerned. The legal doctrine of strict liability makes a person or company responsible regardless of any negligence or fault on their part. It is conventionally applied when such persons engage in inherently dangerous activities. This can be said with regard to social networks as the business models they profit of favour the creation of (illegal) UGC.[110]

Obligations to monitor the UGC establish such a strict liability regime.[111] Compliance with general monitoring obligations proves tremendously difficult in light of the insufficient technical tools.[112] Smaller social networks and start-ups are pushed out due to the high operating costs related to the shielding against risks, thus cementing the market.[113] Innovation and competition are thus hindered by this strict approach, with economic exchange online not being furthered.[114]

This model of law enforcement in social networks creates strong incentives for social networks to block all potentially illegal content in order to avoid any liability. Content carrying the risk of provoking controversy is thus likely taken down pre-emptively or at the first complaint received. There is no significant economic advantage to hosting debatable UGC. Decisions are therefore not primarily made on the legality of the content. Content will readily be removed or blocked before any court involvement. Individual incentives for interventions vary largely and are not sufficient to safeguard the fundamental rights concerned.[115] There is also no incentive for social networks to carry out factual investigations first. This is all the more significant as illegality of UGC in the context of hate speech and fake news is only rarely evident.[116] Accordingly, monitoring obligations lead to incentives to overblock.[117] For that reason, the OSCE Special Rapporteurs on Freedom of Expression spoke out against the imposition of duties to monitor the legality of the activity taking place within the intermediaries' services.[118]

104. *See* 4.
105. *See* 2.2.-2.7.
106. *See* 3.
107. *See* 3.1.
108. *See* 3.2.
109. B. Kleinschmidt, 'An International Comparison of ISP's Liabilities for Unlawful Third Party Content', 18 *IJLIT* 332, at 346 (2010); P. Baistrocchi, 'Liability of Intermediary Service Providers in the EU Directive on Electronic Commerce', 19 *Santa Clara High Tech. L.J.* 111, at 114 (2002).
110. *See* 3.2.
111. Baistrocchi, above n. 109.
112. *See* 3.2.2.
113. *See* Baistrocchi, above n. 109; J. Hornik and C. Villa, 'An Economic Analysis of Liability of Hosting Services: Uncertainty and Incentives Online', 37 *Bruges European Economic Research Papers* 13 (2017) for all ISPs under the ECD.
114. *Ibid.*
115. *See* 2.2.
116. *See* 3.2.3.
117. *Delfi AS* v. *Estonia*, Application no. 64569/09, ECHR, 16 June 2015, Joint Dissenting Opinion of Judges Sajò and Tsotsoria § I.2.
118. Joint Declaration of the Three Special Rapporteurs for Freedom of Expression (2011) 2.b, available at www.oas.org/en/iachr/expression/showarticle.asp?artID=848.

The danger of overblocking leads to chilling effects for the exercise of fundamental rights.[119] Social networks are deterred from hosting content in legal grey areas, and users are discouraged from exercising their fundamental rights such as free speech on social networks in light of the expected quick removal of controversial content. Free speech and potentially also artistic freedom and freedom of the media are most restricted in strict liability systems with an obligation to monitor. This model represents a case of 'collateral censorship' that occurs when the state holds a private party – the social networks – liable for the speech of another private party – the user generating content – and the first private party also has the power to control access to B's speech.[120]

Swedish law foresees an obligation to monitor.[121] However, the social networks' duty to supervise under Swedish law is considerably relativised. Networks do not have to guarantee that their systems are clean.[122] The proactive duty to check for illegal content is limited to areas where UGC is more likely to occur on the basis of past experiences or context. For other areas, a notification system can be sufficient. The Swedish system thus combines the first model of an obligation to monitor with the second model of a notification system.

4.2 Notice-and-Takedown and Negligence-Based Liability Systems

The second model can be described as conditional safe harbour model. Social networks are protected in the safe harbour as long as they comply with the requirements for dealing with unlawful content on their platforms.

With that model, social networks as such have no general monitoring obligation. Their liability for illegal content disseminated via their facilities is limited. It depends on knowledge of the illegal content in question and compliance with duties to take down that content. Examples for notice-and-takedown and negligence-based liability systems are the ECD and the German NetzDG. According to both the ECD and the NetzDG, social network liability is excluded if upon obtaining knowledge or awareness of illegal content, the social media provider acts expeditiously to remove or to disable access to the information, with the German system defining more clear-cut deadlines than does the EU one.[123] Such systems based on knowledge or notice of illegal content mirror the lack of adequate monitoring software.

The rather nebulous concept of expeditious acting, however, risks blurring the lines of the social network's responsibility. In terms of legal certainty, the German model appears to be favourable at first sight as it clearly stipulates deadlines for the takedown. While it appears sensible to tie these deadlines to the time needed to properly assess the illegality of the content, the gradations according to the obviousness of illegality reintroduce elements of legal uncertainty and unpredictability. As a result of legal uncertainty, it is difficult for social networks to weigh how much to invest in the prevention of the publication of illegal UGC on their networks.[124] Legal uncertainty affects the social network's ability to determine a rational investment and an efficient targeted line of attack.[125]

Notice-and-takedown systems can protect the exercise of fundamental rights inasmuch as they drive social networks to actually test the legality of the content before removing or blocking it. Neither the ECD nor the NetzDG foresee specific mechanisms to ensure the test of legality; the tiered deadlines for removal, however, give room for adequate examination.

The option of involving a private outside institution (agency of regulated self-regulation) provided by the German NetzDG does not guarantee correct rulings on the legality of UGC. Even though the examiners' expertise has to be recognised by the Federal Office of Justice, they are part of a private institution offering their services to social networks. As such, their incentives are approximated to those of their clients. There is thus little to no[126] added value in comparison with mere in-house assessments by skilled jurists. Complaint mechanisms are confined to the agency; there is no integration into ordinary jurisdiction. Court reviews are – as with all decisions taken by social networks – limited to the period after the fact, i.e. the removal.

In contrast to the first model with a general obligation to monitor, the incentives to swiftly remove all questionable content are limited to the notified UGC with notice-and-takedown and negligence-based liability systems. They still entail dangers for fundamental rights with regard to the notified content because they cause incentives to overblock as well as considerable chilling effects.[127] These incentives are enhanced by the threat of considerable fines in the NetzDG. Social networks will readily remove content in order to minimise their risks, especially towards the end of the standardised deadlines. The NetzDG has therefore been described as bold gambit with fundamental rights.[128]

This danger is reduced, but far from banned by the limitation of fines to systematic failure rather than to the non-compliance in individual cases by the NetzDG. Standardised deletion upon call minimises the risks to

119. W. Seltzer, 'Free Speech Unmoored in Copyright's Safe Harbor: Chilling Effects of the DMCA on the First Amendment', *Harv J L &Tech* 171, at 175-6 (2010).
120. Balkin, above n. 13, at 2309.
121. *See* 2.6.
122. Koops, Prins & Hijmans, above n. 8, at 165.
123. *See* 2.3 and 2.5.
124. Relying *inter alia* on deterrence theory, Hornik and Villa, above n. 113, at 6.
125. *Ibid*., at 11.
126. M. Liesching, '§ 3', in Erbs/Kohlhaas/Liesching, *NetzDG* (2018), at § 3 Rec. 23.
127. J. Urban and L. Quilter, 'Efficient Process or "Chilling Effects"? Takedown Notices, Under Section 512 of the Digital Millennium Copyright Act', 22 *Santa Clara Tech. L. J.* 621 (2006).
128. E. Douek, 'Germany's Bold Gambit to Prevent Online Hate Crimes and Fake News Takes Effect', published 31 October 2017, available at https://www.lawfareblog.com/germanys-bold-gambit-prevent-online-hate-crimes-and-fake-news-takes-effect.

be fined or prosecuted.[129] The danger actually manifested itself only ninety-six hours after the NetzDG's entry into force, when Twitter blocked the account of a German satirical magazine. The magazine had parodied a far-right politician whose social media accounts were blocked earlier that week due to anti-Muslim posts.[130] There are no data as to how much legal content has been removed and how much illegal content kept.[131] Consequently, the proportionality of measures like the German notification and fining system is hard to assess because of the lack of (reliable) data. According to press reports, Facebook performed 100,000 deletions in Germany in the month of August 2016 alone.[132] Data pertaining to the removal of copyright infringing content support an over-removal of content by Internet hosting providers under a notice-and-takedown system.[133]

4.3 Voluntary Commitments – Code of Conducts and the Like

Voluntary commitments to comply with a code of conduct appear like paper tigers, especially against strict liability or negligence-based systems with severe fines for non-compliance. It must, however, not be forgotten that every deletion of a legal upload, post or tweet violates freedom of speech and expression and possibly also freedom of the media and other fundamental rights. According to the third evaluation of the EU Hate Speech Code of Conduct, whose results were published in January 2018, the signatory IT companies removed on average 70 per cent of illegal hate speech notified to them by non-governmental organisations (NGOs) and public bodies participating in the evaluation.[134] For that reason, EU Commissioner for Justice, Consumers and Gender Equality Jourová found the code of conduct a valuable tool to tackle illegal content quickly and efficiently.[135] The European Commission expressed its conviction that the code of conduct will not lead to censorship, as it does not oblige the signatory companies to take down content that does not count as illegal hate speech.[136]

Against that background, it needs to be reiterated that none of the models and examples presented obliges social networks to take down legal content. As long as non-compliance with voluntary commitments does not lead to any liability or sanction, there is certainly less incentive to overblock than with strict or negligence-based liability systems. Nonetheless, considerable incentives to delete not only illegal but also legal content exist.

As social networks firstly test the content against their individual 'Rules or Community guidelines' according to the Code of Conduct, restrictions on free speech and other fundamental rights are detached from legal prerequisites. The code does consequently not safeguard existing laws that strive to balance free speech and rights of third parties.[137] If policies are significantly stricter than the applicable state law, free speech is unduly limited by deleting legal, albeit undesirable, statements. With social networks increasingly under fire for hate speech and fake news dissemination on their platforms, there is substantial public pressure to act. They can document their efforts with a media-effective signature of a code of conduct and the publication of the percentage of quickly deleted notified content. The figures published by social networks have been recently taken into account by numerous state actors.[138] They also play a substantial role for the businesses' public image. When the image of a company like Facebook or Twitter suffers, this can easily translate to financial loss.

Voluntary commitments gradually include more and more proactive duties.[139] The ECD principle that there is no general obligation to monitor is called into question by the voluntary frameworks set up at EU level. The Commission explicitly demands proactive monitoring: 'Online platforms should, in light of their central role and capabilities and their associated responsibilities, adopt effective proactive measures to detect and remove illegal content online and not only limit themselves to reacting to notices which they receive.'[140] This imposes de facto monitoring obligations[141] with the corresponding dangers for the exercise of fundamental rights. These duties clearly surpass the scope of a notice-and-takedown system, as they also apply to non-notified content. With regard to notified UGC, an overreliance on trusted flaggers is to be feared. Social networks must not refrain from any legal test in cases of notifications from this group of users and institutions.

The encouragement to proactively deploy filtering devices, for example, by the EU fake news initiative, also holds risks for a lawful application of relevant provisions. Fully automated deletion or suspension of content can be particularly effective and deserves support in circumstances that leave little doubt about the illegality of the material, for example, in cases of child pornography. Filtering without an additional case-by-case review equals deletion without any legal test and is therefore

129. Liesching, above n. 55, at 30.
130. Titanic Magazin, 'Twitter sperrt TITANIC wegen Beatrix-von-Storch-Parodie', 3 January 2018, available at www.titanic-magazin.de/news/twitter-sperrt-titanic-wegen-beatrix-von-storch-parodie-9376/.
131. German Federal Ministry of Justice and Consumer Protection, Answer to Written question from André Hunko, No. 10/19 of 6 October 2016, at 1.
132. Zeit Online, 'Facebook nennt erstmals Zahl entfernter Hasskommentare', 26 September 2016, available at https://www.zeit.de/digital/2016-09/hasskommentare-facebook-heiko-maas-richard-allan.
133. A. Marsoof, 'Notice and Takedown: A Copyright Perspective', 5 *Queen Mary J. of Intell. Prop*. 183 (2015); D. Seng, 'The State of the Discordant Union: An Empirical Analysis of DMCA Takedown Notices', 18 *Va. J. L. & Tech.* 369 (2014).
134. Press release, 'Countering Illegal Hate Speech Online – Commission Initiative Shows Continued Improvement, Further Platforms Join' of 19 January 2018, IP/18/261.
135. *Ibid.*
136. Tweet @EU_Justice, Twitter, 07:11 – 19 January 2018.
137. *See* EDRi. 'EDRi and Access Now Withdraw from the EU Commission IT Forum Discussions', *EDRi*, 16 May 2016.
138. *See* 4.2 and 4.3; for the UK *see* House of Commons, above n. 5, at 13.
139. *See* for the fake news initiative 2.4.2.
140. European Commission, above n. 43, at para. 10.
141. G. Frosio, 'The Death of 'No Monitoring Obligations', *CEIPI Research Paper* No. 2017-15, 1, at 25 (2017).

hazardous to the exercise of fundamental rights.[142] Filtering especially leads to significant chilling effects.[143] While voluntary commitments might be paper tigers with regard to their enforcement against the social network's will, they show their teeth when it comes to the endangerment of fundamental right exercise online.

4.4 Conclusions

Social networks have gained a considerable amount of control in areas with high relevance for the enjoyment of fundamental rights like free speech and right of personality. With regard to different models of social network responsibilities, it has been shown that all three of them harbour risks for the safeguard of fundamental rights concerned, especially for the exercise of free speech. All these models delegate the application and interpretation to private entities to an extent endangering the lawful interpretation and application of the criminal provisions penalising hate speech and fake news.

The German negligence-based liability system cannot be recommended as international policy example because of these dangers flowing from its incentive structure. For the same reason, an obligation of social networks to autonomously monitor the content on their platforms has to be dismissed. Voluntary commitments to code of conducts can help integrate social networks in the fight against fake news and hate speech, but not without creating an – albeit mitigated – incentive to overblock UGC.

5 Shaping a Superior Model for Law Enforcement in Social Networks

So far, public policy makers in Europe have largely reacted to the challenge of regulating UGC on social networks with far-reaching delegations of law enforcement to social networks. Territorial governments should realise their regulating potential (5.1). In light of all the above, a multi-player solution with stronger public engagement is favoured (5.2).

5.1 Regulating Potential of Public Policy Makers

The somewhat extraspatial character of the Internet does not mean that online activities shall remain unencumbered by government regulations. The approach that cyberspace 'exists, in effect everywhere, nowhere in particular and only in the Net'[144] and that the Internet is 'not subject to the same laws of reality as any other electromagnetic process'[145] is outdated. Geographically based governmental authority is not inapplicable because of a certain non-physical nature of 'the Internet'. Transmission of online content occurs through physical processes in specific jurisdictions by means of physical infrastructure and processes. It has effect on 'real people and real places'.[146]

Online activities indeed make jurisdictional limits visible, as online content is generally accessible beyond borders. States have personal jurisdiction over Internet users depending on their situation.[147] In case of cross-border crimes, more than one jurisdiction can apply to a single act. While the applicable private law can be determined by the rules on conflict of laws and international civil jurisdiction is established in accordance with international procedural law, this is a significant challenge in practice. Therefore, a further harmonisation and unification of law and policy both in the area of private international law and in the area of substantive laws on fake news, hate speech and other defamatory and illegal content would greatly benefit effective traditional law enforcement.[148] In that context, the level of (international) regulation has to be chosen with particular regard to legal cultures in the participating states, especially the concept of free speech and its limits.

The development of online communication through social media, which has *inter alia* physical, psychological, and cultural effects,[149] brings about major changes for law enforcement. State actors both on national and EU level have extensively criticised social networks for their failure to effectively address fake news, hate speech and defamatory UGC. In spite of this, the adjustment of the law to such major changes is a governmental task rather than a private one. State actors have to meet the regulatory needs created and (re-)evaluate law enforcement strategies in place with regard to the new challenges and actors. The specificities of social networks cannot justify a comprehensive delegation of law enforcement to social networks. State actors need to ensure that the policies they put into place produce a fair balance of rights of personality and honour and free speech rather than legal vacuums.

5.2 Multi-Player Solutions

Many players are involved in the sculpting of social network environments.[150] A superior model for law enforcement on these platforms must therefore not neglect their roles, above all the power relationship between international social media companies and public policy makers, for now mostly nation state governments. State responsibilities can be extended and assumed in cooperation with social networks, whose business models justify their participation in the costs of

142. *See* 4.2; for copyright S. Kulk and F.J.Z. Borgesius, 'Filtering for Copyright Enforcement in Europe after the Sabam Cases', 34 *EIPR* 791 (2012); E. Psychogiopoulou, 'Copyright Enforcement, Human Rights Protection and the Responsibilities of Internet Service Providers After Scarlet', 38 *EIPR* 552, at 555 (2012).
143. Frosio, above n. 141, at 27.
144. D.R. Johnson and D.G. Post, 'Law and Borders, The Rise of Law in Cyberspace', 48 *Stanford Law Review* 1367, at 1375 (1996); *see also* Barlow, above n. 4.
145. M. Wertheim, *The Pearly Gates of Cyberspace* (1999), at 228.
146. *See* Tsesis, above n. 79, at 864.
147. *But see* Johnson and Post, above n. 144, at 1375.
148. N. Alkiviadou, 'Regulating Internet Hate – A Flying Pig?', 7 *JIPITEC* 216, at 217 (2017).
149. *See* Tsesis, above n. 79, at 864.
150. *See* 3.

combating hate speech and fake news. Such multi-player solutions can combine the advantages of the strategic placement of social networks as points of control, while defending law enforcement and the exercise of fundamental rights as basic state task.

Propositions for a superior model of law enforcement can build upon existing concepts. The German NetzDG system already incorporates external assessors for non-obvious cases of UGC legality. While the NetzDG system relegates them to the role of in-house counsel, it shows that a cooperation with an external assessment body is possible. A similar cooperation could be envisioned as private-public partnership. Decisions on the legality of UGC could then be taken by ordinary judges. They possess the necessary expertise and enjoy independence and impartiality. In contrast to private (outside) institutions, their incentives are detached from the ones influencing social networks to overblock content. They would apply the law of the particular circumstances of the case and the fundamental rights affected; their decisions would be subject to review within the ordinary judicial system as well as constitutional review.

Within that proposed scheme, notifications regarding questionable UGC would thus be forwarded to public institutions responsible for the decisions on the take-down of questionable tweets, uploads and other UGC. A timely evaluation could be guaranteed just like swift judicial rulings are provided in the framework of interim legal protection. The referral to the competent judges can happen just as quickly as an in-house transmission. As well as in other contexts, specialised judges can rule within hours or days on the legality of the content, provided sufficient human resources are in place. Such a state intervention obviously requires the attribution of considerable government resources. Costs for this model of law enforcement would be incurred by the state rather than by social platforms as private entities. However, in light of the benefits drawn from the business models, social network responsibility can also be expressed in financial contributions to such a public-private partnership model. The overall cost for law enforcement in social networks would not change. Both public and private investments are worth making in light of the relevance of both social media in today's society and free speech as well as rights of personality in democratic state systems.

Personal Data, Algorithms and Profiling in the EU: Overcoming the Binary Notion of Personal Data through Quantum Mechanics

Alessandro El Khoury*

Abstract

In this paper I propose to analyse the binary notion of personal data and highlight its limits, in order to propose a different conception of personal data. From a risk regulation perspective, the binary notion of personal data is not particularly fit for purpose, considering that data collection and information flows are tremendously big and complex. As a result, the use of a binary system to determine the applicability of EU data protection law may be a simplistic approach. In an effort of bringing physics and law together, certain principles elaborated within the quantum theory are surprisingly applicable to data protection law, and can be used as guidance to shed light on many of today's data complexities. Lastly, I will discuss the implications and the effects that certain processing operations may have on the possibility of qualifying certain data as personal. In other terms, how the chances to identify certain data as personal is dependent upon the processing operations that a data controller might put in place.

1 Introduction

Personal data is any information related to an identified or identifiable natural person.[1] Sometimes it is obvious which information constitutes personal data; some other times the exercise becomes complex and may lead to unexpected results. The paramount principle upon which EU data protection law is based is the possibility of qualifying certain information as *personal data*. Whenever the piece of information carried by data can be separated from the physical person to whom that information refers, the rules and safeguards stemming from EU data protection law become inapplicable. In this sense, data is conceived as binary: it is either personal or not.

The possibility of identifying, directly or indirectly, a person through a number of pieces of information – individual or combined – highlights the complexities that data protection experts are currently experiencing when dealing with technologies and techniques such as Big Data,[2] Cloud Computing,[3] data mining and collection of information through the Internet of Things (IoT).[4] Devices of all sorts around us are constantly collecting information to provide services, yet not all data collected falls within the category of personal data strictly speaking. This amount of non-personal information can, however, quickly lead to the identification of a physical person and reveal very personal aspects such as political orientation or sexual preferences.

In this article, I propose to analyse the binary notion of personal data and highlight its limits in the current EU General Data Protection Regulation (GDPR).[5] *Breyer* v. *Deutschland*[6] shows that from a risk-regulation perspective, the binary notion is not particularly fit for purpose, considering that data collection and information flows are complex processes. This calls for a different conception of personal data, which should go beyond its binary definition, and instead, focus on its inherent, relative nature. Data could indeed be personal and non-personal at the same time: the relevant distinction can be made only in a specific moment, while putting the data in the context of processing operations carried out around it. This article, therefore, purports to show that the use of a binary system to determine the applicability of EU data protection law may be too simplistic an approach.

For this purpose, it employs quantum mechanics as a guide to shed more light on the matter. At the beginning of 1900, when certain observations on matter could not be described through classical physics, physicists began

* Alessandro El Khoury, LLM, Legal and Policy Officer, DG Health & Food Safety, European Commission. The information and views set out in this article are those of the author and do not necessarily reflect the official opinion of the European Commission.

1. The definition we adopt is based on EU data protection law. *See* after the third section.
2. Big Data has been defined as a data set whose size is beyond the ability of typical database software tools to capture, store, manage and analyse. *See* J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh & A.H. Byers. *Big Data: The Next Frontier for Innovation, Competition, and Productivity* (2011).
3. Cloud Computing has to be understood as a methodology through which a vast measure of pooled and virtualised resources can be accessed. *See* A. El Khoury, 'Data Protection and Risk Regulation. Cloud Computing: A Case Study' (LLM thesis on file at LUISS School of Governance, Rome).
4. With the term 'Internet of Things', we refer to a global network infrastructure linking uniquely identified physical and virtual objects, things and devices through the exploitation of data capture, communication and actuation capabilities. *See* A. Guimarães Pereira, A. Benessia and P. Curvelo, *Agency in the Internet of Things*, Publications Office of the European Union (2013), at 7.
5. European Parliament and Council Regulation 2016/679, OJ 2016 L 119/1.
6. Case C-582/14, *Patrick Breyer* v. *Bundesrepublik Deutschland*, ECLI:EU:C:2016:779.

doi: 10.5553/ELR.000114 - ELR december 2018 | No. 3

thinking differently, and quantum mechanics arose as a new branch of physics. Interestingly, it seems that certain principles elaborated within quantum theory may be appropriate for describing data. By drawing inspiration from quantum mechanics, this article aims to ultimately overcome the binary notion of personal data and find a right balance in the application of EU data protection law. This conclusion is also supported by the fact that today it has become rather easy to identify a data subject due to the increasing affordability of certain processing operations and the tools to perform them.

2 Setting the Scene: We Live in a World of Data

The idea behind the quote 'Data is the new oil'[7] is elementary: oil was – and most likely still is – the basis of the world's economy during the twentieth century. Refined to produce plastics, fuel and many other materials, oil can be converted into many different commodities. A legitimate question would be, what does oil have to do with data? Both commodities – oil and data – can be traded and their trade volumes and prices can affect stock markets in different ways. It was demonstrated that changes in oil prices could predict stock market return worldwide,[8] whereas the impact that data can have on stock markets is tied to the reliability that companies feeding off the data can project on the general public.[9]

Moreover, oil is not a self-sufficient commodity: once refined and transformed its sub-products cannot be reverted to oil. This concept was well summarised by Scaruffi:

> [T]he difference between oil and data is that the product of oil does not generate more oil (unfortunately), whereas the product of data (self-driving cars, drones, wearables, etc.) will generate more data (where do you normally drive, how fast/well to drive, who is with you etc.).[10]

Differently from oil, not all data is equal. In this sense, data is more comparable to rocks: there are common and inexpensive, and rare and expensive ones. When a piece of information refers to a human being, it becomes personal data. Not all personal data has the same economic value: there are different values, different pieces of information linked to that data which can make it more or less attractive for business operators according to the type of business they are running.[11] For an advertisement company, geographical data on potential customers is valuable: the company might use that information to target its advertisements and promptly show offers from restaurants to nearby potential customers. This geographical data (technically called 'geotag') needs to be placed in the context of activities that a potential customer is carrying out in a determined time and space. Knowing the potential customer is located close to a restaurant whose advertisement can be shown by the advertisement company is valuable data. If the potential customer is hiking in a forest, however, knowing his specific location does not bring any advertising potential, because there are no restaurants nearby to advertise.[12]

Data is to be understood in broad terms, and according to Ackoff, is raw and does not have a meaning in itself.[13] In the case of geographical data, latitude and longitude are just numbers, coordinates on a map; when matched with a physical person, they become a geotag, an information conveying that a person is physically located somewhere. Therefore, information is data that has been given a meaning by way of relational connection with other data.[14] The meaningfulness of this information has a different degree of appreciation for the subject making use of it.

Another difference between data and oil is that the latter is a scarce resource, whereas the former is virtually infinite, self-sustainable and self-replicable. To understand these concepts we can imagine a timeline, a sequence of events starting at *time 0* and ending at *time 10*. The actual length between *0* and *10* is not relevant. A barrel of oil will always be a barrel of oil throughout the timeline, or until it is transformed into something else. On the contrary, data and the information held within it changes according to its intended use and with time . For example, a person's name is likely to remain unchanged, but if we consider body temperature, its variation throughout a timeline might reveal other pieces of information, such as that the person has a cold or is performing physical activity. This different degrees of information provided by data allows for an interesting observation: data has both an intrinsic and extrinsic value. The intrinsic value is by virtue of the piece of

7. The quote is often attributed to different people. *See* M. Kuneva, European Consumer Commissioner in a 2006's Speech http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm (last visited 24 June 2018); G. Rometty, IBM CEO in a Speech to the Council of Foreign Relations in 2013 https://siliconangle.com/blog/2013/03/11/ibms-ceo-says-big-data-is-like-oil-enterprises-need-help-extracting-the-value/(last visited 24 June 2018).
8. G. Driesprong, B. Jacobsen & B. Maat, 'Striking Oil: Another Puzzle?', 89 *Journal of Financial Economics* 307 (2008).
9. *See* 'Facebook Stock is in the Red for the Year After the FTC Confirms Investigation', http://fortune.com/2018/03/26/facebook-stock-ftc-investigation-cambridge-analytica/(last visited 24 June 2018).
10. P. Scaruffi, *Humankind 2.0* (2016), available at https://www.scaruffi.com/singular/bigdata.html (last visited 18 November 2018).
11. *See also* I.N. Cofone and A.Z. Robertson, 'Privacy Harms', 69 *Hastings Law Journal* 1039, at 1049-1053 (2018) where the concept of the Privacy Bell is discussed. Despite the authors refer to privacy, and not to data protection, the same theory could be used to describe the degree of information on a data subject that data could provide.
12. For an in-depth analysis on the use of geotags and Big Data, *see* J.W. Crampton, M. Graham, A. Poorthuis, T. Shelton, M. Stephens, M.W. Wilson & M. Zook, 'Beyond the Geotag: Situating 'Big Data' and Leveraging the Potential of the Geoweb', 40 *Cartography and Geographic Information Science* 130 (2013).
13. R.L. Ackoff, 'From Data to Wisdom', 16 *Journal of Applied Systems Analysis* 3 (1989).
14. *Ibid.*

information carried by the data. In the previous example, it would be the fact that the body has a certain temperature in a specific moment. When that information is, however, put in correlation with the same data from a different moment of the timeline, it allows inferring a new information.[15]

For companies it makes sense to collect, aggregate and analyse any kind of data, even the one that, prima facie, does not seem to identify a person or highlight a pattern. The reason is that this data could prove useful when put in correlation with other data sets: it could show trends and correlations in those data sets where people are identified, thus transforming into personal data the first data set as well.[16]

So far, this article has focused on the more theoretical aspects of data and information. Now it is time to apply those aspects to concrete cases. The world these days is populated by smart devices capable of collecting and sharing any type of data, by IoT, Cloud Computing and Big Data, which are at the basis of services not even imaginable few years ago. All these technologies are the equivalent of the tools used to extract and refine oil. Tiny sensors collect data, which is shared and processed in the Cloud and ultimately stored in Big Data. Cloud, Big Data and IoT are three different perspectives of complex data processings: IoT *gathers,* Cloud *processes* and Big Data *stores* data. This picture portrays data as a commodity – the fuel running a complex mechanism of systems. Thus, the fundamental question is, how is this commodity regulated? According to EU law, data as such is not regulated, but it becomes strictly regulated when it can be qualified as personal. This leads to another question: are the boundaries of personal data and non-personal data so well defined to justify such a binary approach to data regulation?[17]

3 The Current Notion of Personal Data: From the GDPR to the Case Law of the ECJ

The definition of personal data in Article 4 of the GDPR[18] largely draws from and overlaps with the old definition enshrined in Article 2 of Directive 95/46/CE[19] (Data Protection Directive, or DPD hereinafter), which the GDPR aimed at replacing and updating. The main difference between the two is in the use of the concept of *identifier*: it is used implicitly in the GDPR and explicitly in the DPD. Identifiers are not defined in the GDPR, but they have to be understood as a piece of information holding a particularly privileged and close relationship with the data subject, such as cookies or internet protocol addresses.[20] Recital 30 of the GDPR explains that identifiers are important as they may leave traces of the data subject in a particular environment, which, once combined with other identifiers and information, may be used to create profiles of the data subjects and to identify them.[21]

3.1 The Practical Issues of Identifiers

Some of the issues revolving around identifiers could be understood by analysing the *Cambridge Analytica* scandal, which called into question how Facebook collects and shares personal data.[22] After the scandal became public, Mark Zuckerberg (CEO of Facebook) was summoned before the United States Congress and the European Parliament to answer on how and when Facebook collects and shares data. A recurrent question concerned the so-called *shadow profiles.*[23]

15. Cofone and Robertson, above n. 11. The Privacy Bell shows mathematically how the degree of privacy changes according to the degree of plausible assumptions that can be made on a person: more plausible assumption, less privacy. The same concept is applicable to data protection.
16. This shows why, in academia, some researchers call the debate between anonymous data and personal data a false debate. *See* S. Stalla-Bourdillon and A. Knight, '*Anonymous data* v. *personal data* – a False Debate: an EU Perspective on Anonymization, Pseudonymization and Personal Data', 34 *Wisconsin International Law Journal* 284 (2017) and S.Y. Esayas, 'The Role of Anonymisation and Pseudonymisation Under the EU Data Privacy Rules: Beyond the 'All Or Nothing' Approach', 6 *European Journal of Law and Technology* 1 (2015).
17. Some courts in the United States shared the same perplexity. *See Sanders* v. *ABC*, 978 P.2d 67 (Cal. 1999), where 'privacy, for purposes of the intrusion tort, is not a binary, all-or-nothing characteristic. There are degrees and nuances to societal recognition of our expectations of privacy'. If the case deals with privacy, the same reasoning is valid for data protection if we consider that the presence or absence of privacy is logically linked to the fact that data is personal or not, although the right to privacy and the right to data protection have fundamental differences in their scopes and limitations. *See e.g.* Case C-28/08 *P, Commission/Bavarian Lager*, [2010] ECR I-6055, para. 60, and J. Kokott and C. Sobotta, 'The Distinction Between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR', 3 *International Data Privacy Law* 222 (2013).
18. Personal data is defined as 'any information relating to an identified or identifiable natural person ("data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person', Art. 4, GDPR, above n. 5.
19. Art. 2(a) of Directive 95/46 defines personal data as 'any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity'. European Parliament and Council Directive 95/46/CE, OJ 1995 L 281/31.
20. Compare with Art. 29 Data Protection Working Party, *Opinion 4/2007 on the concept of personal data* (2007) at 14.
21. *See also* R.E. Leenes, 'Do You Know Me? Decomposing Identifiability', *Tilburg University Legal Studies Working Paper No. 001/2008*, where the identifiability is divided in four subcategories: L-, R-, C- and S-identifiability. L-identifiability allows individuals to be targeted in the real world on the basis of the identifier, whereas this is not the case for the other three. In fact, R-identifiability can be further decomposed into the S-type, which is a technical kludge, and C-type, which relates to the classification of individuals as members of some set.
22. The *Cambridge Analytica* scandal concerned the collection of personal data of around 84 million Facebook users by British political consulting Cambridge, which used it to steer the US presidential elections of 2017.
23. Shadow profiles are an aggregation of information concerning a particular data subject who has not yet been formally identified. *See,* in par-

Shadow profiles are based off a basic function of the Internet: most websites collect information on visitors to tailor services such as advertisements or store users' preferences to provide a better browsing experience. Facebook's peculiarity is that whenever one of its features (as simple as a *like and share* button) is embedded in a website, it sends data about its use to Facebook, even if the user's activity on the webpage was only limited to browsing.[24] This data is full of identifiers such as cookies, Internet Protocol (IP) addresses and many others.[25] Facebook counts around 2.2 billion users monthly,[26] and it is not difficult to understand why most websites today embed features from it, thus allowing a large collection of identifiers. A sufficient amount of identifiers can be used to infer information about a virtually unknown person (technically, a not-yet-identified data subject).

What is the use of this aggregated data? In the case of Facebook, when a person registers to it, the platform associates the shadow profile with that person. Without shadow profiles, the database containing personal data of a new user should be empty. Any collection of personal data should begin only at the moment of registration and, in any case, after the user has given explicit consent to it. However, when shadow profiles are used, correlations are done automatically by Facebook, and the already performed data collection and analysis are associated with that data subject. In turn, the platform can immediately offer enhanced services such as suggesting a friend list or displaying advertisements of interest for that user in a surprisingly (or worryingly) accurate fashion.

Identifiers as such do not do have to be understood as personal data: they hold a privileged relationship with the data subject because they can describe certain of his characteristics.[27] Yet, they have the potential to become personal data, at later stages. All the more so when an identifier not conceived to collect personal data could be re-engineered into an identifier carrying a high degree of personal identifiability.[28]

The practical issue of identifiers and personal data has been presented to better understand the impact of the reasoning followed by the European Court of Justice.

3.2 *Breyer* v. *Bundesrepublik Deutschland*

In the landmark judgement delivered on 19 October 2016 in *Patrick Breyer* v. *Bundesrepublik Deutschland*[29] (*Breyer* hereinafter), the Court of Justice of the European Union (ECJ hereinafter) determined that dynamic IP addresses constitute personal data in relation to a certain provider, where it has the legal means that would enable it to identify the data subject through additional data held by another provider.

The case originated from a request for preliminary ruling from the German Federal Court of Justice (FCJ), in relation to an action brought by Mr. Breyer – a former member of the parliament in Schleswig-Holstein – against the Federal Republic of Germany, concerning the registration and storage by the latter of the IP address allocated to him, alongside the date when he accessed several websites run by German federal institutions, the terms entered in the search fields and the quantity of data transferred. Data retained by the German federal institution, no matter how specific, did not allow the identification of Mr. Breyer, thus falling outside the notion of personal data and the protection of the DPD. However, such identification would have been possible if the Internet Service Provider (ISP) had revealed sufficient information to identify the person operating behind a dynamic IP address.[30]

Whether static IP addresses should be considered personal data or not was already answered by the ECJ in 2011. In *Scarlet Extended*,[31] the ECJ concluded that static IP addresses should be considered personal data because they allow the precise identification of the user.[32] According to the Court, there are two elements to consider: one technical and one legal. Technically, the ISP assigns an IP address to a device, and this IP is always the same (static IP); legally, the underlying contract for the Internet service provisioning will be undertaken between the ISP and a natural or legal person, under whose responsibility the connected device is operated. This is why in *Scarlet Extended* the ECJ based its conclusions on the fact that an injunction by a court

ticular, the question asked by New Mexico Representative Ben Lujan (full transcript available at https://techcrunch.com/2018/04/11/facebook-shadow-profiles-hearing-lujan-zuckerberg/?guccounter=1 (last visited 7 July 2018) and by MEP Syed Kamall (*see* Facebook's written answers available at http://www.europarl.europa.eu/resources/library/media/20180524RES04208/20180524RES04208.pdf> (last visited 7 July 2018).

24. *Ibid*.
25. *See* M.D. Ayenson, D.J. Wambach, A. Soltani, N. Good, and C.J. Hoofnagle, 'Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning', Available at SSRN: https://ssrn.com/abstract=1898390; D. Barth-Jones, 'The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now', Available at SSRN: https://ssrn.com/abstract=2076397 and F.J. Zuiderveen Borgesius, 'Singling Out People Without Knowing Their Names – Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation', 32 *Computer Law & Security Review* 256 (2016).
26. Statistics are referred to the second quarter of 2018, https://www.forbes.com/sites/dantedisparte/2018/07/28/facebook-and-the-tyranny-of-monthly-active-users/#383c9c8f6aea (last visited 4 November 2018).
27. It is the case for keystroke dynamics applied for personal authentication, which relies on the fundamental assumption that keystroke dynamics (i.e. how a certain person types on a keyboard) is almost unique for each person. *See* G. Gabla, 'Applying Keystroke Dynamics for Personal Authentication' Available at SSRN: https://ssrn.com/abstract=2508480.
28. *See* Barth-Jones, above n. 25; and Zuiderveen Borgesius, above n. 25.
29. *Breyer*, above n. 6. *Breyer* was ruled under the DPD. Differences with the GDPR will be marked throughout the analysis.
30. An IP address is a logical numeric address assigned to every device connected to a network to identify it. These addresses are assigned by an ISP to a host in a fixed or dynamic fashion. In the former case, a device will always use the same IP address, whereas in the latter case, the IP address is assigned each time the device connects to the network. IP addresses exist to identify a specific device, but they are not necessarily meant to identify the person operating it in a given moment, all the more so when the IP address is a dynamic one. *See* S. Feit, *TCP/IP: Architecture, Protocols, and Implementation with IPv6 and IP Security* (1996).
31. Case C-70/10, *Scarlet Extended*, ECLI:EU:C:2011:77.
32. *Ibid*. para. 51.

to an ISP to install technical means to analyse the traffic generated by a certain IP address, in order to monitor the use of peer-to-peer software[33] used to infringe intellectual property rights, was precluded by EU data protection law.[34]

If the nature and function of static IP addresses are clear, dynamic IP addresses are trickier. The difference in the identifiability features of static and dynamic IP addresses can be understood through an example: if we picture IP addresses as coats of different colours used to identify doctors in a hospital, using a static IP means that each doctor will always wear the same coat; on the contrary, a dynamic IP address entails that each time a doctor enters the hospital's premises, he will be assigned one random coat from the ones available. This latter concept is known in information technology as Dynamic Host Configuration Protocol (DHCP), and it prevents two devices from being assigned the same IP address and thus causing a conflict in the network architecture. That coat alone, however, does not bring sufficient information to enable an identification of the doctor wearing it: it holds a privileged relationship with the data subject, but alone it does not allow its identification.

The DPD enshrines in Recital 26 a key principle to ascertain whether an identifier actually allows for the identification of a data subject: the means likely reasonably to be used by the controller or by any other person to identify the data subject.[35] The GDPR provides the same principle in a same-numbered recital,[36] but it adds to it that to consider those means as 'likely reasonably to be used', account should be given of all objective factors, such as costs and the amount of time required for the identification, taking into consideration the available technology at the time of the processing and the technological developments. In a nutshell: *feasibility* and *capability*. A technical means is likely reasonably to be used according to the *feasibility* of its use and the *capability* of a data controller to use it, which include technical implementation, time and the costs and benefits of doing so. Technical implementation, economic cost and time need to be put in relation to the potential economic benefit of the operation for the data controller.

The concept of *means likely reasonably to be used* generated a large debate in German academia, which polarised around a *subjective* and an *objective* criterion.[37] According to the *objective criterion*, a person can be identified when, regardless of the capability of a certain data controller to identify him, the identification is *feasible* by combining data from different sources. The *subjective criterion* relies on the concrete *capability* of a certain data controller to make use of its means to identify the data subject. The main difference between the two criteria lies in the relevance given to the data controller. For instance, the sheer size of means available might make all the difference in understanding whether certain data is personal or not for that specific controller. Relativity at its best!

In *Breyer*, the two criteria applied as follows: for the *subjective criterion*, IP addresses become personal data only when there is the concrete capacity of a provider who has access to that information to use his own resources to identify the data subject (*e.g.* by performing more correlation with other data sets or even collecting additional data); on less theoretical grounds, by applying the *objective criterion,* IP addresses become personal data only when a data subject can be concretely identified, regardless of the abilities and the means of a provider to do so.[38]

The choice between the two criteria has a fundamental meaning when dealing with dynamic IP addresses. In that case, the means *likely reasonably to be used* to identify the data subject are allegedly more complex, expensive and time consuming to implement. Thus, if theoretically an identification is possible (subjective criterion), it does not mean that this could happen in practice (objective criterion). The question referred to the ECJ by the German FCJ, however, has a remarkable subjective element. What the FCJ fundamentally asks is if a dynamic IP address stored by an online media provider (*i.e.* the owner of a website) has to be considered already personal data for that provider, in the case where only a third party has the additional information necessary to identify the data subject[39] which accessed the online media through that dynamic IP address in a specific moment in time.[40]

The groundbreaking element of *Breyer* does not consist in the ECJ's ruling that, under certain conditions, dynamic IP addresses are personal data, but rather in the legal reasoning followed to reach those conclusions – that same reasoning is applicable *mutatis mutandis* to similar categories of data and subjects. This reasoning is based on three key elements.

33. Peer-to-peer networking is a distributed computing architecture allowing the partitioning of tasks between different devices (peers) connected to a network, thus allowing a substantial degree of anonymity when sharing files of considerable size. *See also* R. Ambrosek, *Shawn Fanning: the Founder of Napster* (2006) where the facts behind the very first peer-to-peer software called 'Napster' are re-construed.
34. Notably, the ECJ ruled, 'Directives 2000/31, 2001/29, 2004/48, 95/46 and 2002/58, read together and construed in the light of the requirements stemming from the protection of the applicable fundamental rights, must be interpreted as precluding an injunction made against an ISP which requires it to install the contested filtering system'. *Scarlet Extended*, above n. 31, para. 55.
35. Recital 26, DPD, above n. 19.
36. Recital 26, GDPR, above n. 5.
37. *See* M. Schreibauer, '§ 11 Telemediengesetz (4 to 10)', in M. Esser, P. Kramer & K. von Lewinski (eds.), *Kommentar zum Bundesdatenschutzgesetz. Nebengesetze* (2014); J. Nink and J. Pohle. 'Die Bestimmbarkeit des Personenbezugs. Von der IP-Adresse zum Anwendungsbereich der Datenschutzgesetze', in *Multimedia und Recht* (9/2015), at 563-67. J. Heidrich and C. Wegener, 'Rechtliche und technische Anforderungen an die Protokollierung von IT-Daten. Problemfall Logging', 8 *Multimedia und Recht* 487 (2015). H. Leisterer, 'Die neuen Pflichten zur Netz– und Informationssicherheit und die Verarbeitung personenbezogener Daten zur Gefahrenabwehr', 10 *Computer und Recht* 665 (2015).
38. See *Breyer*, above n. 6, paras. 52-54.
39. *Breyer*, above n. 6, para. 31.
40. Notably, dynamic IP addresses change at every connection; thus, the reasoning has to be strictly bound to the possibility of identifying a data subject in a specific moment of a timeline.

First, from a technical perspective, dynamic IP addresses belong to the general category of metadata[41]: metadata is not personal data, but may contain data about personal data. For instance, typical metadata applied to personal data would be the date when the personal data *surname* has been changed in a system. If we strictly apply the binary approach adopted in the DPD or in the GDPR, metadata stays outside the protection provided by EU data protection law, meaning that every processing operation on that metadata is possible, including transfers outside the EU and recombination with other data.

The second element stays in the nature of power used by the German federal institutions to obtain information from the ISP. Public entities can act either in their public capacity, representing the public interest (*cum imperio*), or as any other legal entity (*sine imperio*).[42] When acting *cum imperio*, a public administration does not act as a peer towards its counterparts – it exercises a public power with an outreach not possible for private operators – whereas, acting *sine imperio* does not entail an exercise of public power and the outreach is the same as any other private operator. In *Breyer*, the German federal institution acted *sine imperio*.[43]

The third element concerns the outreach of an action *sine imperio,* which, according to the ECJ, consists of any possible channel not prohibited by law to achieve the desired result.[44] These channels could be, for instance, contractual clauses foreseeing the trading of metadata between two entities acting *sine imperio* one against the other.[45] Such clauses could be very easily inserted in a service provisioning agreement between different service providers in a contract for Cloud Computing services and,[46] concerning mere metadata, none of the guarantees foreseen by EU data protection law could prevent such trading.[47]

The three aforementioned key elements have to be tested within the framework of 'means likely reasonably to be used' provided by Recital 26 of the DPD and GDPR. Earlier we used the terms *feasibility* and *capability*, but what the ECJ concluded in a much more complex manner is that the possibility for a data controller to obtain further data from a third party to identify a data subject has to be understood within its capability to do so. In fact, 'that would not be the case if the identification of the data subject was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort'.[48]

Proportionality is another element that has to be accounted for. It entails at least two sub-elements: an effort and a subject performing it. Lifting a hundred kilograms is a remarkable effort for a human, but is a negligible effort for a crane. On those same lines, imposing a certain contractual clause where metadata has to be transferred to a data controller might be a negligible effort, if that data controller is someone the size of Google or Facebook.[49] Moreover, in the proportionality check, a significant role is also played by the reward that those efforts bring.

The conclusion of *Breyer* is that dynamic IP addresses are not personal data per se, but they can become so for a data controller if it has lawful means to obtain any further data that would allow the identification of the data subject. The same reasoning is applicable to any kind of metadata, which brings two questions: I any data potentially personal data? Is the binary notion of personal data adequate to respond to the challenges posed by the complex world of Big Data?

4 Big Data, Anonymisation, Pseudonymisation and Data Analysis

Breyer shows how data is subject to a double relativity. One relativity aspect concerns the very nature of the data (personal or not) against the means that a controller can put in place to reconstruct that data as personal; in this case, the controller performs an identification. The other relativity (hence double relativity) concerns the effort needed to reconstruct non-personal data as personal, which is not relative to the means used, but to the data controller performing it and to its capacity to do so.

To put it in different words, at the beginning of this article I used the example of the timeline, from *time 0* to *time 10*. What *Breyer* shows is that non-personal data located at *time 0* could become personal data in another moment of the timeline, depending on the subjects having access directly or indirectly to it. Moreover, the possibility of non-personal data to mutate its nature depends on the theoretical means that a controller can potentially put in place to do so (if we opt for the subjective criterion), or the means that it actually puts in place, only when it makes use of them (if we opt for the objective criterion).

To add another layer of complexity to this reasoning, we should also take into account the issue of data anonymisation. This practice has been described as the process

41. Metadata has to be understood as data about other data. *See* J. Pomerantz, *Metadata* (2015), at 16.
42. *See* E. Casetta, *Manuale di Diritto Amministrativo* (2008), at 300.
43. M. Reimann and R. Zimmermann, *The Oxford Handbook of Comparative Law* (2007), at 1274.
44. *Breyer*, above n. 6, para. 47.
45. C.J. Hoofnagle, 'Big Brother's Little Helpers: How Choicepoint and Other Commercial Data Brokers Collect, Process, and Package Your Data for Law Enforcement', 29 *N.C.J. Int'l L. & Com. Reg* 595 (2003).
46. *See* C. Reed, 'Information "Ownership" in the Cloud', *Queen Mary School of Law Legal Studies Research Paper No. 45/2010*.
47. Which explains why the data processing put in place by Facebook to perform shadow profiling, despite being despicable, is perfectly compatible with EU data protection rules.
48. *Breyer*, above n. 6, para. 46.
49. *See* S. Bradshaw, C. Millard & I. Walden, 'Contracts for Clouds: Comparison and Analysis of the Terms and Conditions of Cloud Computing Services', 19 *International Journal of Law and Information Technology* 187 (2011) where the authors refer to Terms and Conditions offered by Cloud computing providers in business-to-business contracts.

through which a data controller manipulates data sets in a database in order to make it difficult to identify data subjects.[50] Data anonymisation is also often referred to as 'de-identification'.[51] There are several techniques through which data anonymisation can be achieved, and the difference lies in the cost, complexity, ease of use and robustness.[52] In this sense, we can apply the same proportionality check described for the transformation of metadata in personal data: there will be an initial effort to anonymise personal data, and the anonymisation will be as strong as the effort put in place by the data controller to anonymise that data. Therefore, the robustness of an anonymisation processing is directly proportional to the effort put in place by the data controller, which is also logically impacted by three factors: the degree of robustness that the data controller wants to achieve for those categories of personal data subject to anonymisation; the means likely reasonably to be used to that end and the costs and benefits balance of the anonymisation processing.

Today, the possibility of using virtually unlimited computing power resources, thanks to Cloud Computing[53] and accessing data from tremendously big databases called Big Data, is not reserved for big corporations or governments. The very basis of Cloud Computing is its capability of providing enterprise-like services for any kind of user who can afford the price: the more powerful the service, the higher the price.[54] Data anonymisation is surely a privacy-enhancing technology, but it is also a threatening technology for data protection due to the binary notion of personal data and the so-called accretion problem.[55] The accretion problem postulates that once an adversary has linked two anonymised databases together, he can add the newly linked data to his collection of outside information and use it to help unlock other anonymised databases.[56] Theoretically, the risk increases exponentially for each further database correlated and, as we emphasised earlier, data protection rules are applicable only as long as we are dealing with personal data. If the *personal* element disappears, there are no safeguards for that data. *Pas gráve*, one may argue: if anonymised information suffers a data breach, nobody's rights to data protection or privacy will be violated. From a logical perspective, this is true. The amount of data and metadata present in Big Data, and the simplicity with which they can be computed in a Cloud system by anyone, however, poses a serious risk of reidentification.[57]

There is another interesting debate about anonymisation, and it concerns the 'pseudonymisation' technique. Pseudonymisation involves substituting the real identifying information with a code number or a nickname. Article 29 Data Protection Working Party has described it as 'the process of distinguishing identities'. Such a process aims at collecting additional data related to the same individual without having to know his identity.[58] The problem with pseudonymisation is that it gives the false hope of creating a safe harbour from data protection obligations,[59] thus legitimating high-risk processing operations (such as profiling) under the impression that any claim for damages of unlawful processing could be prevented.[60] Also, the GDPR in Article 6(4)(e) provides that pseudonymisation is an appropriate safeguard,[61] at the same level as encryption.[62] In reality, the means *likely reasonably to be used* are becoming more and more affordable and common thanks to the technologies described earlier. Thus, a correct risk assessment should conclude that re-identification of a data subject is more likely to happen than to retaining a permanent de-identification (or pseudonymisation).

It has been argued that current anonymisation techniques do not favour the data subject's right to self-determination, meaning that the degree of freedom that a data subject can exercise on its personal data is very limited. For instance, when personal data is anonymised, a data subject is faced with difficulty already at the stage of identifying that personal data is being processed. Thus, the data subject cannot verify whether its

50. P. Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', 57 *UCLA Law Review* 1701, at 1707 (2010).
51. S. Latanya, 'Weaving Technology and Policy Together to Maintain Confidentiality', 25 *Journal of Law, Medicine & Ethics* 98, at 100 (1997): 'The term anonymous implies that the data cannot be manipulated or linked to identify an individual'.
52. *See*, for instance, the basic guides to data anonymisation published by the Personal Data Protection Commission of Singapore, *Guide to Basic Data Anonymisation Techniques* (2018); and the European Data Protection Supervisor, *Opinion 3/2018 – EDPS Opinion on Online Manipulation and Personal Data* (2018).
53. S. Chen, H. Lee & K. Moinzadeh, 'Pricing Schemes in Cloud Computing: Utilization-Based versus Reservation-Based', *Production and Operations Management* (2018).
54. For a more detailed overview of Cloud contracts *see* Bradshaw, Millard & Walden, above n. 49.
55. A. Narayanan and V. Shmatikov. 'Robust de-anonymization of large sparse datasets', 111 *IEEE Symposium on Security and Privacy* (2008).
56. *See e.g.* B. Krishnamurthy and C.E. Wills, 'On the Leakage of Personally Identifiable Information Via Online Social Networks', 7 *WOSN '09 Proceedings of the 2nd ACM workshop on online social networks* (2009).
57. *See e.g.* D'Acquisto, J. Domingo-Ferrer, P. Kikiras, V. Torra, Y.A. de Montjoye & A. Bourka, 'Privacy by Design in Big Data: An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics', *ENISA: European Union Agency for Network and Information Security* (2015).
58. Art. 29 Data Protection Working Party, above n. 20, at 18.
59. Esayas, above n. 16, at 6-8.
60. It is the case of the shadow profiling operations performed by Facebook through the placement of cookies, which was fined an incremental penalty of 250,000 EUR per calendar day of non-compliance by the Court of First Instance of Brussels in a judgment of 16 February 2016. See also the joint declaration of the French, Spanish, Belgian and Dutch Data Protection Authorities of 4th December 2015 https://www.cnil.fr/sites/default/files/typo/document/Declaration_commune_Groupe_de_contact_Facebook.pdf (last visited 6 July 2018).
61. The same choice is made in Art. 25(1), where pseudonymisation is presented as a privacy by design measure, Art. 32(1)(a) considering pseudonymisation as adequate safeguard for the security of processing and Art. 40(2)(d), where pseudonymisation becomes a key element of the codes of conducts of enterprises.
62. Encryption can be applied to provide pseudonymisation, but the two processing are logically distinct operations. There is a general understanding that key-coded data may not even be considered personal data so far as there are appropriate measures to exclude re-identification, such as a strong encryption algorithm, a strong encryption key and a secure key. *See* W.K. Hon, C. Millard & I. Walden, 'The Problem of 'Personal Data' in Cloud Computing: What Information is Regulated? – the Cloud of Unknowing', 1 *International Data Privacy Law* 211.

records are getting adequate protection. This kind of dispute is, however, substantially unfounded. Whenever personal information is anonymised, it ceases to be *personal*. Thus, the data subject does not have any legal right over it. It is for this very reason that the real emphasis should be on the moment right before the anonymisation and on the process of anonymisation itself. Once data is anonymised, it can be transferred without boundaries, and as the European Commission stated in 2009, this is not even considered a data transfer in the legal sense.[63] Moreover, other than giving technical advice and guidance on which anonymisation logic exist and what are some of their risks and advantages, and providing examples on their use, public regulatory bodies, such as national data protection authorities, cannot do much more, as anonymisation relies on complex algorithms that are often subject to intellectual property rights.[64]

From a conceptual perspective the distinction between personal and non-personal data is neat; yet, we underlined that this binary approach does not bring a real added value when data protection has to be implemented practically, because the possibilities of identifying a data subject are not the same for every data controller and change according to the circumstances as well. Yet, the legal definition of personal data remains a purely binary one.[65]

If, until now, we were able to substantiate our reasoning without the need to dig into Big Data's technicalities, the next set of issues inevitably demands so. Notably, another set of problems strictly linked to the technical aspects of Big Data – conceptually distinct from data anonymisation and very close to data reidentifiability – are those of data mining and predictive analysis.

Data mining is commonly defined as a set of automated techniques used to extract buried or previously unknown pieces of information from large databases. Data mining makes it possible to unearth patterns and relationships, and then use this new information to make proactive, knowledge-driven business decisions.[66]

From a data protection perspective, data mining is a processing operation and is neutral: the same data mining techniques can be applied to different databases, whether they contain personal data or not. Business operators are increasingly relying on data mining as it allows them to understand the market better and make better decisions.[67] Moreover, thanks to Cloud Computing, the costs of computing services powerful enough to run data mining algorithms are considerably low.[68] The main issues with data mining are that by mining Big Data, the algorithm can find patterns among data sets, thus unveiling further information that was not originally included in those data sets, and de-anonymise personal data that was previously anonymised.[69]

Predictive analysis is a particular type of statistical analysis that can provide, with a certain degree of certainty, answers to certain questions.[70] For instance, by analysing a set of anonymised information, the predictive analysis could tell whether a certain buyer of a product is a man or a woman or if it is a reliable debtor.[71] Once one anonymised information is de-anonymised (remember the accretion problem, and the proportional effort), all the other anonymised information about that (now) identified data subject is immediately correlated to him or her: this is what technically happens behind the curtains of Facebook's shadow profiling.

The relativity of personal data, and the ease with which the virtual border between personal and non-personal data can be disregarded, calls for a different approach, a different conception of personal data – one more attuned with the reality of data processing taking place in today's world – a notion of personal data that draws from quantum mechanics.

5 Overcoming the Notion of Personal Data through Schrödinger's Cat: Quantum Superposition and Quantum Entanglement of Personal Data

Quantum mechanics is a branch of physics developed in the early twentieth century by brilliant minds such as Erwin Schrödinger, Max Planck, Neils Bohr, Albert Einstein and Werner Heisenberg following a series of educated guesses inspired by a thorough knowledge of physics.[72] Quantum theory aimed to describe and explain the behaviour of matter at an atomic and subatomic level, which could not be explained by classical physics, in order to answer very practical questions such as why hot objects glow at a different colour depending

63. European Commission, *Frequently Asked Questions Relating to Transfers of Personal Data from the EU/EEA to Third Countries* (FAQ B.1.9) (2009), available at http://ec.europa.eu/justice/policies/privacy/docs/international_transfers_faq/international_transfers_faq.pdf (last visited 18 November 2018).
64. *See* Art. 29 Data Protection Working Party, *Opinion 5/2014 on Anonymisation Techniques* (2014), at 11 where the analysis revolves around the logic behind certain anonymisation techniques, but it refrains from referring to specific commercial solutions.
65. *See also* Hon, Millard & Walden, above n. 62.
66. A. Cavoukian, *Data Mining: Staking a Claim on Your Privacy* (1998), at 4.
67. J.P. Bigus, *Data Mining with Neural Networks: Solving Business Problems from Application Development to Decision Support* (1996), at 9. I.N. Cofone, Ignacio & A. Robertson, 'Consumer Privacy in a Behavioral World', 69 *Hastings Law Journal* 1471 (2018).
68. P. Ruxandra-Stefania, 'Data Mining in Cloud Computing', 3 *Database Systems Journal* 67 (2012).
69. Art. 29 Data Protection Working Party, above n. 64, at 5.
70. *See* E. Siegel, *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* (2016).
71. On the problem of credit scoring *see* D.K. Citron and F.A. Pasquale, 'The Scored Society: Due Process for Automated Predictions', 89 *Washington Law Review* 1, at 16 (2014).
72. S.M. Barnett, J. Jeffers & J.D. Cresser, 'From Measurements to Quantum Friction', 18 *Journal of Physics: Condensed Matter* S401 (2006).

on their temperature. This article does not claim to redefine quantum physics or enrich its postulations but humbly aims at borrowing specific observations and applying them to data protection law to check whether they could be of help in ultimately providing a more fit-for-purpose definition of personal data.

The basic intuition is that the issue of describing (or measuring) the nature of data as personal or non-personal is very similar to the problems that the illustrious minds behind quantum theory tried to resolve. Bohr wrote, '[A] measurement to a certain degree deprives the information given by a previous measurement of its significance for predicting the future course of the phenomena. Obviously, these facts not only set a limit to the extent of the information obtainable by measurements, but they also set a limit to the meaning which we may attribute to such information'.[73] If we consider a single data in today's interconnected and complex world, its size and velocity of transmission are negligible. Any data (personal or not), regardless of its ability to provide descriptive details of a data subject, is shared between systems at very high speed similar to what happens to protons and electrons in a subatomic system. For the same reason, Bohr concluded that what matters is the unambiguous description of the matter's behaviour, rather than its measurement in a given moment.[74]

Our starting point is the conclusion reached by the ECJ in *Breyer*: data can be personal or non-personal sometimes, according to certain criteria. This emphasises the need to have a notion of personal data capable of providing an unambiguous description, rather than a measurement. The same matter can be better understood through Schrödinger's famous cat experiment.

Schrödinger's cat is a thought experiment imagined in 1935 by the physicist Erwin Schrödinger[75] and used to describe two fundamental principles of quantum mechanics: quantum superposition and quantum entanglement. Specifically, the experiment involves a cat in a sealed box with a bottle of poison, a Geiger counter and a radioactive source. The radioactive source has a 50 per cent chance of decaying. As soon as the Geiger counter detects the decay, a mechanism breaks the bottle of poison in the box, killing the cat. It is not possible to know if the cat is dead or alive before opening the box. Thus, the cat, in the timeline of the experiment, is both dead and alive at the same time. This state of matter is described in quantum mechanics as quantum superposition, and it entails that any two or more quantum states (the cat is dead or alive) can be added together (hence the name superposition) and the result will be another valid quantum state (for the cat, that status would be the cat being dead and alive at the same moment).[76] The main difference with binary systems is that in those, the result can only be true or false, *1* or *0*, but never both together, whereas in quantum mechanics the result can be *1*, *0* or a combination of the two.

Quantum superposition could also be understood through the famous *heads or tail*, where a coin is flipped in the air and the players have to guess on which side the coin is going to land. In a timeline that goes from *0* to *10*, where *0* is the moment just before the coin is flipped and *10* is the moment when the coin lands showing one of the two faces, in any moment between *0* and *10* the coin is potentially showing both *heads* and *tail*.

In our case, the cat or the coin represents data. The fact that the cat is dead or alive or the coin flips on one face or the other represents the fact that data is measured as personal or not. Theoretically, from an observer standpoint, every data not yet identified as personal behaves in the same manner: it is non-personal as long as a data controller does not perform a processing operation suitable of correlating that non-personal data with personal data or an individual, thus converting its nature from non-personal to personal. What puts data in the superposition state is the availability of the *means likely reasonably to be used* by a data controller to identify a data subject from that data. This is why we used the adverb 'theoretically'. Theoretically, we could envisage a set of non-personal data that is kept isolated from any processing operation capable of putting it in correlation with other databases. This is possible either because that non-personal data is collected and stored in a way to be inaccessible or non-compatible with any other data set (thus preventing reidentification) or just because it is swiftly deleted after having achieved its purpose. It was noted, however, that these cases are an exception rather than the rule.[77]

Observing that data is in the quantum superposition state also entails another logical conclusion. Quantum superposition as such is a neutral state: it comprises the case where data becomes personal, but also the opposite, where personal data is anonymised and loses its *identification* properties.[78] This observation is significant for understanding another concept described by quantum mechanics: quantum entanglement.

Quantum entanglement is a very particular quantum mechanical, physical phenomenon[79] in which two particles are so deeply linked that they share the same existence, no matter their physical distance. Once two particles are entangled, even if they are in the superposition status, their measurement will bring the same result.[80] To resume our example of *heads or tails*, if we flip two coins, and these are entangled, any measurement taken during their spin would lead to the same result: the two coins showing the same face. In the case of data, the entanglement consists in the possibility of linking together information from different data sets and pro-

73. A. Plotnitsky, *Niels Bohr and Complementarity. An Introduction* (2012), at 68.
74. J.A. Wheeler and W.H. Zurek, *Quantum Theory and Measurement* (2014), at 5.
75. E. Schrödinger, 'Die gegenwärtige Situation in der Quantenmechanik', 23 *Naturwissenschaften* 807 (1935).
76. P.A.M. Dirac, *The Principles of Quantum Mechanics* (1947), at 1-18.
77. Art. 29 Data Protection Working Party, above n. 64, at 5.
78. Alternatively, pseudonymised with all the *caveats* highlighted before.
79. A. Einstein, B. Podolsky & N. Rosen, 'Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?', 47 *Physics Review* 777 (1935).
80. Wheeler and Zurek, above n. 74, at 422-5.

cessing it in a specific time. At the time of processing, if one data becomes personal, then all the other data from different data sets linked to it exhibit quantum entanglement, and they become personal data too. The bond shared by the different data is their possibility of providing a piece of information sufficient to identify differently the data subject. Then, the fact that this data can be put in correlation provides the entanglement that changes the nature of data that was not personal, in a domino-effect fashion.

Going back to *Breyer*, the collection of dynamic IP address and log files by the online media provider consists of specific data on a particular subject's factual circumstances.[81] The data subject has yet to be identified as Mr. Breyer, and the identification becomes possible only when the ISP reveals information sufficient to achieve the identification. The data held by the ISP and the data held by the online media provider are entangled. They are physically distant, because they are stored in two different systems that are not linked physically or logically to one another; once superposition is triggered by the *means likely reasonably to be used* by the online media provider, the data in the two systems exhibit entanglement and can ultimately be *measured* as personal data. In other words, the entanglement among all the data present in the two data sets allows for an immediate measurement as personal data when a bridge is built between them: this bridge involves the possibility of putting in correlation one non-personal data from a data set with one personal data from another data set. This operation instantly exposes the entanglement (due to the correlations already made within each database), and all data suddenly becomes personal.

Notably, the entanglement – this intangible link or, to use the words of Einstein, this 'spooky action at a distance'[82] – involves the fact that certain data are inherently capable of describing an action, a property or a fact of a data subject. Through this description, data can directly or indirectly contribute to the identification of the data subject. Therefore, the intangible link consists in the fact that all data originates from the same data subject.

6 Quantum Theory and the GDPR

This long and complicated reasoning explained in the previous sections leads to two important conclusions. First, a correct approach to the notion of personal data should aim at providing an unambiguous description of it, rather than a predetermined measurement. In practice, this means taking into account the fact that data is in the quantum superposition state and could exhibit quantum entanglement.

A binary approach fails at grasping these complexities and, above all, fails at describing the true nature of personal data in a world of Big Data and infinite possible processing operations. Quantum superposition and quantum entanglement are a great aid in describing the reality of what can happen to data and personal data when placed in the context of the free-flow of information, and where practically any data controller has access to technical or legal means likely reasonably to be used to achieve the identification of the data subject. Second, rather than measuring the nature of data in a given moment and anchoring to it the applicability of EU data protection law, the focus should be on the processing operations triggering quantum superposition and what surrounds them – meaning that the focus should be on those *means likely reasonably to be used* to transform non-personal data into personal data. The status (personal or not) of data cannot be measured with sufficient certainty or, better, cannot describe the nature of data unambiguously because that status might change in the future depending on the data controller attempting the identification and its available means. If we assume that most data can potentially become personal, from a risk-regulation perspective, it is safer to assume that data is in the superposition status. The focus then shifts on the means used to entangle data and on the safeguards that should apply to those processing operations. In fact, due to those processing operations data exhibits entanglement and can be measured as personal.

On applying quantum superposition to the notion of personal data, the result necessarily moves away from a binary approach and three statuses of data can be observed: personal, non-personal and potentially personal. *Personal data* is data that has already identified (directly or indirectly) a data subject; *non-personal data* is data that does not and cannot (even theoretically) identify a data subject; finally, *potentially personal data* is a residual category, a grey zone, for which identification has not occurred yet, but it has not been excluded either.

If we apply the notion of entanglement to these three new categories, the focus becomes the processing operation that forces data to exhibit quantum entanglement. As a consequence, despite not having measured data as personal in a specific moment, some provisions of the GDPR should be applicable. The rationale behind this consequence lies in the fact that the GDPR foresees obligations and safeguards that could be respected by data controllers without identifying a data subject. This core group of obligations, I argue, would represent a standard for best practice that should be capable of shielding the data controller from liabilities, and possible data subject from damages.[83] Moreover, this conclu-

81. *Breyer*, above n. 6, paras. 23-24.

82. A. Einstein, 'Reply to Criticism: Remarks Concerning the Essays Brought Together in This Co-Operative Volume', in P.A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist* (1949) 665.

83. The data subject is considered as eventual because its identification has not happened yet, but could happen at a later stage, when the damage has already been caused. It will be always the case, for instance, for the transfer of personal data in a third country that does not provide an adequate level of safeguards according to Chapter V of the GDPR. In

sion allows avoiding any measurement *a posteriori* of data as being personal, similarl to the conclusions of the ECJ in *Breyer*. Finally, a solution of this kind would be desirable in a legal framework where administrative fines for unlawful processing of personal data could have significant economic consequences, such as reaching €10,000,000 or 2 per cent of the annual turnover of a company.[84]

We mentioned earlier this *core* group of provisions of the GDPR which should be applicable regardless of the measurement of data as personal in a specific moment. To conclude this section it seems worth presenting a table (Table 1) including these provisions and the rationale behind their applicability.

In conclusion, the core group of provisions listed in Table 1 should provide a fair balance between meeting the need of data controllers to carry out their businesses in a profitable manner without excessive burdens and preventing them from harming data subjects involuntarily. The listed provisions deal with the correct management of data flow in a company and should already be in place for other reasons, mostly linked to the monitoring of the business activities, their profitability and the development of new processing operations.

7 Concluding Remarks

The purpose of this article has been to demonstrate that, despite the best intentions to regulate personal data in a stringent manner, its legal notion has very practical implications. We live in a world dominated by data exchanges, where the saying 'If you are not paying for it, then you are the product' is dramatically fitting. The *Cambridge Analytica* scandal showed that the possibility of transforming data into personal data is very real, and *Breyer* demonstrated its legal implications. In both cases, the binary notion of personal data seemed to be a weak tool to determine the applicability of EU data protection law.

Similar to the problems that physicists had to solve when quantum theory was developed, the notion of personal data has to describe unambiguously the behaviour of personal data in a real-world scenario. The consequences of not doing so are to be mistaken by the measurement of data as personal (or not) in a specific moment, with the certainty that such a result could be reversed at a later stage. This is all the more so when the whole applicability of EU data protection rules depends on that measurement.

The article shows that quantum theory may provide a better point of view, thus enabling the selection of a number of core provisions of the GDPR to avoid the detriment of data subjects, who could suffer damages, and of data controllers, which will have to pay for those damages.

this case, (not yet personal) data can be legally transferred; yet, when that data is used for the identification of the data subject or to enrich a profiling operation that has already taken place, the ultimate result is that the data subject is damaged, but has no legal claim over the data controller that performed the transfer.

84. Art. 83(4) of the GDPR. That amount can be doubled easily according to paras. 5 and 6 of the same article, in case where a company bases its core business on processing data, which only afterwards reveals to be processing of personal data. In fact, paras. 5 and 6 deal with specific cases where either the processing operation went too far and the data subject is irremediably damaged by this or the data controller does not comply with an order of the supervisory authority. In the case where the processing of data is based on the wrong assumption that the data processed is not personal, it is very common to have data transfers towards third countries outside the guarantees of Chapter V. Thus, the processing operations are also engineered on that wrong assumption, and redesigning them is a process that necessarily takes a certain amount of time, during which the company can easily be put out of business.

Tabel 1 Core group of GDPR provisions

Points (b), (c), (d) and (f) of Article 5(1):
Article 5 deals with the principles related to the processing of personal data. In particular, the principles of purpose limitation, data minimisation, accuracy and integrity and confidentiality should be applicable. In turn, those provisions that are strictly related to the presence of a data subject (or the possibility of identifying it) have been excluded.[85] The reasoning is that the data controller might deal with data for which he does not have means likely reasonably to be used to identify the data subject, and may be completely unaware of the fact that that data could lead to the identification of a data subject.[86] On the contrary, the principles we identified as applicable are related to the design of the processing operations and prevent reckless processing of data.
Point (f) of paragraph 1 **and paragraph** 4 **of Article 6:**
Article 6 deals with the lawfulness of processing. Although we deemed as not necessary the provision under point (a) of Article 5(1), the lawfulness referred to in point (f) of Article 6(1) refers to the legitimate interests pursued by the data controller, for instance, its freedom to conduct business. Article 6(4) enriches Article 6(1) and sets further limitations to the processing operations, which include an assessment of the compatibility of the reasons for the further processing, of the need to use encryption or pseudonymisation and an evaluation of the type of data that is being further processed.
Point (f) of Article 14(2):
While Article 14 entails the existence of an identified data subject, the overall goal of the article can be understood from Recital 30 of the GDPR. The idea is that the data controller has to keep track of the personal data it processes. If we apply quantum superposition, and we accept the conclusion that data could turn into personal data at some point, then the data controller should always keep track of where it gets data, where it sends it for further processing, from how long that data is kept and if it transfers it outside the EU.
Section I of Chapter IV, Articles 24 to 31:
Section I of Chapter IV, Articles 24 to 31 establish the obligations between data controllers and data processors. The relationship between the two is fundamental to establishing a good model of governance for the processing operation because although the data processor processes data on behalf of the data controller, it might have a certain degree of flexibility in how certain operations are technically performed.
Article 33:
Article 33 on the notification of data breach towards authorities should be applicable every time a data controller is not able to demonstrate that the data processed under its responsibility is *non-personal data* according to the notion we provided earlier, meaning that the data breach notification should be performed every time the controller has not taken steps to ensure that the data processed is *non-personal* data. Data protection authorities should be put in the position of knowing whether a breach of data that is *potentially personal* could lead different entities, such as cybercriminals, to use the breached data sets with other data sets and ultimately identify data subjects.[87]

[85] In particular, the principle of lawfulness, fairness and transparency and the principle of storage limitation entail obligations that are determined by the data subject. For instance, those two principles will be applied in a very different manner if the data subject is an adult or a minor.

[86] If we consider that data is in the superposition status, and the data controller did not take any measure to make sure that data falls in the *non-personal* data category, then it is legitimate to conclude that another data controller might get access to that data in superposition and make use of its means to combine it with other personal data and ultimately make the data in superposition exhibit entanglement, thus transforming it into personal data.

[87] The accretion problem as such is a neutral process and can be used for legitimate or illegitimate purposes.

Article 37:

The designation of the data protection officer should become the rule where data is processed on a large scale. The designation should be based on the exception provided for in paragraph 4.

Chapter V, Articles 44 to 50:

Transfers of personal data to third countries or international organisations are risky operations by nature because data is transferred to a different jurisdiction with different (or no) safeguards. For this reason, the GDPR allows such transfers only in very limited circumstances and only where the data controller or processor have adopted appropriate safeguards. Therefore, considering the quantum superposition of data, this whole Chapter should be applicable in all cases where the data controller did not put in place mechanisms to ensure that data falls in the *non-personal data* category. The reason for such a stringent conclusion is that once data is transferred outside the EU, it does not matter if it becomes personal: it will still be outside the reach of EU data protection safeguards. All the more so in the case where economic operators amass vast amounts of *potentially personal* data (*e.g.* dynamic IP addresses) and perform the reidentification of subjects outside the EU, in countries where there are no safeguards for personal data and operations like mass-profiling for surveillance reasons are common.[88] The result of that identification can facilitate the use of data mining and predictive analytics techniques, which would ultimately unveil even more personal data on the data subject, with the final goal of using this aggressive profiling on that data subject in the EU.

[88] It is the case for the very recent Social Credit System developed by China. According to this, nothing prevents the fact that China amasses a large amount of *potentially personal* data and performs the identification of tourists or foreigners visiting China, at the border, where biometric data is collected from pictures. *See, for instance,* G. Sgueo, 'Tetris, La Cina e la gamification dei servizi pubblici', available at http://www.forumpa.it/citta-e-territorio/tetris-la-cina-e-la-gamification-dei-servizi-pubblici (last visited 8 November 2018), and A. Cagaan, 'China's Social Credit System raises privacy concerns over surveillance', available at https://www.veridiumid.com/blog/chinas-social-credit-system-raises-privacy-concerns-surveillance/(last visited 8 November 2018). It was also the case for the Prism programme run in the United States by the National Security Agency, which was the main driver behind the ECJ judgment in Case C-362/14 *Maximilian Schrems* v. *Data Protection Commissioner*, where the court stated that the *EU-US Safe Harbour Agreement* was not a legitimate tool for the transfer of personal data from the EU to the United States . *See* A. El Khoury, 'The Safe Harbour Is Not A Legitimate Tool Anymore. What Lies In the Future of EU-USA Data Transfers?', 6 *European Journal of Risk Regulation* 659 (2015).

Right to Access Information as a Collective-Based Approach to the GDPR's Right to Explanation in European Law

Joanna Mazur*

Abstract

This article presents a perspective which focuses on the right to access information as a mean to ensure a non-discriminatory character of algorithms by providing an alternative to the right to explanation implemented in the General Data Protection Regulation (GDPR). I adopt the evidence-based assumption that automated decision-making technologies have an inherent discriminatory potential. The example of a regulatory means which to a certain extent addresses this problem is the approach based on privacy protection in regard to the right to explanation. The Articles 13-15 and 22 of the GDPR provide individual users with certain rights referring to the automated decision-making technologies. However, the right to explanation not only may have a very limited impact, but it also focuses on individuals thus overlooking potentially discriminated groups. Because of this, the article offers an alternative approach on the basis of the right to access information. It explores the possibility of using this right as a tool to receive information on the algorithms determining automated decision-making solutions. Tracking an evolution of the interpretation of Article 10 of the Convention for the Protection of Human Right and Fundamental Freedoms in the relevant case law aims to illustrate how the right to access information may become a collective-based approach towards the right to explanation. I consider both, the potential of this approach, such as its more collective character e.g. due to the unique role played by the media and NGOs in enforcing the right to access information, as well as its limitations.

1 Introduction

The discriminatory potential of automated decision-making solutions has been debated for some time now. Yet, it has only recently received more attention because of the growing, and sometimes contentious, capacities of algorithmic solutions. Publications such as *Weapon of Math Destruction*[1] or *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*[2] inform the broader audience on the threats created by the profiling and algorithms to the most vulnerable groups in society. The fact that algorithms often tend to reproduce human biases and, therefore, to repeat existing discriminatory mechanisms inspires the search for solutions that could guarantee the transparency of automated decision-making processes.

One of these solutions is the right to explanation. The controversy concerning the right to explanation was sparked by colliding opinions on the existence of this right in the General Data Protection Regulation[3] (hereinafter GDPR) and the scope of the GDPR's provisions. The right to explanation can be briefly described as tools that allow the person who is subjected to automated decision-making to be informed about this fact and about the reasoning standing behind this decision. Its function is to provide an individual with instruments that would allow to avoid the discriminatory potential of automated decision-making solutions. The boundaries of this concept's embodiment in the GDPR provoke discussion among scholars triggering the need to search other solutions that may address the threats and challenges posed by the discriminatory potential of automated decision-making solutions.[4] In the article, I present an alternative approach on the basis of perceiving algorithms as information.

I argue that the right to access information could be considered as a more collective-based[5] alternative to right to explanation. The motivation for seeking such an alternative results from limited scope of the right to

* Joanna Mazur, M.A., PhD student, Faculty of Law and Administration, Uniwersytet Warszawski. This research was supported by National Science Centre, Poland: Project number 2018/29/N/HS5/00105 titled *Automated decision-making versus prohibition of discrimination in the European law*.

1. C. O'Neil, *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy* (2016).
2. V. Eubanks, *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor* (2017).
3. Regulation 2016/679, OJ 2016 L 119/1.
4. In favour of a presence of the right to explanation in the GDPR: B. Goodman and S. Flaxman, 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"', *2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*https://bit.ly/2wchh2x (last visited 4 May 2018); against such a possibility: S. Wachter, B. Mittelstadt & L. Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation', 7 *International Data Privacy Law* 76 (2017).
5. Under the term 'collective-based' and 'collective', I understand (1) the special role of media and NGOs, which has been recognised especially by the European Court of Human Rights when realising the right to access information; (2) the character of explanation, which not just refers to a particular individual, but rather offers a model-centric explanation, thus referring to the system, not to the particular decision.

explanation implemented in the GDPR. I examine the legal possibilities of achieving model-centric explanation.[6] Under this term, I understand the solutions that would allow to infer how a system of automated decision-making is structured, for example, inform on all the factors that are taken under consideration in a certain automated decision-making system, their weights, method of assessing the results and so forth. The article is an attempt to examine the possibilities and the limits of applying the right to access information as a way to realise the right to explanation. This would allow us to avoid, to a certain extent, the discriminatory treatment that could result from automated decision-making implemented by the state. The current analysis is indeed strictly focused on automated decision-making solutions that are linked to the state's operations and constitute the examples of state's 'monopoly of information'. Such limitation is warranted by the case law of the European Court of Human Rights (hereinafter ECHR) on which I base my arguments. Even though the ECHR broadened the interpretation of Article 10 of the Convention for the Protection of Human Right and Fundamental Freedoms (hereinafter European Convention),[7] it is debatable if and to what extent the said article is applicable to private entities. Although I do not intend to exclude the possibility of using the approach on the basis of the right to access information in a broader range of situations (*e.g.* concerning horizontal relations), this article focuses specifically and solely on automated decision-making that may occur in the state's operations. In this vein, the article aims to primarily present the reasoning justifying the usage of the right to access information so that a model-centric explanation of automated decision-making solutions used by states is made available for scrutiny.

In order to achieve this goal, the article is structured as follows. The second section starts with some initial remarks on the potential sources of discriminatory treatment in case of automated decision-making. It presents the characteristics of the prohibition on discrimination in EU law and, by doing so, the scope of application of the reasoning developed in the article: this includes exploring the relation between, on the one side, the European Convention and its interpretation and, on the other side, the Charter of Fundamental Rights of the European Union (hereinafter Charter)[8] and its impact on the European law. The third section provides an overview of the possible limitations arising from the approach based on the right to explanation as set out in the GDPR. This would stress the need of having further legal means in order to achieve higher level of automated decision-making transparency. The section ends with the reasons why there is a need to approach the automated decision-making discriminatory potential from a more collective perspective. The fourth section discusses the right to access information in the European law. In this section, the evolution of the interpretation of Article 10 of the European Convention is presented. Its aim is to assess the possibility of using the right to access information whereby states' institutions employ automated decision-making, for example, when providing health services, benefits for the unemployed or the recruitment processes in case of public education. The goal of this section is to present the reasoning standing behind the argument that the right to access information can, to a certain extent, constitute an alternative to the right to explanation. The fifth section concludes.

6. For the explanation of model-centric approach: L. Edwards and M. Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions?"', 16 *IEEE Security & Privacy* 46 (2018).
7. Convention for the Protection of Human Rights and Fundamental Freedoms, 4 November 1950, ETS No. 005.
8. Charter of Fundamental Rights of the European Union, OJ 2012 C 326.

2 Discriminatory Potential of Solutions Using Automated Decision-Making

2.1 Technological Perspective on Discriminatory Potential of Automated Decision-Making Solutions

It is important to notice that the discriminatory potential of automated decision-making has several sources. There are two main sources of concerns, which result from the methods used while preparing solutions allowing automated decision-making. The first one is the character of data used to develop the algorithms. The second one is the choices that are made when deciding which of the collected data should be perceived as important factors influencing the final result of processing.[9] Automated decision-making is – paradoxically – resistant to social changes. Firstly, the input is historically biased: as data on which decisions are based are historical, they can be inherently burdened with prejudice against minorities.[10] Secondly, the decision which of the analysed data should be considered as important is a matter of choice. Machine bias,[11] which is the result of the necessary choices made when testing the program, is the result of the necessity to subject data to generalisation in order to achieve any meaningful

9. Though these two reasons differ, when analyzing certain cases, they usually appear to be linked to each other.
10. 'However, when the input data used by the algorithms are generated by human beings, even algorithms become susceptible to human biases.' – M. Ahsen, M. Ayvaci & S. Raghunathan, 'When Algorithmic Predictions Use Human-Generated Data: A Bias-Aware Classification Algorithm for Breast Cancer Diagnosis', *forthcoming at Information System Research,* at 2 (2017) https://bit.ly/2LQXzj6 (last visited 30 July 2018).
11. This has been subjected to research as early as 1980. Conclusion of the T. Mitchell's study was, 'If biases and initial knowledge are at the heart of the ability to general beyond observed data, then efforts to study machine learning must focus on the combined use prior knowledge, biases, and observation in guiding the learning process. It would be wise to make the biases and their use in controlling learning just as explicit as past research has made the observations and their use.' T. Mitchell, 'The Need For Biases in Learning Generalizations', *Techreport*, at 3 (1980) https://bit.ly/2IkB6t0 (last visited 4 May 2018).

results. Therefore, the categorisation and segmentation, when trying to create automated decision-making solutions, is necessary. However, it must not be forgotten that the choice of what criteria are used for the categorisation are not neutral. Allowing artificial intelligence to analyse the discriminatory present, in order to make automated decisions that determine the future, causes the impression of objectiveness. Lack of human input into this process could be perceived as a tool for making it fairer. However, one should not forget who provides data and tools for analysis.[12]

Referring to the example of algorithms that should support crime prevention, one can say that the selection of a post code as a meaningful variable illustrates the machine bias problems.[13] As it is known that certain districts are inhabited mostly by people of colour, using this variable to assess the risk that the individual may pose in the future has highly discriminatory potential.[14] Another example is the usage of automated decision-making technology to determine what kind of support unemployed person should get: the variables that are taken into account might affect the kind of help that one gets.[15] Arbitrary selection of the meaningful variables may lead to the discrimination of certain groups in the society based on their ethnicity, gender and so forth, thus repeating the discriminatory mechanisms that exist nowadays.

The described mechanisms refer to groups of individuals who share a common characteristic. The discriminatory potential of automated decision-making solutions may therefore have an impact on whole groups, being a potential threat for collective discrimination. However, it can be questioned whether the concept of dividing individuals into groups must necessarily involve discrimination. One could argue that the mechanisms that caused the segmentation of individuals and led to differentiated treatment have always been somehow present. Therefore, the collective discrimination – which can be the result of the above-mentioned mechanisms – is not a unique phenomenon that appears when applying automated decision-making solutions. Moreover, one could argue that it is too early to accuse the technologies that are being developed of discriminatory potential. However, what makes the segmentation in the digital space different from the one in the traditional services sector are the numerous obstacles to the transparency of the divisions that are implemented, for example, intentional concealment by states and corporations or lack of adequate technical and digital literacy of the individuals. From the legal perspective, the obstacles for reaching transparency are, for example, regulations that ensure protection of intellectual property and trade secrets that are necessary to protect the profits of companies developing such solutions.[16] The conflict of interest between subjects making profit – in terms of both monetary character and the efficiency of the processes – thanks to the use of databases and algorithms and the subjects of decisions that are based on big data analysis, will have an impact on the process of spreading automated solutions. As the number of areas in which algorithms are used grows,[17] so grows the disproportion in knowledge on automated decision-making between the broader public and narrow groups of specialists. As a result, these processes produce the need to provide a regulatory framework that would ensure compliance of automated decision-making solutions with the general prohibition on discrimination.

2.2 Discriminatory Potential of Automated Decision-Making Solutions and the Prohibition on Discrimination in European Law

The above-described discriminatory potential of automated decision-making solutions may be perceived as – to a certain extent – a threat to the prohibition on discrimination in the European law. This section presents the character of the prohibition on discrimination in the European law. In doing so, it also presents the scope of the usage of the reasoning, which I present in the article.

On the basis created by the European Convention, the prohibition on discrimination on the grounds indicated in the Article 14 refers to the enjoyment of the substantive rights that are guaranteed by the European Convention itself. To a certain extent, the scope of the prohibition was expanded by Protocol 12 to the European Convention.[18] According to Protocol 12, the ban on discrimination covers any right that is guaranteed at the national level, even where this does not fall within the

12. As V. Eubanks puts it, 'Once the big blue button is clicked and the AFST [Allegheny Family Screening Tool] runs, it manifests a thousand invisible human choices. But it does so under a cloak of evidence-based objectivity and infallibility', above n. 2, at 316 [epub edition].
13. More on the discriminatory character of automated-decision making solutions in the context of crime prevention: '*Profiling* and *data mining* may seem to work up to a point, but inevitably lead to actions against very large numbers of innocent people, on a scale that is both unacceptable in a democratic society…' – D. Korff, 'New Challenges to Data Protection Study', *Working Paper No. 2: Data Protection Laws in the EU: The Difficulties in Meeting the Challenges Posed by Global Social and Technical Developments* 2010: 52; study conducted by ProPublica: J. Angwin, J. Larsona, S. Mattu & L. Kirchner, 'Machine Bias. There's Software used Across the Country to Predict Future Criminals. And it's Biased Against Blacks', *ProPublica* (2016) https://bit.ly/1XMKh5R (last visited 4 May 2018).
14. Abovementioned mechanisms allow scholars to claim, 'The use of algorithmic profiling for the allocation of resources is, in a certain sense, inherently discriminatory: profiling takes place when data subjects are grouped in categories according to various variables, and decisions are made on the basis of subjects falling within so-defined groups. It is thus not surprising that concerns over discrimination have begun to take root in discussions over the ethics of big data' – B. Goodman and S. Flaxman, above n. 4, at 3.
15. For more information on this topic: J. Niklas, K. Sztandar-Sztanderska & K. Szymielewicz, *Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making* (Warsaw 2015) https://bit.ly/1PrMorh (last visited 7 May 2018).
16. For elaboration on some of the obstacles regarding the transparency of automated decision-making: J. Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms', 3 *Big Data & Society* 1 (2016).
17. For complex enumeration of such branches and analysis of the algorithms' impact on society in popular science: O'Neil, above n. 1.
18. Protocol No. 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms, 4 November 2000, ETS No.177.

scope of a European Convention.[19] As only a few countries ratified Protocol 12, the level of protection against discrimination differs across Europe. The consequences of the possible usage of the right to access information in cases referring to the automated decision-making are as follows. In countries that are parties of the European Convention, the case would have to refer to the right to access information on the functioning of discriminatory automated decision-making system in the area covered by the substantive rights guaranteed by the European Convention. The hypothetical example could refer to the usage of the right to access information on the functioning of the automated distribution of cases between judges in relation to the possible threat to the realisation of the right to fair trial.[20] In countries that ratified Protocol 12, the case could additionally refer to rights guaranteed at the national level. In both possibilities, the right to access information would serve as a tool to realise effectively another right that must have been endangered due to the possible discriminatory treatment.

In terms of the prohibition on discrimination in the EU, the relevant provision is set out in Article 21 of the Charter. The scope of prohibition on discrimination refers to the EU's institutions and bodies actions and the actions of the Member States when implementing the EU's law.[21] It is necessary to note that according to the Charter the content of rights should be understood in accordance with the ones guaranteed by the European Convention.[22] Additionally, selected areas and grounds of potential discrimination are covered by the equality directives: the Employment Equality Directive,[23] the Racial Equality Directive,[24] the Gender Goods and Services Directive[25] and the Gender Equality Directive.[26] The character of the prohibition of discrimination for the EU law may also be enshrined by the fact of recognising it as a general principle of the EU law: 'The principle of equal treatment is a general principle of EU law, enshrined in Article 20 of the Charter, of which the principle of non-discrimination laid down in Article 21(1) of the Charter is a particular expression.'[27] However, it must be noted that the overall material scope of prohibition on discrimination in the EU law remains limited:

> the material scope of specific non-discrimination provisions in EU law is often quite limited and uneven. For example, whilst Directive 2000/78/EC only applies in the field of employment and occupation, the material scope of Directive 2000/43/EC is considerably broader, also including e.g. employment-related social security, further access and supply of goods and services, and other matters such as education and social advantages. The only exception to this is the prohibition of discrimination on grounds of nationality, which applies in the full scope of EU law.[28]

As the result of such a character of the prohibition on discrimination in the EU law, the reasoning presented in the article might be used in case of automated decision-making implemented by the EU's institutions and bodies. Moreover, it could be used in case of the EU's Member States in the areas covered by the EU law. The scope of the possible discriminatory treatment resulting from the usage of automated decision-making solutions would have to refer to the grounds on which discrimination is prohibited in this area. The exception would be, as indicated in the quote above, discrimination on grounds of nationality. The character of ban on discrimination on grounds of nationality is more general. If interpreted in accordance with the case law analysed in this article, the right to access information might provide a tool to check if the automated decision-making solutions implemented by the state is within the area of the EU law. The right to access information could provide an insight into the question if automated decision-making solutions implemented by the state and concerning, for example, employment or access to vocational education as guaranteed by Directive 2000/43/EC are not a source of a discriminatory treatment on the basis of sexual orientation, religion or belief, age or disability.

Before presenting the arguments that support such a hypothesis, it is necessary to present the regulatory solutions proposed so far to deal with the issue of potential discrimination resulting from the automated decision making. Such a solution is the right to explanation as implemented in the GDPR. The analysis of the said right is the heart of the next section.

19. European Union Agency for Fundamental Rights/Council of Europe, *Handbook on European Non-Discrimination Law. 2018 edition* at 18 (2018).
20. For a similar argument *see* M. Matczak, 'List do Trybunału Sprawiedliwości Unii Europejskiej ws. praworządności w Polsce' (2018) https://bit.ly/2Fw6pRz (last visited 4 November 2018).
21. Art. 51, Charter of Fundamental Rights of the European Union, above n. 8.
22. Art. 52, *ibid.* This is also the reason why in the article I focus on the analysis of the content of the ECHR's case law referring to the relevant article.
23. Which prohibited discrimination on the basis of sexual orientation, religion or belief, age and disability, in the area of employment: Council Directive 2000/78/EC, OJ 2000 L 303.
24. The Directive prohibits discrimination on the basis of race or ethnicity in the context of employment. Moreover, it refers also to the access to the welfare system, social security, and goods and services: Council Directive 2000/43/EC, OJ 2000 L 180.
25. The Directive Council Directive 2004/113/EC, OJ 2004 L 373.
26. The Directive refers to the equal treatment in relation to social security: Council Directive 2006/54/EC, OJ 2006 L 204.
27. Para. 43, Case C-356/12, *Wolfgang Glatzel* v. *Freistaat Bayern*, [2014], ECLI:EU:C:2014:350.
28. Ch. Tobler, 'Equality and Non-Discrimination under the ECHR and EU Law A Comparison Focusing on Discrimination against LGBTI Persons', 74 *Zeitschrift für auslandisches öffentliches Recht und Völkerrecht* at 532 (2014).

3 Right to Explanation: An Approach Based on the Data Protection Framework

3.1 Right to Explanation in the GDPR

The term and concept of the right to explanation has been developed as a tool to ensure privacy protection and should – for now – be understood mainly as an element of data protection law. The discriminatory character of automated decision-making procedures is to a certain extent addressed at the EU level by the GDPR. The data subject, according to the Articles 13-15 of the GDPR, should be informed about:

> the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.[29]

The Articles 13-15 of the GDPR refer to, respectively, information that is to be provided where personal data are collected from the data subject, information that is to be provided where personal data have not been obtained from the data subject and the right of accessing data by the data subject. The common provision regarding 'meaningful information', which should be delivered to the data subject, can be perceived as a step towards increasing the level of consciousness of individuals in the area of automated decision-making. To a certain extent, these obligations may address the above-mentioned issue of insufficient digital literacy. However, the lack of precision regarding the scope of 'meaningful information about the logic involved' leads to a broad informational obligation that seems difficult to pin down.

Moreover, the possibility of combating online discrimination on the basis created by the GDPR is weakened by the fact that – as a general rule – the GDPR allows both automated individual decision-making and profiling.[30] According to the GDPR, the data subject is granted the right 'not to be subjected to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'.[31] The threshold set for the possibility of opposing the automated decision-making is relatively high. Firstly, this right refers to a decision, not to the processing itself. Therefore, it allows developing technologies that may be discriminatory and introduces its control on the last level of the process, when the decision in question has been already made. The adopted form of the GDPR does not address the problems that result from the lack of the automated decision-making technologies' transparency, from the perspective of the individual. Secondly, Article 22 of the GDPR refers to a decision based *solely* on automated processing. As a result of such phrasing, decisions predominantly based on automated processing would be excluded from its scope.[32] This may significantly limit the number of decisions that may be questioned on the basis guaranteed by the GDPR. Thirdly, doubts should be raised with regard to the understanding of the denotation 'similarly significantly affects'. The impact of the decision may differ depending on the individual conditions of, for example, economic or social character. The phrasing implemented in the GDPR can strengthen the role of discretion in the process of assessing the decision's character. Moreover, there are three grounds on which automated individual decision-making can be justified[33] – including the user's consent – which make it even more difficult to visualise the potential impact of Article 22 as threatening the practices of automated decision-making and profiling in the web. Even though the GDPR provides grounds to debate the right to explanation and its character, it seems to offer limited possibilities to effectively address the challenges linked to the discriminatory potential of automated decision-making technologies.

Having said that, it is necessary to note two additional factors that provide motivation for searching alternative legal means to ensure a non-discriminatory character of the digital space. The first is the extent to which the logic involved in automated processing should be revealed to the data subject. As is stated in recital 63 of the GDPR, 'that right should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software'.[34] The unrestrained development of data-driven[35] economy and high level of personal data protection is hardly achievable, which can be illustrated by the above-mentioned example of limiting the initial scope of the GDPR's Article 22: the protection against automated decision-making refers to the decision based solely on automated processing, which leaves aside the decisions based predominantly on automated processing. On the one hand, it does not impede the possibility of developing solutions using automated decision-making as the vital factor influencing certain decision. On the other hand, due to such phrasing the individual's right to explanation may cease to have any real effect.

The second problem is predominantly individual character of the right to explanation included in the GDPR.

29. Arts. 13-15, above n. 3.
30. Profiling in GDPR is presented as a special category of individual decision-making: Art. 22, *ibid.*
31. *Ibid.*
32. The authors of 'Why a right to explanation of automated decision-making does not exist in the general data protection regulation' point out the evolution of the proposed scope of the Art. 22. The legislative process led to the exclusion of denomination 'predominantly' from the final version of this legal act: S. Wachter, B. Mittelstadt & L. Floridi, above n. 4, at 92.
33. Art. 22(2), above n. 3.
34. *Ibid.*, Rec 63.
35. M. Mandel, 'Beyond Goods and Services: The (Unmeasured) Rise of the Data-Driven Economy', *Progressive Policy Institute: Policy Memo* (2012) https://bit.ly/2FLBcVk (last visited 4 May 2018).

Even the phrasing, namely the term 'automated individual decision-making', shows its focus on an individual perspective: it is the individual who is subjected to the decision in question. It is the individual who can object to the decision based on automated decision-making. Such an approach somehow leaves aside the question of a possible collective character of discriminatory practices, which are based on big data analysis. Simultaneously, so-far-identified and described impact of the machine bias when implementing the automated decision-making solutions shows that it affects the minorities and the most vulnerable groups in the society.[36] The possibility of collective discrimination resulting from automated decision-making should provoke questions about the legal means in the GDPR, which can allow to combat such threats.

3.2 Doubts Concerning the Collective Dimension of the Right to Explanation in the GDPR

In case of automated decision-making one should ask: what if 'I' is also a 'we'? What if this particular decision that has been taken in one case is in fact representative for a whole group in the society, which has been defined on the basis of big data analysis? The tension between personalisation and big data–based technologies becomes more evident nowadays: the individualisation of content presented to individuals is only possible due to the analysis of data of millions. Defining common characteristics allows to undertake actions in scale of millions of individuals. Effectiveness of profiling is the result of the algorithms' being fed enormous data collections. Therefore, one could wonder what law can offer in terms of applying right to explanation in order to address the collective dimension of discriminatory potential and risks posed by the automated decision-making technologies. In terms of the GDPR's provisions, one could evoke Article 35. It refers to carrying out a data protection impact assessment if it is likely to result in high risk to the rights and freedoms of natural persons.[37] However, it must be noted that impact assessment is not addressed to the broader public. It does not empower the users or groups of users, and it does not allow the users or groups of users to take any control over the process of assessing the potential impact of data processing. As such, it does not constitute an element of the right to explanation.

Considering the GDPR's collective dimension, it is necessary to examine Article 80.[38] It allows the data subject to mandate a not-for-profit body, organisation or association to lodge the complaint on its behalf. Moreover, Article 80(2) provides the Member States of the EU with the opportunity to grant anybody, an organisation or an association referred to in Article 80(1) independently of the data subject's mandate, the right to lodge a complaint and to exercise certain rights included in the GDPR.[39] However, this representation refers to the rights granted in the GDPR and therefore the limits to the right to explanation apply to the proceedings initiated on the basis of Article 80. They focus on the particular decision referring to the individual. The abstract control, understood as a legal equivalent of the above-described model-centric explanation, potentially performed by an NGO may, but does not have to, be allowed by the Member States. This can lead to a conclusion that in the GDPR there are no obligatory legal means that ensure transparency of the overall mechanisms standing behind automated decision-making solutions. There is only a slight possibility for single individuals to receive information on the grounds of a decision about their own individual case. However, it is not possible for a potentially discriminated group to examine *in abstracto* the systemic dimension of automated decision-making solutions and their discriminatory potential. The discretional power of the Member States on this matter could prevent the potential development of tools which would allow wide engagement of the civil society organisations in issues related to the right to explanation. Therefore, I propose to analyse to what extent the right to access information may fill the GDPR's shortcomings. Does focusing not on 'data' itself but on 'information' may strengthen the users' position? May it result with providing the individuals with the insight into the logic standing behind the automated decision-making solution? May it be a tool used for receiving model-centric explanation instead of one focused on a particular decision?

3.3 Right to Explanation in the GDPR and Right to Access Information: The Necessity of Shifting from Individual- to Collective-Based Approach

It is necessary to note that the above-mentioned right to explanation in the GDPR technically could refer both to the overall system functionality focused on a certain group (model-centric explanation)[40] and to the specific decisions concerning an individual.[41] The term used in Articles 13-15 of the GDPR, namely, 'logic involved', could – if interpreted broadly – provide the user with more general information on the system that allows automated decision-making. However, it might as well refer solely to the elements of the system, which had an impact on the decision concerning individual in the particular case. As the approach presented in the GDPR seems to suggest, the information on the logic involved in the processing should predominantly help to understand why this particular 'one' was subjected to a certain decision. This approach – more probable when one takes into account the valuable character of programmes

36. For detailed case study, see: Eubanks, above n. 2.
37. Art. 35, above n. 3. Art. 35(3) includes list of three cases in which impact assessment shall be required.
38. Art. 35, *ibid.*
39. For detailed analysis of this issue: L. Edwards and M. Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?', 16 *IEEE Security & Privacy* 46 (2018) https://bit.ly/2IDsBcO (last visited 12 February 2019).
40. L. Edwards and M. Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For', 16 *Duke Law & Technology Review* 18 (2017).
41. Wachter, Mittelstadt & Floridi, above n. 4, at 78.

used to perform activities leading to automated decision-making – contradicts the attitude presented by some scholars regarding the specific character of big data analysis: to a certain extent collecting and processing data may lead to 'learning nothing about an individual while learning useful information about a population'.[42] Far from espousing such a one-sided approach, I would argue that big data–based technologies cause a feedback loop effect: as growing collection of data on individuals increases the possibilities of identifying group characteristics, the detailed characteristic of a group allows to complete an individual profile on the basis of information about the group, to which one seems to belong. Referring to the term used by M. Hildebrandt,[43] this can lead to the creation of 'non-distributive group profiles': assigning one to a certain group on the basis of selected characteristics of an individual (selected personal data). Even though there may be significant determinants that are not taken into account, and which could change the way in which one is classified, they are not considered as valid for such a classification.[44]

The limitations of the approach based on the personal data protection can be stressed by evoking the Court of Justice of the European Union's (hereinafter ECJ) case law concerning personal data. In the case *YS* v. *Minister voor Immigratie, Integratie en Asiel* the ECJ notices that 'the data in the legal analysis contained in that document, are "personal data" within the meaning of that provision, whereas, by contrast, that analysis cannot in itself be so classified'.[45] The analogy with automated decision-making system shows that the individual may receive access to the personal data used to make a decision and to the decision itself; however, the analysis remains outside the scope of the term 'personal data' and therefore cannot be subjected to such an access. The concept of personal data involves the possibility of linking certain information with a particular individual, for example, one's name and surname, e-mail address containing one's surname and place of work or IP address.[46] As explained earlier, the source of potential discrimination in case of automated decision-making solution may not be linked to the individual and his or her personal data: it may be the result of how the particular automated decision-making system was structured.

In order to achieve effective protection against the possible discrimination, it is necessary to shift from the perspective focused on an individual and personal data to the perspective that focuses on a group and the information on how the automated decision-making system works. The advantage of the solution based on the right to access information is its more systemic approach towards the prohibition on discrimination. Taking into consideration the material scope of the non-discrimination provisions in the EU law explained earlier, its possible usage might be illustrated with the following example of the potential discrimination on grounds of nationality. The approach based on the right to access information would allow, for example, to check if the automated decision-making solution, which is implemented by the state, is somehow determined to result with the unequal treatment of the country's citizens and the nationals of the other Member States due to the factors that are taken into account when analysing data. It would allow to determine whether the systemic solutions based on automated decision-making and implemented by the Member State, are in accordance with the prohibition on discrimination.

The next section presents reasoning standing behind the hypothesis that the right to access information might be a tool to achieve such a model-centric explanation, focused on exploring the discriminatory potential of automated decision-making solution, instead of being focused on protection of individual's personal data, which in fact only fuels the automated decision-making solution.

4 Right to Explanation: An Approach Based on the Right to Access Information

4.1 Right to Access Information as a Human Right: Evolution of Interpretation of the European Convention's Article 10

Recognising the right to access information as a human right is not obvious. Even though Article 10 of the European Convention and Article 11 of the Charter provide the individuals with the 'right … to receive and impart information and ideas without interference by public authority and regardless of frontiers',[47] only in

42. C. Dwork and A. Roth, 'The Algorithmic Foundations of Differential Privacy', 9 *Theoretical Computer Science* 211, at 215 (2013). Similarly: 'We should acknowledge the change, and accept that privacy is a public and collective issue' – P. Casanovas, L. De Koker, D. Mendelson & D. Watts, 'Regulation of Big Data: Perspectives on Strategy, Policy, Law and Privacy', 7 *Health and Technology* 1, at 13 (2017); and 'predictions based on correlations do not only affect individuals, which may act differently from the rest of the group to which have been assigned, but also affect the whole group and set it apart from the rest of society' – A. Mantelero, 'Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection', 32 *Computer Law & Security Review* 238, at 239 (2016).
43. M. Hildebrandt, 'Profiling: From Data to Knowledge. The Challenges of a Crucial Technology', 30 *Datenschutz and Datensicherheit* at 548 (2006).
44. Which is the effect of above-mentioned source of potential discrimination, namely the choices made during the meaningful variables data selection.
45. Joined Cases C-141/12 and C-372/12, *YS* v. *Minister voor Immigratie, Integratie en Asiel and Minister voor Immigratie, Integratie en Asiel* v. *M and S.*, [2014] ECLI:EU:C:2014:2081.
46. Case C-582/14, *Patrick Breyer* v. *Bundesrepublik Deutschland*, [2016], ECLI:EU:C:2016:779.
47. The phrasing of ECHR and the Charter is in this regard the same. The content of the Articles is similar to the Art. 19 of the Universal Declaration of Human Rights ('to seek, receive and impart information and ideas through any media and regardless of frontiers') – Universal Declaration of Human Rights, 10 December 1948, General Assembly resolu-

2006 has the ECHR begun to interpret Article 10 of the European Convention broadly. The ECHR's judgements stress the conditionality of the right to access information and therefore remain behind other human right bodies, for example, Inter-American Court of Human Rights, which have already recognised a self-standing right to access information.[48] The reason for such temperance is the grounds on which the broad interpretation of Article 10 is based. The ECHR's interpretation results not from the literal reading of the European Convention. It is mostly the result of broad consensus that can be observed regarding the right to access information both on the international level and on the level of the domestic laws of the overwhelming majority of Council of Europe Member States.[49] In this section, I present the selected case law that illustrates the change in the ECHR's approach towards the right to access information and the general tendencies concerning the ECHR's interpretation of the right to access information, which can be drawn from the analysed cases.

The recognition of a right to access information in the ECHR's case law dates back to 2006. The case *Sdružení Jihočeské Matky* v. *Czech Republic*[50] concerned information demanded by a non-governmental organisation about a nuclear power plant. Even though the ECHR decided that essentially technical information about the nuclear power station[51] did not reflect a matter of public interest, it opened the possibility of interpreting Article 10 of the ECHR as a source of demanding access to administrative documents from public institutions. The shift that came with *Sdružení Jihočeské Matky* v. *Czech Republic* is unprecedented. Even though Article 10 offers several reasons for which the scope of information shared publicly may be limited,[52] the overall attitude towards the right to access information has changed. The right to access information has been recognised as an element of Article 10: as a rule – under certain conditions – the public should be given access to the relevant information, and as an exception the limitations to this rule could be evoked.

The confirmation of such a notion can be found in *Társaság a Szabadságjogokért* v. *Hungary*.[53] The Hungarian NGO requested the Constitutional Court to grant them access to the complaint pending before it. The Constitutional Court denied the request, explaining that a complaint could not be made available to outsiders without the approval of its author on the basis of the protection of the Member of Parliament's personal data. The ECHR explicitly stated, 'The Court has recently advanced towards a broader interpretation of the notion of freedom to receive information and thereby towards the recognition of a right of access to information'.[54] Due to the public character of the information requested by the NGO, the ECHR confirmed that denying access to the complaint was a violation of Article 10.

The occasion to strengthen the trend of broad interpretation of Article 10 resulted from the proceeding initiated by the Austrian non-governmental organisation demanding access to decisions regarding transfers of ownership of agricultural and forest land in Tirol: *Österreichische Vereinigung zur Erhaltung, Stärkung und Schaffung Eines Wirtschaftlich Gesunden Land- und Forst-Wirtschaftlichen Grundbesitzes* v. *Austria*.[55] According to the judgement,

> the applicant association was therefore involved in the legitimate gathering of information of public interest. Its aim was to carry out research and to submit comments on draft laws, thereby contributing to public debate.[56]

The ECHR stated that the reasoning standing behind such an interpretation can be based on the fact that the state's monopoly on information actually interferes with the activities performed by NGOs as social 'watchdogs'.[57]

When explaining the threshold criteria, which need to be fulfilled in order to evoke the right to access information in the case *Magyar Helsinki Bizottság* v. *Hungary*, the ECHR enumerates four conditions. Firstly, 'the purpose of the person in requesting access to the information held by a public authority is to enable his or her exercise of the freedom to "receive and impart

tion 217 A; and the Art. 19(2) of the International Covenant on Civil and Political Rights ('this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice') – International Covenant on Civil and Political Rights, 16 December 1966, General Assembly resolution 2200A (XXI). Lack of the verb 'seek' in the European Convention results with doubts concerning the possibility of interpreting the Art. 10 as containing the right to access information. These doubts are illustrated by the evolution of the case law presented in the article.

48. '…the Court finds that, by expressly stipulating the right to "seek" and "receive" "information," Article 13 of the Convention protects the right of all individuals to request access to State-held information, with the exceptions permitted by the restrictions established in the Convention' – Inter-American Court of Human Rights, *Claude Reyes et al.* v. *Chile*, Judgment, 19 September 2006, para. 77.
49. 'The Convention cannot be interpreted in a vacuum and must, […], be interpreted in harmony with other rules of international law, of which it forms part. Moreover, […] the Court may also have regard to developments in domestic legal systems indicating a uniform or common approach or a developing consensus between the Contracting States in a given area' – *Magyar Helsinki Bizottság* v. *Hungary* (2016) No. 18030/11, para. 138.
50. *Sdružení Jihočeské Matky* v. *Chech Republic*, ECHR (2006) No. 19101/03.
51. It is worth noticing that the roots of direct recognition of the right to access information can be linked to the protection of the environment. It has been implemented in Art. 4 of Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters, 25 June 1998, UNTS 2161 at 447.
52. Analysed in detail below.
53. *Társaság a Szabadságjogokért* v. *Hungary*, ECHR (2009) No. 37374/05.
54. *Ibid.*, para. 35.
55. *Österreichische Vereinigung zur Erhaltung, Stärkung und Schaffung Eines Wirtschaftlich Gesunden Land- und Forst-Wirtschaftlichen Grundbesitzes* v. *Austria*, ECHR (2013) No. 39534/07.
56. *Ibid.*, para. 36.
57. '…stating that the most careful scrutiny was called for when authorities enjoying an information monopoly interfered with the exercise of the function of a social watchdog' – *ibid.*, para. 41.

information and ideas" to others'.[58] This illustrates subsidiary character of the right to access information as a provision included in the Article, which reflects on the freedom of expression. Therefore, as explained in the Sub-Section 4.3., the special role of media and NGOs when executing the right to access information must be stressed. Secondly, the information, data or documents to which access is sought must meet a public interest test.[59] The ECHR does not elaborate on the conditions that shall be fulfilled in order to comply with this test, claiming that this definition 'depend[s] on the circumstances of each case'.[60] I hypothesise on the possible meaning of this test in regard to the algorithms in the Sub-Section 4.2. Thirdly, 'an important consideration is whether the person seeking access to the information in question does so with a view of informing the public'. This functional approach towards the information requested was envisioned in the above-mentioned case law. It also strengthens the position of media and NGOs as natural candidates for seeking access to the information in purpose of informing the public (see Sub-Section 4.3.). Additionally, the ECHR notes that

> the fact that the information requested is ready and available ought to constitute an important criterion in the overall assessment of whether a refusal to provide the information can be regarded as an 'interference' with the freedom to 'receive and impart information' as protected by that provision.[61]

I refer to this condition in Sub-Section 4.2.
Such conditions provide an argument that is crucial when analysing the possibility of using the right to information as an alternative to the right to explanation. The role of the state as a guarantee of the right to information – seen from the perspective of the ECHR judgements – has evolved. From being viewed as a purely passive actor, whose function was not to disturb the flow of information,[62] state may be considered more active player if state monopoly of information is under consideration.[63] The shift in the ECHR's interpretation of Article 10 of the European Convention and towards the relationship between the state and the guards of democratic values embodied by the media and NGOs could have an impact on the right to access information in regard to digital space. However, the possibilities and limits of such a concept in regard to algorithms need to be explored. In the next sub-section, I present the issues that should be considered in order to apply Article 10 to scrutinise or prevent discriminatory treatment when applying automated decision-making technologies.

4.2 Right to Access Information in Digital Space: Algorithms as Information of Public Interest

In order to examine the legal viability to apply the right to access information to issues resulting from the development of digital economy, three issues shall be considered. Firstly, I analyse whether algorithms on which automated decision-making is based can be viewed as information. Secondly, I examine the condition of being information of public interest, as it may limit the extent to which Article 10 can apply in regard to automated decision-making. Thirdly, the character of information that could potentially be received in case of automated decision-making technologies should be identified.
The possibility of understanding an algorithm as an information is based on the view that algorithms, in their broad – and original – meaning, are chains of commands, or, as Robin K. Hill briefly puts it, 'finite, abstract, effective, compound control structure'.[64] Their characteristics include 'accomplishing a given purpose under given provisions'.[65] However, nowadays a semantic shift from this purely theoretical sense towards a more pragmatic meaning is taking place. In public discourse, the term algorithm usually refers to 'the implementation and interaction of one or more algorithms in a particular program, software or information system'.[66] In both cases – the mathematical approach and the one represented in public discourse – an algorithm can be presented as a nexus: it allows for analysis of data and gaining meaningful results. Therefore, it may be perceived as information on how the process is organised. The key element of applying Article 10 to automated decision-making technologies is to disenchant algorithms and view them simply as information on how the architecture of automated decision-making processes – irrespective of the level of their complexity – has been designed, that is, which variables are considered as meaningful. This perspective on the algorithm complies

58. *Magyar Helsinki Bizottság* v. *Hungary*, above n. 49, para. 158.
59. *Ibid*., para. 161.
60. *Ibid*., para. 162.
61. *Ibid*., para. 170.
62. The example of such an approach: 'The Court observes that the right to freedom to receive information basically prohibits a Government from restricting a person from receiving information that others wish or may be willing to impart to him' – *Leander* v. *Sweden*, ECHR (1987) No. 9248/81, para. 74; or: 'That freedom cannot be construed as imposing on a State, in circumstances such as those of the present case, positive obligations to collect and disseminate information of its own motion' – *Guerra and Others* v. *Italy*, ECHR (1998) No. 14967/89, para. 53. The fact that state is under no circumstances obliged to disseminate information of its own motion has been confirmed in *Magyar Helsinki Bizottság* v. *Hungary*, above n. 49, para. 156. The tension between lack of positive obligations from the state's side and its more active role promoted by the above-mentioned judgements probably will result with continuation of the case law explaining the conditions that should be met when using the right to access information, for example, what is information of public interest? How to address the state's monopoly of information?
63. Simultaneously not being obliged to perform information activities out of its own motion, see above n. 62.
64. R.K. Hill, 'What an Algorithm Is', 29 *Philosophy & Technology* 35, at 44 (2016).
65. *Ibid.*, at 47. This understanding of algorithms implies that they do not have to be even digitized: 'Algorithms need not be software: in the broadest sense, they are encoded procedures for transforming input data into a desired output, based on specified calculations' – T. Gillespie, 'The Relevance of Algorithms', in T. Gillespie, P. Boczkowski & K. Foot (eds.), *Media Technologies*, *Essays on Communication, Materiality, and Society* (2014) 167, at 167.
66. B.D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, & L. Floridi, 'The Ethics of Algorithms: Mapping the Debate', *Big Data & Society* at 2 (2016).

with the above-described condition of the requested information being 'ready and available'. On the basis of the relevant case law, it is impossible to argue that the state should provide the analysis of how automated decision-making solution works. Nevertheless, it could be obliged to provide the access to the raw algorithm itself. This might be perceived as a path to ensuring model-centric explanation of automated decision-making solutions to broader public.

Considering the second issue, this analysis is limited to information of public interest, even though the automated decision-making process can refer to an infinite number of issues. Due to, for example, the dominant character of the ECHR's case law regarding the right to access information as well as above-mentioned conflict of rights between the intellectual property rights and the right to access information, my argument here is strictly limited to the automated decision-making technologies used by the state's institution (the algorithms that underpin *operations of the state*).[67] Following the case law of the ECHR, the condition that would have to be fulfilled on demanding the access to information in question is the existence of state's 'monopoly of information,'[68] which is described in the ECHR's case law as a form of censorship. The logic presented in the ECHR's case law runs as follow: in case of the refusal of access to the information on how the system works, the state who possessed 'monopoly of information' would limit the possibilities on media and NGOs to exercise their function of conducting informed public debate. Therefore, the hypothesis of this article could be applied to algorithms that determine the knowledge about issues that constitute matters of public interest, as their importance for the public debate may not be questioned.

I would suggest that the automated decision-making technologies used to determine access to social benefits or automatically assign juries could serve as possible examples. I would argue that in case of automated decision-making solutions used to provide public services, such as public insurance, public education or public health services, the relevant algorithms could be subjected to the more proactive interpretation of the right to information, which has been developed by the ECHR. Not only do the states exercise information monopoly in these areas, but their impact on public matters of special interest to the society could also be considered as a reason for ensuring the transparency of the organisation process.

This characteristic of the right to access information shows the differences between approaching the right to explanation from the perspective of data protection and from the perspective of the right to access information. Contrary to the GDPR-based approach, which ultimately is focused on the effects of automated decision-making for a particular individual, the approach based on the right to information would allow for a more abstract and general control of the mechanisms determining automated decision-making. Firstly, it could justify access to the documents that regulate decision-making procedures concerning groups of people, allowing to apply a more collective perspective than the one focused solely on the individual, as is the case with the GDPR.[69] Secondly, the collective dimension of the right to access information is strictly linked to the special position of the media and NGOs in executing the freedoms and rights guaranteed in Article 10 of the European Convention, to which is dedicated the next sub-section.

4.3 Who Is a 'We'? Media and Non-Governmental Organisations as Citizens' Representatives

It should not be overlooked that the processing of big data is based on mechanisms that allow for dividing individuals into groups that have certain common characteristics. The collective character of the potential discrimination seems to be an inescapable argument, tilting the scale for the possibility of recognising the right to information as an alternative to the tightly restricted right to explanation implemented in the GDPR. The special position of the media and NGOs has been stressed by the ECHR in numerous judgements and has been approached from the functional perspective:

> However, the function of creating forums for public debate is not limited to the press. That function may also be exercised by non-governmental organisations, the activities of which are an essential element of informed public debate.[70]

67. It might be possible to broaden the scope of right to access information: 'The Court has further emphasised the importance of the right to receive information also from private individuals and legal entities. While political and social news might be the most important information protected by Article 10, freedom to receive information does not extend only to reports of events of public concern, but covers cultural expressions and entertainment as well….': European Court of Human Rights, *Internet: Case-Law of the European Court of Human Rights*, 2011 (update: 2015), at 43 https://bit.ly/2HYhITm (last visited 7 May 2018).
68. 'The Constitutional Court's monopoly of information thus amounted to a form of censorship.; *Társaság a Szabadságjogokért* v. *Hungary*, above n. 53, para. 28.
69. Even though data protection may provide tools that to certain extent allow auditing the processes standing behind automated decision-making, they are mostly of voluntary or self-regulatory character: above-mentioned data-processing impact assessments and codes of conduct or the possibility of establishing certification mechanisms, the latter two not having obligatory character. For presentation of this possibilities see: B.W. Goodman, 'A Step Towards Accountable Algorithms? Algorithmic Discrimination and the European Union General Data Protection', *29th Conference on Neural Information Processing Systems (NIPS 2016)*, at 4-5 https://bit.ly/2rlBzSf (last visited 7 May 2018).
70. *Österreichische Vereinigung zur Erhaltung, Stärkung und Schaffung Eines Wirtschaftlich Gesunden Land- und Forst-Wirtschaftlichen Grundbesitzes* v. *Austria*, above n. 55, para. 34. *See also*: 'However, the realisation of this function is not limited to the media or professional journalists. In the present case, the preparation of the forum of public debate was conducted by a non-governmental organisation. The purpose of the applicant's activities can therefore be said to have been an essential element of informed public debate' – *Társaság a Szabadságjogokért* v. *Hungary*, above n. 53, para. 27.

The unique position of media and NGOs in regard to the right to access information is firmly embedded in case law concerning the right to access information. As the actors whose function is enabling and participating in the informed debate, their primary task is to provide the information to the broad public. Therefore, they fulfil the conditions set out by the ECHR in regard to recognition of the right to access information. The judiciary practice of the ECHR continuously recognises a special role of the media and non-governmental organisations as guards of democracy and somehow privileged actors in terms of executing rights included in Article 10 of the European Convention.[71] Not only are they perceived by the ECHR as actors whose mission is to inform the public on most important issues, but they are also legitimised to demand access to public information from the governmental institutions in order to inform broader public. They seem to be the subject most befitting this function: as they are the representatives of civil society, the impact of their actions should be more fruitful than legal actions undertaken solely by individuals. Moreover, one of the obstacles mentioned in the introduction to this article that limits the transparency of the implemented solutions is the lack of adequate digital literacy of individuals. Specialised NGOs[72] or well-informed journalists could instead act as intermediaries between the individual and the decision makers (or shall we say, decision-making automated solutions).

The presence of such representatives as NGOs is crucial for ensuring fairness and non-discriminatory treatment when applying automated decision-making.[73] The potential of using traditional importance of the media and NGOs in regard to the right to access information allows, as I argue, for the possibility of bringing the issue of automated decision-making to the collective dimension understood as a right to model-centric explanation. Instead of focusing on the explanation of a decision referring to one particular individual, it could focus on the architecture of the system used to determine the automated decision-making rules. It answers the systemic challenges created by the automated decision-making solutions. It provides the organisations representing certain groups with power to question the fairness of the system created to determine automated decision-making solutions. However, even their privileged position should be subjected to certain limitations, which I examine in the next sub-section.

4.4 Limits of Right to Access Information in Digital Space

The consequences of applying Article 10 to algorithms that determine automated decision-making in case of state's operations bring up the necessity to analyse limitations imposed on the right to information by the European Convention itself. I would argue that the right to access information, as understood by the ECHR, can refer to the state's areas of activity. The examples of operations included in the scope of this article's hypothesis could include automated decision-making systems, which determine access to public services (*e.g.* unemployment benefits).[74]

However, according to Article 10(2) the exercise of freedoms guaranteed by Article 10 may be subject to restrictions prescribed by law and necessary in a democratic society, among others, in the interest of national security and public safety, for the prevention of disorder or crime, for the protection of health, and for preventing the disclosure of information received in confidence.[75] In the above-mentioned case *Österreichische Vereinigung zur Erhaltung, Stärkung und Schaffung Eines Wirtschaftlich Gesunden Land- und Forst-Wirtschaftlichen Grundbesitzes* v. *Austria*,[76] the ECHR analysed in detail if the interference with the applicant association's right to receive and to impart information as enshrined in Article 10(1) was justified on grounds offered by Article 10(2), namely, prescribed by law, pursuing one or more of the legitimate aims set out in that paragraph[77] and necessary in a democratic society. The conclusion of the judgement in this aspect may be perceived as a test of conditions that have to be met in order to be able to lawfully refuse providing the information: according to the ECHR, the refusal was prescribed by law and pursued

71. For the analysis of the role of media in the ECHR's case law concerning Art. 10 *see*: T. Mendel, *A Guide to the Interpretation and Meaning of Article 10 of the European Convention on Human Rights* (2017), at 14-17 https://bit.ly/2OwOACd (last visited 30 July 2018).
72. It is worth to note that in the Art. 80 of the GDPR the conditions that the organisation representing the individual has to meet include: '…and is active in the field of the protection of data subjects' rights and freedoms with regard to the protection of their personal data'. This element may limit the number of organisations that would be allowed to represent the individuals in cases initiated in order to ensure the execution of the right to explanation based on the GDPR's provisions – Art. 80(1), above n. 3.
73. Moreover, it is necessary to admit that the analysis is partly inspired by a ruling of the Polish Voivodship Administrative Court in Warsaw, which decided that algorithms could be treated as public information and which was initiated by the Polish non-governmental organisation Panoptykon. The case regarded algorithms that are involved in providing services for the unemployed. It allowed dividing them into three groups, which determined the scope of support granted to each individual. The administrative court decided that the mechanism that formed the basis for the classification should be revealed accordingly to the regulations concerning public information. Judgement of WSA in Warsaw, II SAB/Wa 1012/15, 5 April 2016. Moreover, recently the case concerning the access to the algorithm determining the System of Random Allocation of Cases has been initiated: K. Izdebski, 'Algorithms of Fairness', *Medium*, 15 February 2018 https://bit.ly/2GeO7zH (last visited 30 July 2018). In the moment of preparing this article, the outcome of the proceeding has been unknown.
74. Niklas, Sztandar-Sztanderska & Szymielewicz, above n. 15.
75. Art. 10(2), above n. 7.
76. *Österreichische Vereinigung zur Erhaltung, Stärkung und Schaffung Eines Wirtschaftlich Gesunden Land- und Forst-Wirtschaftlichen Grundbesitzes* v. *Austria*, above n. 55.
77. The catalogue of the legitimate aims is included in the Art. 10(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.': Art. 10(2), above n. 7.

the legitimate aims. However, it was not considered by the ECHR as necessary in a democratic society:

> the reasons relied on by the domestic authorities in refusing the applicant association's request for access to the Commission's decisions – though 'relevant' – were not 'sufficient'. While it is not for the Court to establish in which manner the Commission could and should have granted the applicant association access to its decisions, it finds that a complete refusal to give it access to any of its decisions was disproportionate.[78]

Tensions between the technological possibilities offered in areas such as health insurance (*e.g.* adjusting an offer) or crime prevention and the execution of the right to information are impossible to avoid. Time will show how the ECHR will resolve the issue of setting the boundaries between the right to access information and the state's justified interests to protect its activities. Nevertheless, the outcome of the test of conditions that should be met when justifying the refusal of information applied in *Österreichische Vereinigung zur Erhaltung, Stärkung und Schaffung Eines Wirtschaftlich Gesunden Land- und Forst-Wirtschaftlichen Grundbesitzes* v. *Austria*[79] proves that the condition of being 'necessary in a democratic society' included in Article 10(2) may lead to possible restrains of the right to access information.

5 Conclusions: The Right to Access Information and the Rule of Law in the Digital Space

The necessity to rethink what is information and how it should be treated is growing because of the spreading of automated decision-making technologies and big data analyses. Moreover, the datasets used for such analyses are constantly growing, and 'the Big Data of today can easily become the little data of tomorrow.'[80] There is a strong need to confront the methods applied to such analyses with the general prohibition on discrimination, which is crucial to ensure the democratic fundaments of European countries. As Hildebrandt claims,

> The Rule of Law aims to create an institutional environment that enables us to foresee the legal effect of what we do, while further instituting our agency by stipulating that such effect is contestable in a court of law – also against big players (…) Such a – procedural – conception of the Rule of Law implies that both automation and autonomics should be constraint in ways that *open them up to scrutiny* [emphasis of the author] and render their computational judgements liable to being nullified as a result of legal proceedings.[81]

The usage of right to access information could 'open up to scrutiny' at least certain automated decision-making solutions and provide the citizens with the answers whether the decisions that are made in their cases have been taken on grounds, which include potentially discriminatory criteria. The ongoing digital transformation seems to leave no time for the adequate *lex speciali* regulatory solutions to develop. Therefore, it is worth considering if the ones already existing cannot provide us with innovative answers to the new challenges, using their dynamic interpretation. I argue that when facing the challenges created by the automated decision-making solutions, the existing right to information can serve as a way of improving the current state of the art. Rethinking the character of the right to access information in the light of the debate on the right to explanation may be seen as a step towards an updated, dynamic interpretation of a well-known human rights acts' provision. In absence of solutions focused strictly on automated decision-making technologies, the right to access information sets the fundaments for a technologically neutral regulatory framework that may prove to be useful when preventing discriminatory treatment by technological solutions, which few seem to understand whilst all may be subjected to their decisions.

78. *Österreichische Vereinigung zur Erhaltung, Stärkung und Schaffung Eines Wirtschaftlich Gesunden Land- und Forst-Wirtschaftlichen Grundbesitzes* v. *Austria*, above n. 55, para. 47.
79. *Ibid.*
80. P. Casanovas, L. De Kokerl, D. Menderson, & D. Watts, 'Regulation of Big Data: Perspectives on Strategy, Policy, Law and Privacy', 7 *Health and Technology* 335, at 337 (2017).
81. M. Hildebrandt, 'The New Imbroglio – Living with Machine Algorithms', in L. Janssens (ed.), *The Art of Ethics in the Information Society. Mind You* at 56 (2016) https://bit.ly/2wn0b1I (last visited 8 May 2018).

Fostering Worker Cooperatives with Blockchain Technology: Lessons from the Colony Project

Morshed Mannan*

Abstract

In recent years, there has been growing policy support for expanding worker ownership of businesses in the European Union. Debates on stimulating worker ownership are a regular feature of discussions on the collaborative economy and the future of work, given anxieties regarding the reconfiguration of the nature of work and the decline of standardised employment contracts. Yet, worker ownership, in the form of labour-managed firms such as worker cooperatives, remains marginal. This article explains the appeal of worker cooperatives and examines the reasons why they continue to be relatively scarce. Taking its cue from Henry Hansmann's hypothesis that organisational innovations can make worker ownership of firms viable in previously untenable circumstances, this article explores how organisational innovations, such as those embodied in the capital and governance structure of Decentralised (Autonomous) Organisations (D(A)Os), can potentially facilitate the growth of LMFs. It does so by undertaking a case study of a blockchain project, Colony, which seeks to create decentralised, self-organising companies where decision-making power derives from high-quality work. For worker cooperatives, seeking to connect globally dispersed workers through an online workplace, Colony's proposed capital and governance structure, based on technological and game theoretic insight may offer useful lessons. Drawing from this pre-figurative structure, self-imposed institutional rules may be deployed by worker cooperatives in their by-laws to avoid some of the main pitfalls associated with labour management and thereby, potentially, vitalise the formation of the cooperative form.

1 Introduction

There has been a long-running policy-level discussion on the role of worker ownership and management of firms in the European Union.[1] Labour-managed firms (LMFs) are firms in which the suppliers of labour, rather than capital, have ultimate control rights in the governance of a firm, including the right to collectively hire and dismiss directors.[2] The suppliers of labour also receive the residual earnings of the firm on the basis of their labour input.[3] LMFs offer an appealing governance structure for firms due to their perceived positive effects on employee behaviour for firms[4] as well as high survival rates during times of recession.[5] From the workers' perspective, LMFs provide job security,[6] 'positive energy'[7] resulting from the knowledge that they work for their own benefit rather than for non-worker shareholders and act as 'sites of solidarity'[8] in a neoliberal economy where workers' rights are gradually eroded.[9] As a consequence, LMFs such as worker cooperatives have regained attention in recent times[10] in view of the anxieties regarding job quality, income inequality,

* Morshed Mannan, LLM (Adv.), PhD Candidate, Company Law Department, Institute of Private Law, Universiteit Leiden. I wish to thank the anonymous referees, the issue editors and my PhD supervisor for their helpful comments on drafts of this article. I wish to show my appreciation to my wife for her support. Usual disclaimers apply.

1. From improving working conditions to providing start-up support, administrative and accounting spaces as well as workspaces for self-employed persons, *see* EP Resolution, OJ 1983 C 128/51; A. Bock, L. Bontous, S. Figueiredo do Nascimento, & A. Szczepanikova, *The Future of the EU Collaborative Economy — Using Scenarios to Explore Future Implications for Employment* (2016), at 27.
2. G.K. Dow, 'The Theory of the Labor-Managed Firm: Past, Present, and Future', 89 *Annals of Public and Cooperative Economics* 65, at 65 (2018).
3. H. Hansmann, *The Ownership of Enterprise* (2000), at 11. Workers also contribute capital, but their decision-making and financial rights are not predicated on the extent of their capital contribution.
4. Cf. I. Basterretxea and J. Storey, 'Do Employee-Owned Firms Produce More Positive Employee Behavioural Outcomes? If Not Why Not? A British-Spanish Comparative Analysis', 56 *British Journal of Industrial Relations* 292 (2018); R. Brown, R. McQuaid, R. Raeside, M. Dutton, V. Egdell, J. Canduela, 'Buying into Capitalism? Employee Ownership in a Disconnected Era', forthcoming in *British Journal of Industrial Relations* (2018), doi:10.1111/bjir12309.
5. V. Pérotin, 'Workers' Cooperatives: Good, Sustainable Jobs in the Community', 2 *Journal of Entrepreneurial and Organizational Diversity* 34, at 40 (2013); J. Birchall and L.H. Ketilson, *Resilience of the Cooperative Business Model in Times of Crisis* (2009) at 7, 13-14.
6. I. Heras-Saizarbitoria, 'The Ties That Bind? Exploring the Basic Principles of Worker-Owned Organizations in Practice', 21 *Organization* 645, at 656, 658 (2014).
7. Basterretxea and Storey, above n. 4, at 300.
8. J. Itzigsohn and J. Rebón, 'The Recuperation of Enterprises: Defending Workers' Lifeworld, Creating New Tools of Contention', 50 *Latin American Research Review* 178, 189-90 (2015).
9. P. Raffaelli, 'Social and Solidarity Economy in a Neoliberal Context: Transformative or Palliative? The Case of an Argentinian Worker Cooperative', 5 *Journal of Entrepreneurial and Organizational Diversity* 33, at 34 (2016); X. de la Barra, 'Sacrificing Neoliberalism to Save Capitalism: Latin America Resists and Offers Answers to Crises', 36 *Critical Sociology* 635, at 655 (2010).
10. CICOPA-COOP, *The Future of Work: Where do Industrial and Service Cooperatives Stand?* (2018); M. Sandoval, 'Fighting Precarity with Co-operation? Worker Co-operatives in the Cultural Sector', 88 *New Formations* 51, at 62 (2016).

diminishing worker protections, and worker participation raised by the collaborative economy and the 'future of work'.[11]

Yet, LMFs continue to be relatively rare in developed economies compared to capital-managed firms (KMFs),[12] barring famous exceptions in regional economies such as that of the Basque country of Spain,[13] the Emilia Romagna region of Italy[14] and the Buenos Aires province of Argentina.[15] While interest in worker cooperatives has surged in South Korea[16] and certain states in the United States of America,[17] their number in all of these instances still remains in the hundreds. The most common reasons attributed for their relative scarcity are acquiring start-up capital, workers' apprehension about not being able to spread their investment risk,[18] the risk of absenteeism and free-riding on the efforts of other workers,[19] the inability to meet the high ideological and economic expectations set when the LMF was formed[20] and a perceived tendency to 'degenerate' into KMFs, by replacing retiring worker-members with employees in a bid to maximise individual member remuneration, thereby diminishing worker voice and losing its democratic character.[21] Degeneration is seen as a particularly acute concern when a worker cooperative tries to internationalise its operations.[22]

Taking its cue from Hansmann's hypothesis that organisational innovations may make labour management and ownership viable in previously untenable circumstances,[23] this article explores how organisational innovations, such as those embodied in the capital and governance structure of, can potentially facilitate the growth of LMFs. D(A)Os refer to organisations that rely on blockchain technology and smart contracts as their source of governance and respond to both digital and human input.[24] In recent years, D(A)Os and platforms to create D(A)Os have emerged as ways to coordinate the supply of capital and labour in a globally distributed manner.[25] An important aspect of creating such organisations has been the design of governance systems that align incentives in a manner that promotes high-quality input as well as active member participation. This has prompted an outpouring of interest in decentralised governance,[26] and consequently led to proposals which employ game theory and technology to achieve, *in abstracto*, the formation of organisations, the financing of projects and high-quality and active member participation. In essence, these proposals strive for *corporate governance-by-design*.[27] This bears a strong resemblance to the start-up and coordination issues faced by LMFs. It is hypothesised that LMFs, particularly those operating online workplaces, may draw beneficial lessons from these experiments in decentralised governance. This is the first study that seeks to bridge the gap between worker cooperative and blockchain technology.

To explore this hypothesis, this article is structured as follows. The second section of the article elaborates on the governance structure of an archetypical LMF, a worker cooperative,[28] their main advantages according

11. T. Balliester and A. Elsheikhi, 'The Future of Work: A Literature Review', *International Labour Office Research Department Working Paper* 2018: 29, at 20, 26-27, 33.
12. F. Fakhfakh, V. Pérotin & M. Gago, 'Productivity, Capital and Labor in Labor-Managed and Conventional Firms: An Investigation on French Data', 65 *ILR Review* 847, at 850 (2012).
13. S.P. Thompson, 'Is the Mondragón Co-operative Experience a Cultural Exception? The Application of the Mondragón Model in Valencia and Beyond', 47 *Journal of Co-operative Studies* 19, at 19 (2014).
14. S. Zamagni and V. Zamagni, *Cooperative Enterprise: Facing the Challenge of Globalization* (2010), at 58.
15. P. Ranis, 'Argentine Worker Cooperatives in Civil Society: A Challenge to Capital-Labor Relations', 13 *WorkingUSA: The Journal of Labor and Society* 77, at 83 (2010).
16. M. Ji, 'The Worker Cooperative Movement in South Korea: From Radical Autonomy to State-Sanctioned Accommodation', 59 *Labor History* 415, at 428 (2018).
17. California, Massachusetts, New York, Ohio, Vermont, Washington and Wisconsin being particularly prominent. *See*, A. Johnson and M. Hoover (eds.), *Democracy at Work: U.S. Directory of Worker Cooperatives & Guide to Democratic Business Resources* (2015), at 10, 74-78.
18. J.M. Podivinsky and G. Stewart, 'Why is Labour-Managed Firm Entry So Rare? An Analysis of UK Manufacturing Data', 63 *Journal of Economic Behavior & Organization* 177, at 188 (2007); J.M. Podivinsky and G. Stewart, 'Modeling Proportions: Random Effects Models of UK Firm Entry', 54 *The Singapore Economic Review* 367, at 374 (2009).
19. Basterretxea and Storey, above n. 4, at 302-3, 307-8.
20. Cf. S. Arando, M. Gago, D.C. Jones, T. Kato, 'Efficiency in Employee-Owned Enterprises: An Econometric Case Study of Mondragon', 68 *ILR Review* 398, at 417, 421 (2015). They find that LMFs can be highly demanding and stressful workplaces due to (self-imposed) high expectations of their work.
21. This is an argument that has been made for over a century, starting with B. Potter, *The Cooperative Movement in Great Britain* (1891). An overview of the degeneration thesis is provided in K. Langmead, *Exploring the Performance of Democracy and Economic Diversity in Worker Cooperatives* (2017), at 24-27.
22. Cf. I. Bretos, A. Errasti & C. Marcuello, 'Ownership, Governance, and the Diffusion of HRM Practices in Multinational Worker Cooperatives: Case-Study Evidence from the Mondragon Group', 28 *Human Resource Management Journal* 76, at 76-77, 81-82, 85 (2018); P. Battilani and H.G. Schröter, 'Conclusion: The Decisive Factors of Cooperatives' Future – Their Nature, Longevity, Role, and Environment', in P. Battilani and H.G. Schröter (eds.) *The Cooperative Business Movement, 1950 to the Present* (2012), at 266-7.
23. H. Hansmann, 'When Does Worker Ownership Work? ESOPs, Law Firms, Codetermination and Economic Democracy' 99 *The Yale Law Journal* 1749, at 1816 (1990). These untenable circumstances are discussed in Section 2.4 on the scarcity of worker cooperatives.
24. Cf. Most recently, P. De Filippi and A. Wright, *Blockchain and the Law* (2018); P. Hacker and C. Thomale, 'Crypto-Securities Regulation: ICOs, Token Sales and Cryptocurrencies under EU Financial Law' (2018) https://ssrn.com/abstract=3075820; I.M. Barsan, 'Legal Challenges of Initial Coin Offerings (ICO)', 3 *Colloque* 54 (2017).
25. S. Davidson, P. De Filippi & J. Potts, 'Blockchains and the Economic Institutions of Capitalism', 14 *Journal of Institutional Economics* 639, at 643 (2018).
26. Cf. W. Reijers, F. O'Brolcháin & P. Haynes, 'Governance in Blockchain Technologies & Social Contract Theories', 1 *Ledger* 134 (2016); M. Atzori, 'Blockchain Technology and Decentralized Governance: Is the State Still Necessary?' (2016), https://ssrn.com/abstract=2709713.
27. This is distinct from public regulation by design and privacy by design, discussed in D.K. Mulligan and K.A. Bamberger, 'Saving Governance-By-Design', 106 *California Law Review* 697 (2018). Corporate governance by design is of legal and political interest as such technological innovations can shape public orders in lasting ways. *See* L. Winner, 'Do Artifacts Have Politics?', 109 *Daedalus* 121, at 128 (1980).
28. As with most corporate entity forms, there are jurisdictional differences in the characteristics of a worker cooperative. Therefore, this archetype is based on the Principles of European Cooperative Law (PECOL) which were published in 2017 and are derived from a synthesis of the cooperative laws of the UK, Finland, France, Germany, Italy, Portugal, Spain and the EU. G. Fajardo, A. Fici, H. Henry, D. Hiez, D. Meira, H.-H. Muenker & I. Snaith. *Principles of European Cooperative Law:*

to theoretical and empirical literature and the policy-level support for their growth, which has gained urgency with the emergence of the platform-mediated, collaborative economy. This section is concluded with a consideration of the central causes of the scarcity of LMFs. The third section of the article provides a brief overview of smart contracts and D(A)Os, as they are key to understanding the governance and incentive system of decentralised organisations. The fourth section presents a case study of one D(A)O platform, Colony, created by Collectively Intelligent Ltd. that seeks to create decentralised, open, self-organising companies where decision-making power is intertwined with high-quality labour input. The case study was conducted by reviewing Colony's legal and technical documentation, software development platform (Github), social media posts and presentations through which information about the project is shared.[29]First, the aspirations of the Colony project are mentioned, along with its proposed governance structure. Second, its governance features are assessed against that of a worker cooperative. This permits a tentative analysis of the Colony protocol's potential to address some of the perceived governance shortcomings of worker cooperatives, particularly when operating across borders. In view of this sample governance structure, self-imposed institutional rules may be deployed by worker cooperatives in their by-laws to avoid some of the main pitfalls associated with labour management[30] and thereby vitalise the use of an alternate form of business organisation. The fifth section sums up and concludes.

2 Labour Management and Ownership of Businesses

2.1 The Archetypical LMF: The Worker Cooperative

In a bid to distinguish cooperatives from other legal entity forms, the International Co-operative Alliance (ICA), a representative body of the international cooperative movement, and the International Labour Organization (ILO) promote a set of core values and principles integral to the cooperative identity. All cooperatives, including worker cooperatives, value 'self-help, self-responsibility, democracy, equality, equity and solidarity; as well as ethical values of honesty, openness, social responsibility and caring for others'.[31] This is implemented through seven principles: (1) voluntary and open membership; (2) democratic member control; (3) member economic participation; (4) autonomy and independence; (5) education, training and information; (6) cooperation among cooperatives; and (7) concern for the community.[32] In particular, worker cooperatives seek to create and maintain sustainable jobs and wealth, which will dignify human work, improve worker-members' quality of life, allow democratic self-management and enable local and community development.[33] This is reflected in the capital and governance structure of worker cooperatives.

In a worker cooperative, most, if not all, of the capital of these firms is held by worker-members.[34] While worker cooperatives are generally permitted to have non-member employees, this is usually set at a low threshold and employees are often given the option of becoming members.[35] To become a member, an employee must not only complete a certain amount of hours of work (*i.e.* a probation period) but must usually contribute a 'buy in' to the cooperative as well, which may be redeemable at face value upon exit from the cooperative.[36] As the purpose of the business is to undertake economic activities in the interest of its worker-members, rather than to make a profit for the cooperatives itself or external investors,[37] cooperatives make allocations to mandatory and voluntary reserves from their cooperative transactions (*i.e.* surplus of revenue over costs) and profitable non-cooperative transactions (*e.g.* holding shares in other companies).[38] Most often, surplus, if discretionarily distributed as refunds, is received by members in proportion to their work (measured in hours worked) for the worker cooperative.[39] In the event of a loss being incurred, they are first covered through the reserves of the cooperative before turning to the members, in proportion to 'the quantity and/or quality of their participation in cooperative transactions within the limit of the value of the goods and services received'.[40] In case of business failure, as the assets and reserve of the worker cooperative are commonly held, if the worker cooperative is liquidated, the residual net assets are distributed according to the principle of disinterested distribution, that is, to associated cooperatives or the community.[41]

Principles, Commentaries and National Reports (2017), at 2-4. It also incorporates the description of V. Perotin, 'What Do We Really Know About Workers' Co-Operatives?', in A. Webster, L. Shaw & R. Vorberg-Rugh, *Mainstreaming Co-operation: An Alternative for the Twenty-First Century?* (2016).

29. The author also had conversations with two of the authors of the Colony White Paper, Jack du Rose and Dr. Aron Fischer, about the project.
30. Dow, above n. 2, at 76.
31. Art. 3(a), ILO Recommendation 193 concerning the Promotion of Cooperatives, 2002.
32. ICA, Statement on the Cooperative Identity, 1995; ICA, *Guidance Notes to the Co-operative Principles*, 2015.
33. CICOPA-COOP, *World Declaration on Worker Cooperatives* (2005), at 2.
34. Section 3.1, PECOL acknowledges the possibility that cooperatives can 'use shares, reserves, loans and other financial instruments as sources of capital, providing they are compatible with their cooperative nature'.
35. Section 1.5(3), PECOL. In some jurisdictions, like the UK, it is mandatory for individuals who are eligible (*i.e.* have worked a minimum number of hours) to be offered membership. Footprint Workers' Co-operative Ltd. and Seeds for Change Lancaster Co-operative Ltd., *How to set up a Workers' Co-op*, 4th ed. (2015), at 110.
36. Sections 3.2(2), 3.3, PECOL.
37. Section 1(1) PECOL.
38. Sections 3.6-3.7, PECOL.
39. Section 3.6(3)(a), PECOL.
40. Section 3.6(6)(b), PECOL. This is in keeping with members' limited liability under Section 3.5, PECOL.
41. Section 3.8(2), PECOL. *Also see* Fajardo *et al.*, above n. 28, at 94. This requirement has helped LMFs avoid the theorised problem of underinvestment (*i.e.* a horizon problem) – workers choosing to maximise the

These firms share the characteristic of providing worker-members a voice in governance,[42] either on a one-member, one-vote basis or based on the extent of their non-capital contribution.[43] In many of these firms, delegated management still exists, but the directors are elected by workers and the latter retain an extensive right to ask questions and be informed and consulted.[44] In some cases, they may have the right to vote on issues of major corporate interest.[45] In certain firms, members may be involved in a range of strategic decisions, from setting trading hours to exploring new markets to introducing a product.[46] What is notable, however, is that it appears that there is a risk for worker participation to become more shallow as cooperatives internationalise.[47]

While worker cooperatives continue to be marginal organisational forms in developed economies, the appeal of worker cooperatives endures. An estimated 11 million people presently work in such cooperatives as worker-members.[48] Across the globe, they are present in a variety of industries, from sheet metal factories[49] to media,[50] from the cultural sector[51] to cutting-edge ICT.[52] In France[53] and Italy,[54] there is a relatively high proportion of worker cooperatives in manufacturing and construction respectively. However, the predominant view is that capital-intensive sectors, involving tasks with a high degree of standardisation, will continue to be predominated by KMFs while those in which personal relations and human creativity feature heavily are more amenable to worker ownership and management.[55] This coincides with the view of organisational theorists, who observe that those engaged in knowledge-intensive work tend to be less indifferent about hierarchical employment relations and believe that 'the locus of decisions has to coincide with the locus of knowledge'.[56]

2.2 The Appeal of Worker Cooperatives to Workers

From the non-executive workers' perspective, worker cooperatives hold the promise of lower wage differentials than KMFs[57] and improved benefits, such as collective private health insurance.[58] Based on cross-cultural evidence, it would appear that LMFs also provide stronger guarantees of employment stability, as LMFs tend to prefer reducing hours of work, rather than laying off worker-members, in response to recessions.[59]

An ideal-type worker cooperative allows workers an involvement in organisational decision-making that goes far beyond the voluntarist human resource management practices (*e.g.* agile management) used by KMFs.[60] Along with being given a voice in production processes, workers are also given a say in key governance decisions, which reduces information asymmetry between labour and management. Instead of viewing workers as a monolithic group with uniform interests, individual preferences and views can be better communicated. In short, as workers hire managers, rather than the other way round, labour management and ownership avoids the dishonouring of workplace bargains[61] – such as the unilateral termination of certain rights to voice. This allows workers to develop, simultaneously, a sense of self-determination in how they work[62] and solidarity with each other.[63] This is manifested in how worker coopera-

firm's present value instead of pursuing long-term gain. *See* Fakhfakh, Pérotin & Gago, above n. 12, at 855.

42. Section 2.3(4)(b), PECOL.
43. Section 2.4(8)(a), PECOL.
44. Potentially extending beyond the minimum information and consultation rights ordinarily enjoyed by workers in the EU under Directive 2002/14/EC, OJ 2002 L 80/29, industry-specific legislation and legislation concerning changes of corporate control.
45. B. Bakaikoa, A. Errasti and A. Begirstain, 'Governance of the Mondragon Corporación Cooperativa', 75 *Annals of Public and Cooperative Economics* 61, at 68 (2004).
46. Cf. A. Cathcart, 'Directing Democracy: Competing Interests and Contested Terrain in the John Lewis Partnership', 55 *Journal of Industrial Relations* 601, at 611 (2013); S. Hernandez, 'Striving for Control: Democracy and Oligarchy at a Mexican Cooperative', 27 *Economic and Industrial Democracy* 105, at 122 (2006).
47. Particularly if that host state does not have a solid cooperative tradition. Bretos, Errasti & Marcuello, above n. 22, at 82.
48. CICOPA, *Industrial and Service Cooperatives: Global Report 2015-2016*, at 9 (2017).
49. S. Jaumier, 'Preventing Chiefs from Being Chiefs: An Ethnography of a Co-Operative Sheet-Metal Factory', 24 *Organization* 218 (2017).
50. In Greece, there are examples of cooperatives newspapers (*e.g.* Efsyn), online media (*e.g.* Alterthess) and radio stations (*e.g.* Flash FM). E. Siapera and L. Papadopoulou, 'Entrepreneurialism of Cooperativism?', 10 *Journalism Practice* 178, at 185 (2016).
51. One of the leading symphony orchestras in the world, the London Symphony Orchestra, is a LMF and has been so for over a hundred years. C.P. Mulder, *Transcending Capitalism Through Cooperative Practices* (2015), at 35-37.
52. RChain Coop is a cooperative building a blockchain platform, www.rchain.coop/.
53. Fakhfakh, Pérotin & Gago, above n. 12, at 852.
54. J. Pencavel, L. Pistaferri, and F. Schivardi, 'Wages, Employment, and Capital in Capitalist and Worker-Owned Firms', 60 *Industrial and Labor Relation Reviews* 23, at 28 (2006).
55. V.N. Zamagni, 'The Co-operative Enterprise: A Valid Alternative for a Balanced Society', in S. Novkovic and T. Webb, *Co-operatives in a Post Growth Era* (2014), at 196; Dow, above n. 2, at 78.
56. A. Grandori, 'Knowledge-Intensive Work and the (Re)emergence of Democratic Governance', 30 *Academy of Management Perspectives* 167, at 173 (2016).
57. C. Heales, M. Hodgson & H. Rich, *Humanity at Work: Mondragon, a Social Innovation Ecosystem Case Study* (2017), at 51; G.K. Dow, *Governing the Firm: Workers' Control in Theory and Practice* (2003), at 76.
58. Mulder, above n. 51, at 42.
59. Dow, above n. 2, at 74, summarising evidence from the USA, Italy and Uruguay.
60. T. Dobbins and T. Dundon, 'The Chimera of Sustainable Labour-Management Partnership', 28 *British Journal of Management* 519 (2017).
61. *Ibid.*, at 521-2; P. Thompson, 'Financialization and the Workplace: Extending and Applying the Disconnected Capitalism Thesis', 27 *Work, Employment and Society* 472, at 478-479 (2017).
62. Having more decision-making powers allows workers to develop a feeling of being trusted. *See* B.S. Frey and R. Jegen, 'Motivation Crowding Theory', 15 *Journal of Economic Surveys* 589, at 601 (2001); T. Ellingsson and M. Johannesson, 'Paying Respect', 21 *Journal of Economic Perspectives* 135, at 139 (2007); V.H. Bernstrøm and H. Svare, 'Significance of Monitoring and Control for Employees' Felt Trust, Motivation, and Mastery' 7 *Nordic Journal of Working Life Studies* 29, at 43 (2017). The authors also note how worker perceptions of being monitored due to a managerial fear of shirking can engender unpleasant feelings and counterproductive behaviour.
63. M. Parker *et al.*, 'Imagining Alternatives', in M. Parker *et al.* (eds.), *Routledge Companion to Alternative Organizations* (2014) 31, at 32, 36-37. The Editors of this book see worker cooperatives as one of the

tives, and LMFs in general, are able to account for quality-of-life issues and individual and team well-being.[64]

As a consequence, it is easy to understand why labour management and ownership has gained particular resonance in the context of the 'collaborative economy', given the effects it has had on the nature of work.[65] The actors in this space include individuals providing services, users of these services and the online platforms that mediate their interactions by offering access and executing tripartite contracts.[66] Economic theorists have characterised such online platforms as being multisided markets[67] which enable value-creating transactions by facilitating service providers and users finding each other and developing interdependence. In a labour intermediation platform, such as Etsy or Uber, the greater the number of workers on the platform, the more that platform appeals to other workers (*i.e.* a direct network effect). Conversely, the presence of a large number of potential clients persuades more workers to join the platform (*i.e.* an indirect network effect).[68]

The collaborative economy accounted for 26.5 billion EUR in gross revenue in 2016 and created approximately 394,000 jobs across the European Union member states.[69] While creating employment opportunities and consumer value, from the perspectives of those who work on, or through these platforms, they create a downward pressure on permanent, full-time, subordinated employment relationships towards nonstandard employment and self-employment.[70] This creates new pressures on worker representation institutions, such as trade unions and works councils, that have been built around the employment relationship.[71] This reversion to pre-twentieth century employment practices serves some well,[72] particularly those who have highly coveted skills and scope for job mobility, but it exposes many others to job precarity and income insecurity.[73] This trend can also be seen as cynical exploitation of workers' own frustrated desires for freedom and self-determination.[74]

Firms representing such cooperative qualities have begun to emerge in the collaborative economy, with the ambition of providing less precarious workplaces and more broadly accountable organisations.[75] These platforms put the interest of the user-members at the forefront, by involving them in the financing and management of the platforms. These range from cooperative platforms like Doc Servizi,[76] a 8,000-person creative workers' cooperative in Italy, to Stocksy,[77] a platform cooperative that accepts and provides royalty-free stock footage.

2.3 Worker Cooperatives as Competitive Firms

In addition to these potential benefits for worker-members, worker cooperatives are also competitive businesses in their own right. Agency theory suggests that worker ownership aligns the economic interests of the organisation and individual workers, thereby promoting productivity and organisational loyalty.[78] This is in contrast to KMFs where information asymmetries and differing interests may lead to a fear that employment bargains will be reneged at a future date or that optimal firm-specific investments will not be made by either labour or management.[79] Providing feedback and suggestions on production processes allows firms to benefit from the workers' experience and knowledge of the technology, organisation and market environment.[80] Moreover, the costs of monitoring diminish, in comparison to KMFs, as workers are incentivised to monitor each other.[81] Going beyond agency theory, motivation crowding theory suggests that feelings of independence and self-governance can act as intrinsic motivation to work in the interest of the organisation, even where there may be little or no direct financial reward on offer.[82] This is of particular relevance in knowledge-intensive and creative

alternative organisations that can potentially embody the principles of autonomy, solidarity and responsibility.

64. M. Atzeni and M. Vieta, 'Between Class and Market: Self-management in Theory and in the Practice of Worker-Recuperated Enterprises in Argentina', in M. Parker *et al*. (eds.), *Routledge Companion to Alternative Organization* (2014) 47, at 56. The authors highlight how workers are able to modulate production in keeping with the needs of the team.
65. A. Ben-Ner, 'The Life-Cycle of Worker-Owned Firms in Market Economies', 10 *Journal of Economic Behavior and Organization* 287, at 296 (1988). Ben-Ner hypothesised that organisational and technological innovations that affect the workplace would drive the demand for worker-owned firms. According to the EU Agenda for the Collaborative Economy, the term 'refers to the business models where activities are facilitated by collaborative platforms that create an open marketplace for the temporary usage of goods or services often provided by private individuals'. EC Communication, 'A European Agenda for the Collaborative Economy', COM (2016) 356 final, at 3.
66. V. Hatzopoulos, *The Collaborative Economy and EU Law* (2018), at 7.
67. D.S. Evans and R. Schmalensee, *Matchmakers* (2016), at 8.
68. Hatzopoulos, above n. 66, at 9-10.
69. Technopolis Group, VVA Consulting and Trinomics, *Study to Monitor the Economic Development of the Collaborative Economy at Sector Level in the 28 EU Member States* (2018), at 12.
70. CICOPA-COOP, *The Future of Work: Where do Industrial and Service Cooperatives Stand?* (2018), at 11.
71. J. Prassl, *Collective Voice in the Platform Economy: Challenges, Opportunities, Solutions* (2018), at 14.
72. S. Deakin, 'The Contract of Employment: A Study in Legal Evolution', 11 *Historical Studies in Industrial Relations* 1, at 29 (2001).
73. This can range from manual labourers to creative workers, cutting across generations and disproportionately affecting women. G. Standing, *The Precariat* (2011), at 59; U. Huws, 'iCapitalism and the Cybertariat: Contradictions of the Digital Economy', *Monthly Review*, 1 January 2015.
74. P. Frase, 'Beyond the Welfare State', *Jacobin*, 11 December 2014; E. Chiapello, 'Evolution and Co-optation: The "Artist Critique" of Management and Capitalism', 18 *Third Text* 585, at 593 (2004).
75. N. Schneider, 'An Internet of Ownership: Democratic Design for the Online Economy', 66 *The Sociological Review Monographs* 320 (2018).
76. www.docservizi.it/(last visited 1 December 2018).
77. www.stocksy.com (last visited 1 December 2018).
78. J.P. Bonin, D.C. Jones & L. Putterman, 'Theoretical and Empirical Studies of Producer Cooperatives: Will Ever the Twain Meet?', 31 *Journal of Economic Literature* 1290, at 1303 (1993); G. Nuttall, *Sharing Success: The Nuttall Review of Employee Ownership* (2012), at 22-28.
79. Ben-Ner, above n. 65, at 293.
80. Dow, above n. 2, at 77.
81. This fundamentally differs from hierarchical monitoring as worker cooperatives preserve the right of individual members to challenge authority and commands. *See* Jaumier, above n. 49, at 223.
82. Frey and Jegen, above n. 62, at 595, 597-8.

industries where workers may have to work extra hours, without compensation, to complete a project.[83]

The recent empirical evidence on this offers a nuanced picture of the commercial benefits of labour management and ownership and the conditions needed to achieve it. One study that compared sales per employee between 300 US firms that are majority or fully employee owned, with similarly sized comparator firms that are investor owned, substantiates the idea that growth in employee stake in firms and influence in decision-making lead to improvements in productivity.[84] Another study, examining a panel of 7,000 French firms, 500 of which were employee owned, reveals that worker cooperatives (SCOPs) in France are as productive, if not more, than KMFs.[85] The fact that worker cooperatives prioritise job stability means that they are willing to introduce wage flexibility, if it will ensure the survival of the firm.[86] However, in a longitudinal study of two of the largest employee-owned retailers in Europe, the John Lewis Partnership and Eroski, it was found that the former had lower absenteeism and higher job satisfaction rates among worker-members than their capital-managed counterparts, while the latter had higher absenteeism rates and lower job satisfaction rates. The authors of the study attribute this to differences in the quality of management across the two firms; in balancing the need to respond to crises with agility and decisiveness, with the goal of invigorating and implementing a culture of shared ownership.[87] While workers in LMFs may be willing to take on more responsibility, a lack of vigilance in monitoring performance and ineffectively communicating business needs – including engaged member participation – may hamper these goals.

It is for these perceived advantages that worker ownership has long received policy-level attention at the European level. During the 1980s and 1990s, the European Parliament recognised the role of cooperatives in improving working conditions,[88] regional development through job creation and preservation in local communities[89] as well as contributing to women's integration into the workplace.[90] In view of this, the Parliament called for, *inter alia*, investigations into how the formation of worker cooperatives can help rescue distressed businesses[91] and for incentives to be 'provided for innovative sectors and that steps should be taken to facilitate access by women to new technologies'.[92] In parallel to these developments, the idea of creating a transnational European cooperative was also promoted, the origin of which dates back to the earliest consultations on establishing a European commercial company in the 1960s.[93] It was noted in policy discussions, and subsequently in the recitals of the European Cooperative Society (SCE) Regulation, that cross-border cooperation between cooperatives was inhibited by legal and administrative barriers – given the lack of harmonisation of national cooperative laws – and that the community was 'anxious to ensure equal terms of competition' for cooperatives with limited liability companies.[94] Following the enactment of the SCE Regulation, the European Commission issued a far-reaching Communication[95] to promote the visibility and use of cooperatives. More recently, the role that cooperatives may have in providing start-up support, administrative and accounting spaces as well as workspaces for self-employed persons was particularly noted in a 2016 study commissioned by the European Commission.[96] The European Parliament has also observed the interest in developing cooperative alternatives to collaborative economy companies.[97]

Notwithstanding the appeal of worker cooperatives and their positive reception, it still remains difficult for entrepreneurs to establish cooperatives, nationally and especially transnationally, in comparison to KMFs. The next section discusses this further.

2.4 The Scarcity of Worker Cooperatives

There has been theoretical and empirical research into the reasons for the scarcity of worker cooperatives and other LMFs for at least sixty years.[98] Over this period, a number of hypotheses have been tested, most notably – whether worker-members tend to underinvest in the firm ('horizon problem'), whether workers are less productive ('shirking' and 'free-riding' problems), whether members seek to replace exiting members with employees so as to maximise individual refunds ('degeneration problem') and whether there are fewer LMFs being born in comparison to KMFs ('birth rate problem'). As indicated by the empirical research described in Section 2.3, it would appear that worker cooperatives are not inherently dysfunctional. They have the capacity to be as productive as KMFs and have high survival rates. In contrast to the shibboleth that worker cooperatives inevitably degenerate into KMFs, researchers have

83. Cf. A. Alacovska, 'Informal Creative Labour Practices: A Relational Work Perspective', 71 *Human Relations* 1563, at 1585-1586 (2018). Alacovska offers a relational perspective on creative labour practices, emphasising how feelings of friendship and kinship motivate non/under-remunerated work.
84. B. Kramer, 'Employee Ownership and Participation Effects on Outcomes in Firms Majority Employee-Owned Through Employee Stock Ownership Plans in the US', 31 *Economic and Industrial Democracy* 449, at 466-467 (2010).
85. In the printing and publishing, paper and wood industries, worker cooperatives have been found to be more productive (in terms of output) than KMFs. Fakhfakh, Pérotin & Gago, above n. 12, at 867.
86. *See* G. Burdín, 'Are Worker-Managed Firms More Likely to Fail Than Conventional Enterprises? Evidence from Uruguay', 67 *ILR Review* 202, at 224, 226 (2015).
87. Basterretxea and Storey, above n. 4, at 315-7.
88. EP Resolution, OJ 1983 C 128/51.
89. Recital 12 EP Resolution, OJ 1994 C 61/231; Recitals 3-4 EP Resolution, OJ 1987 C 246/94.
90. EP Resolution, OJ 1998 C 313/234; EP Resolution OJ 1989 158/380.
91. Recital 3, EP Resolution, OJ 1983 C 128/51.
92. Recital 3, EP Resolution, OJ 1998 C 313/234.
93. C. Chomel, 'The Long March of the European Cooperative Society', *Recma*, no. 291, 1, at 2 (2004).
94. Recitals 6 and 11 Regulation (EC) 1435/2003, OJ 2003 L 207/1.
95. EC Communication on the promotion of co-operative societies in Europe, COM (2004) 18 final.
96. Bock *et al.*, above n. 1, at 27.
97. Recital 11 EP Resolution on a European Agenda for the collaborative economy (2017/2003(INI)).
98. Starting with B. Ward, 'The Firm in Illyria: Market Syndicalism', 48 *American Economic Review* 566 (1958).

found that time-tested cooperatives undergo periods of cyclical degeneration and regeneration.[99] In areas where they do have shortcomings – such as lower average wages compared to peers in comparable KMFs[100] – it can often be attributed to the fact that worker cooperatives are different by design from their capitalist counterparts. For instance, empirical research in Italy has found that worker cooperatives have (marginally) lower and more volatile wages compared to peers in comparable KMFs. This is complemented with having more stable employment.[101] It would therefore seem that worker cooperatives prioritise stability and retention of members over wage certainty.

Instead, at present, it would appear that the two major reasons for the scarcity of worker cooperatives is a very low birth rate[102] and, if and when created, coordination problems as the entity scales across borders.

The low birth rate has three major factors: a lack of information about the worker cooperative option, the lack of a conducive legal environment and scarcity of financing options.[103] An example can illustrate how visibility continues to be a pertinent problem for potential cooperators. A recent study commissioned by the European Commission acknowledges the importance of digital tools in supporting the platform-mediated labour market, and noted instances of good practices that include platform cooperatives,[104] yet the new Proposal for a Directive regarding the use of digital tools and processes in company law falls short in making the cooperative form a visible and viable alternative for entrepreneurs. If the Proposal is adopted in its current form, member states will only be required to provide online templates of company constitution instruments for company forms mentioned in a proposed Annex IIA, such as the UK Private Company Limited by Shares or Guarantee. The provision of templates for other limited liability company forms,[105] such as a cooperative,[106] remains optional.[107] This appears to be the result of path dependence – as entrepreneurs have shown a preference for the company forms specified in Annex IIA – yet this may make such entities a default choice, especially for start-ups. In short, cooperatives and companies will no longer be in equal competition, as set out in the aforementioned recitals of the SCE Regulation.

This lack of familiarity with the worker cooperative form also makes it difficult to finance their formation. In the absence of sufficient collateral, the workers' own savings or loans from friends and family, worker cooperatives traditionally have difficulty in obtaining debt financing. As a consequence of legal regulation and/or ideological principle, worker cooperatives can only accept limited non-member equity investment.[108] In any case, conventional financiers, such as private equity funds, are dissuaded from investing in worker cooperatives as they are not profit-oriented and the requirement to be majority member-controlled inhibits the grant of substantial equity positions to external investors. Instead, they often have to rely on a single, large private customer,[109] a sympathetic public authority[110] and/or community contributions, through mechanisms such as crowdfunding.[111] (Admittedly, the quality and value of LMF membership is hard to estimate even for the most ideologically committed capital contributor.[112]) This financing challenge is also seen as one of the major deterrents to the formation of SCEs,[113] as a minimum capital of EUR 30,000 is required,[114] which is beyond the scope of many small businesses that may wish to operate across borders.[115]

Turning to the coordination issues that occur upon the formation of worker cooperatives, collective action theory suggests that the heterogeneous preferences of equal worker-members make it difficult to arrive at decisions expeditiously.[116] Competing with capitalist firms means that there are time constraints on decision-making and worker-members may not respond to the market rapidly enough.[117] This is borne out by the studies on the larger worker cooperatives, such as Eroski, discussed in Section 2.3.[118] In view of this, worker-members have to work longer hours, under more stress, with serious consequences for their own health.

This coordination problem is accentuated as cooperatives scale or internationalise. With advances in modern technology, such as those discussed in Section 3, it is

99. C. Cornforth, 'Patterns of Cooperative Management: Beyond the Degeneration Thesis', 16 *Economic and Industrial Democracy* 487, at 494 (1995); Y. Stryjan, 'Understanding Cooperatives: The Reproduction Perspective', 65 *Annals of Public and Cooperative Economics* 59, at 62-65 (1994).
100. J. Pencavel, L. Pistaferri & F. Schivardi, 'Wages, Employment, and Capital in Capitalist and Worker-Owned Firms', 60 *Industrial and Labor Relations Review* 23 (2006).
101. *Ibid*.
102. Dow, above n. 2, at 78.
103. Ben-Ner, above n. 65, at 289-90. This is particularly true when worker cooperatives are formed 'defensively' – as a last resort by workers to prevent business closure and maintain jobs. T. Kerswell and S. Pratap, *Worker Cooperatives in India* (2019), at 80. This makes the durability of Argentina's *empresas recuperadas* (worker-recuperated enterprises) all the more remarkable.
104. Bock *et al*., above n. 1.
105. The broader ambit of this term can be seen in Art. 119(1) Directive (EU) 2017/1132, OJ 2017 L 169/46.
106. The fact that Directive (EU) 2017/1132, OJ 2017 L 169/46 explicitly countenances cooperatives qualifying as a limited liability company form is clear from Art. 120(2).
107. Art. 13(g), Proposal for a Directive amending Directive (EU) 2017/1132 as regards the use of digital tools and processes in company law.
108. K. Mikami, 'Cooperatives, Transferable Shares and a Unified Business Law', 87 *Annals of Public and Cooperative Economics* 365, at 374 (2016).
109. Jaumier, above n. 49, at 219.
110. Mulder, above n. 51, at 83-86.
111. Community Wealth Building, *An Introduction to Financing for Cooperatives, Social Enterprises, and Small Businesses* (2015), at 12-14.
112. Dow, above n. 2, at 79.
113. Only 41 are in operation as of 2018. *See* Libertas Institut, 25 August 2018, www.libertas-institut.com/wp-content/uploads/2018/08/sce-list.pdf.
114. Art. 3(2), Regulation (EC) 1435/2003, OJ 2003 L 207/1.
115. A. Fici, 'The European Cooperative Society Regulation', in D. Cracogna *et al*. (eds.), *International Handbook of Cooperative Law* (2013) 115, at 120, 145, 149.
116. Hansmann, above n. 23, at 1772-1779.
117. Atzeni and Vieta, above n. 64, at 53.
118. Basterretxea and Storey, above n. 4.

possible for workers to cooperate across borders even if their enterprise is small in scale. In certain sectors, like the creative and tech industry, it is difficult to avoid as the workplace is globalised.[119] However, coordinating such business practices in a distributed manner, without the use of a third-party platform intermediary, involves high transaction costs. The evidence from the few worker cooperatives that have grown in scale[120] and internationalised[121] their operations indicates a negative trend in participatory management, mutual monitoring and solidarity. It has been seen that contrasting cooperative cultures and restrictive legislation on worker organising in the host state inhibit the replication of cooperative practices.[122]

Having canvassed the appeal and drawbacks of worker cooperatives, the remainder of the article explores how the organisational innovations developed by D(A)O platforms would potentially address some of these start-up and coordination problems. This analysis is predicated on the understanding of blockchain as an institutional technology, which can coordinate economic activity in novel ways.[123] To do so, the next section sketches how smart contracts and D(A)Os work, before presenting a particular D(A)O platform and the governance structure it has designed for D(A)Os created through its platform.

3 Understanding the Technology: Smart Contracts and D(A)Os

Developers of D(A)Os[124] draw inspiration from transaction cost economics and the nexus of contracts theory of corporations, where the corporation is viewed as a 'complex set of contracts among managers, workers, and contributors of capital' that mediate relationships in a hierarchical structure to internalise and diminish transaction costs.[125] This is reflected in their belief that decentralised (autonomous) organisations can emerge from a complex set of 'smart contracts'. Smart contracts are software deployed on a blockchain (most famously, Ethereum) which, for a small transaction fee ('gas'), is capable of receiving and storing cryptocurrency (*e.g.* 'Ether') and tokenised representations of assets. They also contain conditions subject to which an exchange of assets and transactions will take place (*e.g.* passage of time, a certain event). As such, a smart contract can act as an escrow account, as well as automate certain functions of ordinary contracts. A simple example of a smart contract involves a transfer of cryptocurrency for an asset. Once the payment is made to the smart contract, for the contract to be executed, the nodes of the blockchain will verify that the transferees' wallets respectively hold the claimed sum of cryptocurrency and the asset. If validated, the smart contract will receive a message to automatically self-execute and the exchange will take place. The blockchain will then be updated to reflect the transfer of asset ownership as well as the change in cryptocurrency amounts in the participants' wallets.[126] As a result, third parties – whether they be title registries or courts – are not required to enforce the transaction. Unless the smart contract has a dispute resolution 'safety valve' built in, the parties will not be able to stop the performance of the contract.[127] Moreover, smart contracts do not need to be triggered ('called') by human parties to a contract but can also respond to inputs from off-chain third parties (oracles) that a certain event has occurred.

Following the creation of smart contracts, the idea soon arose of an algorithmically governed organisation which responds automatically to inputs from both digital and analogue sources.[128] The organisation would be composed of a collection of smart contracts which would have internal capital, discourage collusion among members, focus on automating transactions and, ultimately, have a peripheral role for human involvement. This idea was operationalised through the creation of The Decentralized Autonomous Organization (The DAO), for the purpose of decentralised crowdfunding. The DAO would allow participants to manage invested funds directly and for governance rules to automatically self-execute, once certain conditions were met.[129]

The DAO set a minimum fundraising goal to be achieved within a defined period, failure to achieve which would have resulted in the funds being returned. During this 'creation phase', units of Ether could be sent to The DAO's smart contract address, in exchange for which The DAO would create and transfer 'DAO tokens'. These tokens conferred voting rights on their holders, in proportion to the number of tokens held.

119. V. Lehdonvirta, Otto Kässi, Isis Hjorth, Helena Barnard & Mark Graham, 'The Global Platform Economy: A New Offshoring Institution Enabling Emerging-Economy Microproviders', 45 *Journal of Management* 567 (2019).
120. T. Webb and G. Cheney, 'Worker-Owned-and-Governed Enterprises and the Wider Co-Operative Movement', in M. Parker *et al*. (eds.), *Routledge Companion to Alternative Organization* (2014) 64 at 76-77; Ben-Ner, above n. 65, at 297.
121. Cf. A. Errasti, I. Bretos & E. Etxezarreta, 'What do Mondragon Coopitalist Multinationals Look Like? The Rise and Fall of Fagor Electrodomésticos S. Coop. and its European Subsidiaries', 87 *Annals of Public and Cooperative Economics* 433 (2016).
122. Bretos, Errasti & Marcuello, above n. 22, at 85.
123. Davidson, De Filippi & Potts, above n. 25, at 641.
124. Hence, why projects like Colony cite Coase's seminal article on the Nature of the Firm on the first page of their White Paper. A. Rea, A. Fischer & J. du Rose, 'Colony: Technical White Paper', 27 July 2018, at 1, https://colony.io/whitepaper.pdf.
125. F.H. Easterbrook and D.R. Fischel, 'Limited Liability and the Corporation', 52 *University of Chicago Law Review* 89, at 89 (1985); O.E. Williamson, *The Economic Institutions of Capitalism* (1985).
126. www.ethdocs.org/en/latest/introduction/what-is-ethereum.html#ethereum-virtual-machine.
127. De Filippi and Wright, above n. 24, at 75.
128. Q. DuPont, 'Experiments in Algorithmic Governance: A History and Ethnography of "The DAO", a Failed Decentralized Autonomous Organization', in M. Campbell-Verduyn (ed.), *Bitcoin and Beyond* (2018) 157, at 159.
129. C. Jentzsch, 'Decentralized Autonomous Organization to Automate Governance' (2016), at 3 https://download.slock.it/public/DAO/WhitePaper.pdf.

They would be freely transferable and divisible.[130] As an entity, creating, storing and transferring tokens was the limit of what The DAO could achieve autonomously.[131] For creating and voting on funding proposals, it required human Contractors. The off-chain projects that would result from successful funding proposals would be directly governed by token-holders, in proportion to the tokens they held, and returns would be distributed pro rata. These tokens could also be sold for fiat currencies through exchanges.

The creation of The DAO was met with a great deal of enthusiasm and during its initial creation phase, it raised US$ 150 million worth of Ether.[132] It was intended that The DAO would be an archetype for future decentralised organisations and in a sense, it was successful. The successful crowdfunding of The DAO – and the subsequent siphoning of over US$ 50 million of Ether and investigation by the US Securities and Exchange Commission (SEC) – has served as a cautionary tale for everyone involved in the blockchain ecosystem. While its name is a misnomer, as key decision-making powers resided in certain humans, it continues to be the prime example of a decentralised organisation. The ambition of creating DOs and DAOs persists[133] but tempered with the knowledge that they are exposed to governance risks endogenous to decentralised systems operating under the logic of smart contracts and are subject to an array of off-chain risk and regulation.

4 Case Study of Colony

4.1 What Colony Does

Colony is a platform that provides the infrastructure for creating an ecosystem of self-organising companies (*i.e.* 'colonies'),[134] by lowering the costs of a diverse group of people coordinating their efforts and resources to realise shared goals, even when they do not necessarily know or trust each other. The ambition of Colony is that this coordination will occur in the organisation created through its platform in a meritocratic manner through the dynamic allocation of *reputation*.

Reputation is a number that is associated with a person, reflecting the value of their recent contributions to a colony. It may be earned by bootstrapping colonies, successfully completing tasks and constructively resolving disputes.[135] This figure affects the extent of a person's control rights in the organisation as well as their share of rewards. Significantly, unlike currencies or securities, reputation cannot be transferred and is non-negotiable in crypto-capital markets.[136]

Colony is still at an early stage of development and much of what is described below is based on its white paper, setting out the features the development team expects the layers of Colony to have. The development team have been building the Colony Network and Colony JS, a software library that enables independent developers to develop applications (dApps) that can interact with the underlying smart contracts. These colonies may be established to create software but also for tangible goods, such as jewellery. As one of the founders of Colony, Jack du Rose, began developing the platform as a way of solving problems he encountered while coordinating persons in a global, high-end jewellery supply chain,[137] the illustrative examples in the following subsection draw from the jewellery industry.

4.2 The Governance of Colony

To understand the governance of the Colony platform, it is necessary to consider the Colony Network, the Meta Colony and individual colony layers separately.

The Colony protocol[138] is built on the Colony Network, a collection of smart contracts deployed on the Ethereum blockchain by the Colony development team. These contracts provide the broad parameters in which colonies may be created, such as the fees charged to use the Network, upgrades of its functionality and the reputation mining mechanism.[139] Management of the Colony Network will be gradually ceded to a Meta Colony, the first, parent colony to be created on the Network.[140] When this has occurred, tokens in the Meta Colony (CLNY) will have been distributed and reputation can be earned in the Meta Colony through the completion of tasks, such as making updates to individual colony smart contracts. CLNY and reputation holders get to vote on the fundamental parameters of the Network (control rights) and receive a portion of the fee charged by the Network when individuals are paid.[141] Moreover, CLNY holders act as reputation miners, calculating rep-

130. Securities and Exchange Commission, 'Report of Investigation Pursuant to Section 21(a) of the Securities Exchange Act of 1934: The DAO' [DAO Report], 25 July 2017, at 6.
131. Jentzsch, above n. 129, at 2.
132. DAO Report, above n. 130, at 16.
133. *See* Colony, DigixDAO; MakerDAO and Hutten DDO, among others.
134. The name was inspired by the archetypical ant colony, a complex adaptive system that may be found in nature. *See* G. Rosenblatt, 'Is Colony a Glimpse of the Blockchain-Based Future of Work?', www.the-vital-edge.com/colony-blockchain/.
135. Rea, Fischer & du Rose, above n. 124, at 15. In the Meta Colony, reputation can also be earned through reputation mining.
136. *Ibid.*, at 14.
137. Blockchain Review, How Blockchain Technology is Enabling the Future of Work, https://www.youtube.com/watch?v=o_erLhcDqMU (last visited 1 December 2018).
138. In general, protocols are a set of rules and steps that facilitate effective communication between computers. As with the internet, the Colony protocol is one of several layers of protocols arranged in a stack through which information travels from one computer to another. The Colony protocol is in between the Ethereum decentralised data processing layer and the layer of applications that are deployed using Colony. In short, the Colony protocol provides the rules for the division of labour, decision making and financial management of decentralised organisations.
139. Rea, Fischer & du Rose, above n. 124, at 5. Individual colonies can opt in to the upgrades.
140. *Ibid.*, at 7-8.
141. *Ibid.*, at 7, 46-47. If the fee is paid in CLNY tokens, it is burned. If it paid in white-listed external crypto-currencies such as Ether and DAI, it will be distributed to a reward pot and a working capital pot. If the fee is paid in a native colony token that is illiquid, monthly Dutch auctions will be held in which the native token can be acquired in exchange for CLNY tokens. These CLNY tokens are then burned (destroyed). I thank Jack du Rose for this information.

utation scores off-chain and updating reputation scores on-chain, for which new CLNY tokens and reputation are conferred as rewards.[142] The functionality of CLNY tokens will be set initially by the Colony development team and the Ethereum community but eventually by the Meta Colony.

Individual colonies may be created to achieve a single goal or multiple goals, over a short or long time frame. They are entities with discrete purposes but act within the broad parameters set by the Colony Network. Regardless of the goal, they will substantially share the membership and governance rules described below due to the underlying smart contract code. As these rules are embodied in code, when they are being used they are much harder to skirt than institutional and social rules in a worker cooperative, where they may be under-enforced.[143] When a colony is created, it will generate its own native token that will primarily have financial value.[144] To achieve its goal(s), the work needed can be broken down into tasks and (sub) domains (*e.g.* assembly) in which tasks can be clustered. This is analogous to departments in an organisation. Domains can also be nested within wider domains, with the widest domain being the colony itself. Along with allocating a task to a domain, tasks will be tagged with relevant skills needed for its completion (*e.g.* #casting, #soldering). This may be a specific skill within a broader skill set (*e.g.* #design). Thus, there is an organisational tree and a skills tree, with participants able to earn and lose reputation in both.

To create and define a task, a person with sufficient reputation must deposit ('stake') colony tokens proportionate to the amount of reputation in the domain.[145] Reputation and colony tokens may be initially assigned as control rights and working capital at the time a colony is created to allow certain persons to set up tasks.[146] Otherwise, usually, a task initiator will submit a funding proposal from the pot (wallet) of a parent domain.[147] The proposal will specify the amount of funds needed and can be denominated in the colony's own currency or in Ether. If there is only one funding proposal for a task, there are sufficient funds in the pot and there are no objections, the smart contract will begin to release funds to the pot of the task. This materialises Colony's emphasis on completing work efficiently rather than voting on every decision. Once the funds needed for payment are in place (the bounty), the manager will have to enter into a tentative agreement with a worker who has the necessary skill set and reputation. When joining the Colony platform, workers would have tagged their skill sets and managers can use this to search for one who is most appropriate for a task. After an agreement is reached, a task may be specified to them along with working guidelines, a due date and payment terms (for the worker, evaluator and manager).[148] While the manager may also act in the capacity of evaluator, this role can be delegated to a separate person as well. The evaluator may be unknown to the worker, as they may only be identifiable by their public key.

Following the completion and evaluation of the task, there will be three days to raise objections and disputes regarding the quality of the task performed. When there are no objections, the worker gets paid in the colony's native token or another approved cryptocurrency.[149] If paid in native tokens, the workers' reputation in their domain increases, as well as all the wider domains of which it is part, including the colony itself (*i.e.* the top-level domain). Simultaneously, their reputation for performing the tagged skill increases, as well as any wider, parent skills of which the skill is a part.[150] The sum of their top-level domain and top-level skills reputations determines their influence on decisions that affect the individual colony. To avoid disproportionate gains in reputation following the completion of a task, the bounty initially set should be consistent.[151]

If there is an objection, an objector must be able to defend his/her objection. Its content should not only specify why a task is inadequate and what could be done better, but also suggestions as to the 'reputations' (*i.e.* Colony members with a certain level of reputation) that should vote if a dispute arises and reasoning for why these reputations should vote. This allows objections to be scaled to a larger group of peers, whether at a domain, colony or Meta Colony level. This objection can only be made if an objector has a certain reputation score and stakes some of their own tokens.[152] If no one makes a counter-stake to object to the objection, then the objection will pass and the worker will receive less/no pay. If someone does sufficiently counter-stake within three days, then a dispute will arise. The staking of tokens is needed not only to avoid frivolous objections but also to compensate the persons involved in settling a dispute through voting. The weight of their votes is contingent on a person's reputation in the skill and domain in dispute.[153] Being on the winning or losing side of a dispute has the corresponding effect of enhancing or diminishing reputation scores. The payment and reputational scores allotted to the worker or evaluator depends on the final score received after disputes are resolved. If the work is found to be inadequate, the worker will receive diminished payment and lose repu-

142. *Ibid*, at 7, 19, 22. Calculating reputation scores off-chain saves costs incurred by Ethereum blockchain transactions.
143. Reyes analogises these parameters with choosing a corporate statute. C.L. Reyes, 'If Rockefeller were a Coder', 87 *George Washington Law Review* 1 (forthcoming 2018), at 34.
144. They will have a vote on changing the supply of native tokens in a colony, Rea, Fischer & du Rose, above n. 124, at 12. They will also be entitled to vote on arbitrary transactions, that is, actions that are unforeseen by the colony and the Meta Colony, at 49.
145. *Ibid.*, at 9.
146. *Ibid.*, at 17.
147. *Ibid.*, at 32-33.
148. *Ibid.*, at 9.
149. *Ibid.*, at 10.
150. The manager's token-holding and domain reputation rises or falls in the same manner, but their skill rating is not affected, *ibid.*, at 16.
151. *Ibid.*, at 13. The White Paper indicates that the tokens allocated could represent the hours worked.
152. Rea, Fischer & du Rose, above n. 124, at 39, Annex A.
153. *Ibid.*, at 42.

tation in their domain and their tagged skill, as well as parent and child domains and parent and child skills.

In addition to payment for completed tasks to workers, managers and evaluators, persons in the colony holding native colony tokens and reputation are entitled to rewards from the revenue earned by the colony.[154] This means that a worker in a colony, waiting for the next task to be assigned to them, can continue to earn (for a while) from the revenue they had helped generate.

4.3 Worker Cooperatives: Learning from the Colony Project

A close reading of the governance structure of colony reveals a startling resemblance to LMFs, such as worker cooperatives. Firstly, the economic activities are carried out primarily for the benefits of its participants. Secondly, most, if not all, of the capital of the organisation is held by the participants. This is indicated by the fact that tokens and reputation are issued exclusively to the participants of a new colony,[155] before gaining potential investors, and as such can only be gained through various forms of work: production, evaluation and management. This is akin to the common practice in the start-up technology sector of granting employees stock options,[156] but in this instance it is coupled with the right to have a voice in significant strategic decisions.

Thirdly, as currently designed, colonies have voluntary, open membership by default. Restricted membership is not mentioned in the Colony White Paper. This is characteristic of initiatives in open source communities, where objective peer review is critical and where, instead, there are concerns about keeping participants motivated and committed.[157] However, the key difference with open source communities is that colonies may not be limited to the private provision of public goods,[158] for which values such as the long-term striving for excellence may come into play.[159] Colonies may be used for the production of private goods as well.

Fourthly, Colony has what can be broadly described as dynamic meritocratic governance, where the weight of one's vote is dynamically adjusted according to one's contributions to a task, domain or colony. In itself, this is not contrary to cooperative principles as there are cooperatives which weigh voting power according to, for example, production.[160] Participants still have a voice in the governance and strategic decision-making of the colony, as exemplified by the fact that anyone can set up a task for the colony to complete.

Fifthly, it is clear from the White Paper that the assets of a colony are conceptually distinct from that of the participants, as they are escrowed in a smart contract and associated pots. Access to these pots is conditional on a successful funding proposal. There is also a separate revenue pot from which rewards may be distributed or working capital replenished.[161] Notionally, colony smart contracts can subsist indefinitely with tokens in escrow, even after it has been abandoned, indicating that it is technologically possible for the colony to have its own capital. Moreover, the payment of Network fees, which is reinvested to maintain the Network and to do useful supportive work (*e.g.* build applications) is also reminiscent of the cooperative practice of building financial reserves and investing in useful services (*e.g.* training) to sustain the mission of the business.

While taking these similarities into account, there are certain functionalities in Colony, which can potentially overcome the start-up and coordination costs that worker cooperatives often face, especially when operating across borders.

Decentralised organisations prefigure ready-made governance structures that are easily accessible online and are native to globally distributed blockchains. While the governance mechanism is technically complex, as with other digital applications, once launched its use will be intuitive and user-friendly. As such, these organisations can provide capital and governance structures for digitally native worker cooperatives to adopt.

In terms of financing, worker cooperatives can consider implementing a system in which financial rewards and decision-making power are generated through useful patronage, represented as separate quantified units, but with only the financial rewards being exchangeable – as they are with native tokens and reputation on the Colony platform.[162] If the token gains use-value, then it can be sold or swapped for other, more widely used cryptocurrencies, which can tide over those who only have intermittent work. The relative transferability of a token compared to a partnership interest, a standard cooperative membership, or an employee share held in a trust, allows workers to diversify their risks, in the event their cooperative fails. At the same time, this allows for a certain amount of external investment to flow into the business. As (most) decision-making rights are not attached to native tokens independent of reputation, it may be acquired and held by third parties without diluting the

154. *Ibid.*, at 44-45.
155. *Ibid.*, at 13.
156. Cf. Index Ventures, *Rewarding Talent: A Guide to Stock Options for European Entrepreneurs* (2017), at 13.
157. *See, e.g.* G. von Krogh *et al.*, 'Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development', 36 *MIS Quarterly* 649 (2012).
158. As open source software is often characterised as, *see* M.A. Rossi, 'Decoding the Free/Open Source Software Puzzle: A Survey of Theoretical and Empirical Contributions', in J. Bitzer and P.J.H. Schröder (eds.), *The Economics of Open Source Software Development* (2006) 15, at 33.
159. *See* the discussion on the social philosophy of Alasdair Macintyre in van Krogh *et al.*, above n. 157, at 661ff.
160. Section 2.4(8)(a), PECOL.
161. Rea, Fischer & du Rose, above n. 124, at 44.
162. Financial reward here refers to both a cryptocurrency for work done and a token from the revenue of the colony. Reputation, like labour, is inalienable from the worker-member. The development of online reputation systems allows skills, organisational contributions and organisational value to be represented more tangibly, homogenously and dynamically than capital shares and labour membership. On the limitations of a LMF membership market due to the inalienability of labour, *see* G.K. Dow, *The Labor-Managed Firm: Theoretical Foundations* (2018), at 8.

decision-making rights of worker-members, as is the predominant concern with non-member investment.[163]

In terms of collective action problems, a frequent criticism of worker cooperatives is time spent on meetings to reconcile heterogeneous interests,[164] and as such taking actions on the basis of tacit consent, rather than majority voting or unanimity, may in fact be preferable. Similarly, the requiring of staking of reputation and tokens in raising an objection can help avoid trivial disagreements about the quality of work. Turning to the aforementioned cross-border coordination issues, the fact that workers are drawn from different backgrounds prevents them from having a shared background in terms of politics, work and culture, which are usually associated with worker cooperatives.[165] Instead, reputation-weighted governance may be especially suited for organisations seeking to coordinate a heterogeneous, pseudonymous group of actors[166] who operate across a wide geographical territory with limited trust and state policing. While blockchain communities have only emerged in recent years,[167] history is replete with examples of such organisations. Examples range from the Amsterdam Stock Exchange in the seventeenth century[168] to modern Moroccan bazaars.[169] Contemporaneous examples include Usenet newsgroups, massive multiplayer online gaming and open source software developer communities. A common theme appears to be finding counterparties with desirable qualities (*e.g.* a certain set of skills and experience), while at the same time coordinating these individuals to ensure contractual performance and the pursuance of the collective interest.

This does not necessarily require external enforcement, through judges or regulators, but can be achieved through the threat of diminished reputation. The risk of losing reputation is sufficient motivation for performance by a party, especially when it is in their interest to have continuous transactions with a counterparty,[170] on a regular[171] or irregular basis.[172] As such, the fear of lost reputation will 'crowd in' honesty in the long run.[173] This is true of online communities and project-based work, particularly in creative industries.[174] This, however, assumes that parties have sufficient information and knowledge of each other's reputations. Online reputation systems are able to address these information asymmetries to an extent, as user reviews and ratings provide granular information about a potential counterparty in a digestible form. Yet, peer-to-peer systems are vulnerable to manipulation by platforms that host them and biased reviewers, raising concerns about the system's own trustworthiness.[175]

However, the manner of its deployment in the Colony protocol makes the system less prone to cronyism. Managers of tasks are incentivised to intuitively and objectively choose workers based on a quantification of their demonstrated skills and recent contributions, rather than personal characteristics, as they stake their own tokens when initiating a task. This score is not generated through ratings by (potentially) anonymous individuals with little to lose. Instead, evaluators stand to receive diminished payment and a reduced reputation score for inadequate evaluations, while contesting a task or decision through the dispute resolution mechanism requires risking tokens and reputation. A teething concern about the democratisation of reputation systems is that it will ultimately not be sustained, with its growing complexity leading to the emergence of oligarchy. One empirical study has already observed this trend with regard to peer-production projects, leading to structural changes in authority and a reorientation of organisational goals.[176] A key distinguishing feature of Colony's reputation system, however, is its degradability, which prevents early movers from resting on their laurels and incentivises the continuous, useful engagement of all members in the governance of colonies. To embed such a system in a worker cooperative, a link to a user-friendly portal that provides up-to-date individual reputation scores and accrued financial rewards may be provided in the section of the by-laws concerning membership.

5 Conclusion

Colony is one of a handful of blockchain projects currently exploring how to design organisations that work

163. It is also less clear-cut that a token, as described herein, will constitute a security as compared to tradable shares in a worker cooperative, which generally will. *See* K. Mikami, 'Are Cooperative Firms a Less Competitive Form of Business? Production Efficiency and Financial Viability of Cooperative Firms with Tradable Membership Shares', 42 *Economic Systems* 487, at 501 (2018); S. Zamagni and V. Zamagni, *Cooperative Enterprise: Facing the Challenge of Globalization* (2010), at 87-88.
164. Cf. G.F. Davis, 'Can an Economy Survive Without Corporations? Technology and Robust Organizational Alternatives', 30 *Academy of Management Perspectives* 129, at 137 (2016).
165. Z.F. Gamson and H.M. Levin, 'Obstacles to the Survival of Democratic Workplaces', in R. Jackall and H. Levin, *Worker Cooperatives in America* (1984) 220, at 225.
166. I. Bohnet, B.S. Frey & S. Huck, 'More Order with Less Law: On Contract Enforcement, Trust, and Crowding', 95 *American Political Science Review* 131 (2001).
167. DuPont, above n. 128, at 175.
168. E. Stringham, 'The Extralegal Development of Securities Trading in Seventeenth-Century Amsterdam', 43 *The Quarterly Review of Economics and Finance* 321, at 324 (2003).
169. C. Geertz, 'The Bazaar Economy: Information and Search in Peasant Marketing', 68 *The American Economic Review* 28, at 29 (1978).
170. Stringham, above n. 168, at 323-4, 336.
171. R.C. Ellickson, *Order Without Law* (1991), at 55-58, 214.
172. P.R. Milgrom, D.C. North & B.R. Weingast, 'The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs', 2 *Economics and Politics* 1, at 7-8 (1990); D.W. Brown, *When Strangers Cooperate* (1995), at 18.
173. Bohnet, Frey & Huck, above n. 166, at 132, 138.
174. J. Lerner and J. Tirole, 'Some Simple Economics of Open Source', 50 *The Journal of Industrial Economics* 197, at 218 (2002); P. Schörpf *et al.*, 'Triangular Love-Hate: Management and Control in Creative Crowdworking', 32 *New Technology, Work and Employment* 43, at 46 (2017).
175. S. Ranchordás, 'Online Reputation and the Regulation of Information Asymmetries in the Platform Economy', 5 *Critical Analysis of Law* 127, at 134-8 (2018).
176. A. Shaw and B.M. Hill, 'Laboratories of Oligarchy? How the Iron Law Extends to Peer Production', 64 *Journal of Communication* 215, at 219, 229 (2014).

in the interest of its multi-stakeholder organisations.[177] These decentralised organisations reconfigure ownership within firms, enabling greater rights to the residual profits of the firm and control rights. In doing so, they bear a remarkable resemblance in the crypto-space to the early pioneers of worker cooperativism.

Undoubtedly, such projects (including Colony) entail risks and proactive cooperators should be wary of them when experimenting with blockchain technology. The regulatory status of crypto-tokens is still in flux[178] and sudden classification as a security can have deeply unpleasant, costly securities liability consequences for members.[179] This article has concentrated on the capital and governance structures of cooperatives, but it is still unsettled which legal structure would be most suitable for the goals of DOs while still providing the benefits of limited liability.[180] It is therefore important to be open to the idea of also using technologies other than blockchain in creating the governance and capital structure recommended in this article. Moreover, for the promoters of such businesses, as well as interested participants, it is necessary to challenge and grapple with the complexity of these governance structures in which corporate governance-by-design is sought, as it potentially embeds power structures in new and unexpected ways. Decades of research into cooperative degeneration and regeneration highlight the importance of being alive to the possibility of oligarchy emerging.

On a more optimistic note, blockchain projects such as Colony provide considerable insight into the technological and theoretical possibilities (and limitations) of decentralised governance. The proposed capital and governance structure of colonies may hold lessons for LMFs, such as worker cooperatives, in the process of being formed and those confronted with cross-border coordination problems as they expand overseas. These decentralised governance structures allow us to imagine self-employed persons or small businesses in Bangladesh, Uzbekistan and the Netherlands collaborating together in a joint venture, where power is not distributed according to capital or bargaining power, but reputation tied to the quality of their non-capital contributions. As blockchain technology is adopted more widely, this may be a part of a broader movement to achieve a more engaged, more effective participatory democracy across nation states.[181] By providing the contours of how worker cooperatives may draw lessons from these blockchain projects, this article has sought to contribute to the realisation of alternative economies[182] in which there is greater scope for worker ownership.

177. DAOStack, https://daostack.io/; Aragon, https://aragon.org; Steem, https://steem.com, among others (last visited on 8 December 2018).
178. W. Hinman, 'Digital Asset Transactions: When Howey Met Gary (Plastic)', 14 June 2018, https://www.sec.gov/news/speech/speech-hinman-061418.
179. For a case involving securities classification of a purported utility token, see *In Re: Munchee*, Administrative Proceeding File No. 3-18304, 11 December 2017, at 5-6.
180. Reyes, above n. 143, at 43 suggests the business trust. Indeed, some blockchain projects have incorporated as a cooperative, *see* Lars, 'ARK Creates a Unique Business Entity', *Medium*, 27 November 2017.
181. M.-L. Marsal-Llacuna, 'Future Living Framework: Is Blockchain the Next Enabling Network?', 128 *Technological Forecasting & Social Change* 226, at 232 (2018).
182. J.K. Gibson-Graham and G. Roelvink, 'The Nitty Gritty of Creating Alternative Economies', 30 *Social Alternatives* 29 (2011).

Abbreviations and Glossary

Blockchain Technology	A resilient, near-immutable, distributed and transparent database that can pseudonymously execute economic transactions. It can be public or private, thereby affecting who can interact with the blockchain. (*See* De Filippi and Wright, at n. 24, at 2)
CLNY	Meta Colony of the Colony protocol. The Meta Colony also has its own tokens, referred to as CLNY tokens.
Cryptocurrency/Currency Token	Tokens that are a unit of account and are used as a means of payment
D(A)O	Decentralised (Autonomous) Organisations that use blockchain technology and smart contracts as their primary or exclusive source of governance and respond to both digital and human inputs. (*See* De Filippi and Wright, at n. 24, at 136-7)
Investment Token	Tokens that have the characteristics of an equity instrument and embody expectations of future profit through the managerial efforts of others
KMF	Capital-Managed Firm
LMF	Labour-Managed Firm
PECOL	Principles of European Cooperative Law
Off-chain	All transactions that are not represented on the blockchain
Oracle	A third party, trusted by parties of a smart contract, that relays information from the outside world to a smart contract
SCE	European Cooperative Society
SEC	Securities and Exchange Commission of the United States of America
Smart Contract	Software that embodies an agreement between parties and then (self-)executes when certain conditions are met
Utility Token	Tokens that give a right of access to an online platform, product or service

doi: 10.5553/ELR.000113 - ELR december 2018 | No. 3