Regulating Algorithms' Regulation? First Ethico-Legal Principles, Problems, and Opportunities of Algorithms

Giovanni Comandè

Abstract Algorithms are regularly used for mining data, offering unexplored patterns, and deep non-causal analyses in what we term the "classifying society". In the classifying society individuals are no longer targetable as individuals but are instead selectively addressed for the way in which some clusters of data that they (one or more of their devices) share with a given model fit in to the analytical model itself. This way the classifying society might bypass data protection as we know it. Thus, we argue for a change of paradigm: to consider and regulate anonymities—not only identities—in data protection. This requires a combined regulatory approach that blends together (1) the reinterpretation of existing legal rules in light of the central role of privacy in the classifying society; (2) the promotion of disruptive technologies for disruptive new business models enabling more market control by data subjects over their own data; and, eventually, (3) new rules aiming, among other things, to provide to data generated by individuals some form of property protection similar to that enjoyed by the generation of data and models by businesses (e.g. trade secrets). The blend would be completed by (4) the timely insertion of ethical principles in the very generation of the algorithms sustaining the classifying society.

Abbreviations

AI	Artificial intelligence
CAL. BUS & PROF. CODE	California business and professions code
CAL. CIV. CODE	California civil code
CONN. GEN. STAT. ANN.	Connecticut general statutes annotated
DAS	Domain awareness system
DHS	U.S. Department of Homeland Security
DNA	Deoxyribonucleic acid
EDPS	European data protection supervisor
EFF	Electronic Frontier Foundation

G. Comandè (✉)
Scuola Superiore Sant'Anna Pisa, Piazza Martiri della Libertà 33, 56127, Pisa, Italy
e-mail: g.comande@santannapisa.it

© Springer International Publishing AG 2017

T. Cerquitelli et al. (eds.), *Transparent Data Mining for Big and Small Data*,
Studies in Big Data 32, DOI 10.1007/978-3-319-54024-5_8

EU	European Union
EU GDPR	European Union general data protection regulation
EUCJ	European Union Court of Justice
FTC	Federal Trade Commission
GA. CODE ANN.	Code of Georgia annotated
GPS	Global positioning system
GSM	Global system for mobile communications
GSMA	GSM Association
ICT	Information and communications technology
NSA	National Security Agency
PETs	Privacy-enhancing technologies
PPTCs	Privacy policy terms and conditions
SDNY	United States District Court for the Southern District of New York
ToS	Terms of service
WEF	World Economic Forum
WPF	World Privacy Forum

> "Today's trends have opened an entirely new chapter, and there is a need to explore whether the principles are robust enough for the digital age. The notion of personal data itself is likely to change radically as technology increasingly allows individuals to be re-identified from supposedly anonymous data. In addition, machine learning and the merging of human and artificial intelligence will undermine concepts of the individual's rights and responsibility" [1, p. 13].

1 Introduction

Today's technologies enable unprecedented exploitation of information, be it small or big data, for any thinkable purpose, but mostly in business [2, 3] and surveillance [4] with the ensuing juridical and ethical anxieties.

Algorithms are regularly used for mining data, offering unexplored patterns and deep non-causal analyses to those businesses able to exploit these advances. They are advocated as advanced tools for regulation and legal problem-solving with innovative evidence gathering and analytical capacities.

Yet, these innovations need to be properly framed in the existing legal background, fit in to the existing set of constitutional guarantees of fundamental rights and freedoms, and coherently related to existing policies in order to enable our societies to reap the richness of big and open data while equally empowering all players.

Indeed, if the past is normally the prologue of the future, when our everyday life is possibly influenced by filtering bubbles and un-verifiable analytics (in a word: it is heteronomously "pre-set"), a clear ethical and legal response is desperately needed to govern the phenomenon. Without it, our societies will waver between two extremes—either ignoring the problem or fearing it unduly. Without a clear transnational ethical and legal framework, we risk either losing the immense possibilities entailed by big and small data analytics or surrendering the very gist of our praised rights and freedoms.

To secure the benefits of these innovations and avoid the natural perils every innovation brings along, this paper makes a call for regulating algorithms and expanding their use in legal problem-solving at various levels by first exploring existing legal rules. Accordingly, this paper, building on existing literature: (1) frames the main legal and ethical issues raised by the increasing use of algorithms in information society, in digital economy, and in heightened public and private surveillance; (2) sets the basic principles that must govern them globally in the mentioned domains; (3) calls for exploring additional uses of algorithms' evidence-based legal problem-solving facility in order to expand their deployment on behalf of the public interest (e.g. by serving pharmacovigilance) and to empower individuals in governing their data production and their data flow.

It is important to stress that the tremendous benefits of big data—for instance, of the interplay among data mining and machine learning—are not questioned here. To the contrary, this paper asserts that in order to reap all the promised richness of these benefits it is important to correctly frame the reality of the interplay of these technologies. It is also important to emphasize that data protection, whatever form it takes in a given legal system, can be a key to development, rather than an obstacle.[1] There is no possible alternative to sustaining such a bold statement, since in the digital age static profiles are basically disappearing due to the expansion of machine learning and artificial intelligence. Among other things, such an evolution implies that the classification model used in any given moment no longer exists, as such, in a (relatively contemporaneous) subsequent one. Accountability, then, requires both the technical (already extant) and the legal ability to establish with certainty in each moment which profile has been used in the decision process.

Thus, the emerging regulatory approach must necessarily blend together various legal, technological, and economic strategies for which the time frame is of crucial importance.

The EDPS [1, p. 13] rightly stressed that "The EU in particular now has a 'critical window' before mass adoption of these technologies to build the values into digital structures which will define our society. This requires a new assessment of whether the potential benefits of the new technologies really depend on the collection and analysis of the personally-identifiable information of billions of individuals. Such an

[1] "Contrary to some claims, privacy and data protection are a platform for a sustainable and dynamic digital environment, not an obstacle" [1, p. 9].

assessment could challenge developers to design products which depersonalise in real time huge volumes of unorganized information making it harder or impossible to single out an individual." Yet, at this stage we should wonder whether it is more a matter of privacy protection or dignity threatened by the *biased targeting of anonymities*, not just individuals.

2 The Classifying Society

The Free Dictionary defines the verb "to classify" as: "to arrange or organize according to class or category." Google, faithful to its nature, clarifies: "[to] arrange (a group of people or things) in classes or categories according to shared qualities or characteristics".

According to the Cambridge Advanced Learner's Dictionary, the verb "classify" originally referred, however, to "things", not to humans.[2] The societies in which we live have altered the original meaning, extending and mostly referring it to humans. Such a phenomenon is apparent well beyond big data[3] and the use of analytics. For instance, in medicine and research the term "classification" now applies to old and new (e.g. emerging from DNA sequencing) shared qualities and characteristics that are compiled with the aim of enabling increasingly personalized advertising, medicine, drugs, and treatments.[4]

The rising possibilities offered by machine learning, data storage, and computing ability are entirely changing both our real and virtual landscapes. Living and being treated according to one or more class is the gist of our daily life as personalized advertising[5] and the attempt by companies to anticipate our desires and needs[6] clearly illustrate.

Often every group of "shared qualities or characteristics" even takes priority over our actual personal identity. For instance, due to a group of collected characteristics,

[2]Cambridge Advanced Learner's Dictionary & Thesaurus of the Cambridge University Press specifies as the first meaning: "to divide things into groups according to their type: The books in the library are classified by/according to subject. Biologists classify animals and plants into different groups".

[3]On the risks and benefits of big data see e.g., Tene and Polonetsky [5]; *contra* Ohm [6].

[4]"Personalization is using more (demographic, but also behavioral) information about a particular individual to tailor predictions to that individual. Examples are Google's search results based on individual's cookies or GMail contents" [7, p. 261]. See also https://www.google.com/experimental/gmailfieldtrial.

[5]"Algorithms nowadays define how we are seen, by providing a digital lens, tailored by statistics and other biases." [7, p. 256].

[6]Amazon for instance is aiming at shipping goods to us even before we place an order [8]. This approach is very similar to Google attempting to understand what we want before we know we want it. "Google is a system of almost universal surveillance, yet it operates so quietly that at times it's hard to discern" [9, p. 84].

we can be classified and targeted accordingly as (potential) terrorists, poor lenders,[7] breast cancer patients, candidates for a specific drug/product, or potentially pregnant women.[8] These classifications can be produced and used without us having the faintest clue of their existence, even though we are not actual terrorists (perhaps, for instance, we like to travel to particular areas and happen to have an Arab-sounding name), we are affluent (but prefer to live in a much poorer neighbourhood), we do not have breast cancer, we have no need for or interest whatsoever in a drug/product, we are not even female, or we are not a sex addict (see infra).

The examples referred to above are well documented in literature and have been chosen to illustrate that classifications characterize every corner of our daily life.[9] They expose most of the legal problems reported by scholars and concerned institutions. Most of these problems revolve around notions of privacy, surveillance, and danger to freedom of speech,[10] ...

Yet, literature to date has failed to discuss the very fact that the classifying society we live in threatens to make our actual identities irrelevant, fulfilling an old prophecy in a cartoon from a leading magazine. This cartoon displayed a dog facing a computer while declaring: "on internet nobody cares you are actually a dog". With hindsight, we could now add "if you are classified as a dog it is irrelevant you are the human owner of a specific kind of dog". Indeed, Mac users are advertised higher prices regardless of whether they are personally identified as affluent [18]. Although in some countries (such as the USA) price discrimination and customer-steering are not forbidden unless they involve prohibited forms of discrimination,[11] we should begin to question the ethics of such processes once they are fully automatic and

[7] See for more examples Citron and Pasquale [10].

[8] By using previous direct interaction, Target knew a teenage girl was pregnant well before her family did [11].

[9] See, for instance, the following list of horrors in Gray and Citron [12, p. 81, footnotes omitted]: "Employers have refused to interview or hire individuals based on incorrect or misleading personal information obtained through surveillance technologies. Governmental data-mining systems have flagged innocent individuals as persons of interest, leading to their erroneous classifications as terrorists or security threats. ... In one case, Maryland state police exploited their access to fusion centers in order to conduct surveillance of human rights groups, peace activists, and death penalty opponents over a 19 month period. Fifty-three political activists eventually were classified as 'terrorists,' including two Catholic nuns and a Democratic candidate for local office. The fusion center subsequently shared these erroneous terrorist classifications with federal drug enforcement, law enforcement databases, and the National Security Administration, all without affording the innocent targets any opportunity to know, much less correct, the record."

[10] On the chilling effect of dataveillance for autonomy and freedom of speech see, for instance, in literature [13–17].

[11] The limits of antidiscrimination law to cope with data-driven discrimination have been already highlighted by Barocas and Selbst [19].

unknown to the target.[12] In addition, they lock individuals into myriad anonymous models based upon traits that they might not be able to change, even if, theoretically, such traits are not as fixed as a fingerprint—as, for instance, where the use of an Apple product is needed for the specific kind of work performed, or happens to be required by ones' employer.

Even in those instances in which the disparate classifying impact is technically legal it can be ethically questionable, to say the least. Such questionability does not depend so much on the alleged aura of infallibility attributed to automatic computerized decision-making [23, p. 675] as on its pervasiveness and unavoidability.

Apparently there is nothing new under the sun. Humans, for good or bad, have always been classified. Classes have been a mode of government as well. What is different about the classificatory price discrimination (and not only price!)[13] that can be systematically produced today—different from the old merchant assessing the worth of my dresses adjusted according to the price requested—is discrimination's dimension, pace, pervasiveness and relative economy.[14]

Literature has also failed to address another key issue: due to the role of big data and the use of algorithms, actual personal identification as a requirement for targeting individuals is becoming irrelevant in the classifying society. The emerging targets are clusters of data (the matching combination between a group of data and a given model), not physical individuals. The models are designed to identify clusters of characteristics, not the actual individuals who possess them—and even when the model incorrectly identifies a combination, it is refining its own code,[15] perfecting its ability to match clusters of shared characteristics in an infinite loop.

Classification based on "our" "shared characteristics" covers every corner of us (what we are and what we do, let alone how we do it) once we can be targeted at no cost as being left-handed, for instance. Yet, the power (resources, technology, and network economies) to do such classifying is quickly becoming concentrated in fewer hands than ever.[16] The pace of classification is so rapid as to have begun happening in real time and at virtually no cost for some players. Its pervasiveness is evidenced by the impact our web searches have on our personalized

[12] Some forms of notification at least have already been advocated in the context of the debate surrounding the USA's Fourth Amendment [20]. In the EU, specific rules on automation are in place [21]. However, some authors claim that automation as such does not require higher scrutiny [22].

[13] See also Moss [24], stressing the ability of algorithms to discriminate "in practically and legally analogous ways to a real world real estate agent".

[14] "It is not just the amount of data but also novel ways to analyze this data that change the playing field of any single individual in the information battle against big companies and governments. Data is becoming a key element for profit and control, and computers gain in authority" [7, p. 256; 25].

[15] See infra footnotes 73–85 and accompanying text.

[16] According to EU Competition Commissioner Margrethe Vestager, the EU Commission is considering the proposal of a specific directive on big data.

advertising on the next web site we visit.[17] The progressive switch to digital content and services makes this process even faster and easier: "Evolving integration technologies and processing power have provided organizations the ability to create more sophisticated and in-depth individual profiles based on one's online and offline behaviours" [27, p. 2].[18] Moreover, we are getting increasingly used to the myth of receiving personalized services for free[19] and to alternatives to the surrender of our data not being offered [29].

Classifications are not problematic by definition, but some of their modes of production or uses might be. In particular, this applies extensively to the digital domain, wherein transactions are a continuum linking the parties (businesses and "their" customers)—a reality that is clearly illustrated by the continuous unilateral changes made to Terms of Service (ToS) and Privacy Policy Terms and Conditions (PPTCs) that deeply alter the content of transactions[20] after (apparent) consumption and even contemplate the withdrawal of the product/service without further notice.[21]

On a different account, the expanding possibility of unlocking information by data analysis can have a chilling effect on daily choices when the virtual world meets the real one at a very local level. For instance, a clear representation of the outer extreme of the spectrum introduced at the beginning (fear of technology), could be the hesitation to join discussion groups (on drugs, alcoholism, mental illnesses, sex, and other topics) for fear that such an affiliation can be used in unforeseen ways and somehow disclosed locally (maybe just as a side effect of personalized marketing).[22]

Even when cookies are disabled and the web navigation is "anonymous", traces related to the fingerprints of the devices we use are left. Thus, the elaboration and enrichment of those traces (that are by definition "anonymous") could be related to one or more identifiers of devices, rather than to humans. At this point there is no need to target the owner of the device. It is simpler and less cumbersome from a legal standpoint to target the various evolving clusters of data related to a device or a group of devices instead of the personally identified/able individual. Such a state of affairs calls for further research on the need to "protect" the anonymous (and, to the extent that we are unaware of its existence, imperceptible) identity generated by data analytics.

[17] See in general [26].

[18] See also Rajagopal [28].

[19] This is the way in which data collection and sharing is supposedly justified in the eyes of customers.

[20] For a general description, see Perzanowski [30].

[21] Apple [31] for instance imposes the acceptance of the following: "Notwithstanding any other provision of this Agreement, Apple and its licensors reserve the right to change, suspend, remove, or disable access to any products, content, or other materials comprising a part of the Service at any time without notice. In no event will Apple be liable for making these changes. Apple may also impose limits on the use of or access to certain features or portions of the Service, in any case and without notice or liability."

[22] Actually, companies already extensively use algorithms to select employees. For documented cases, see Behm [32].

In this unsettling situation the legal and ethical issues to be tackled could be tentatively subsumed under the related forms of the term "classify".

The adjective "classifiable" and its opposite "non-classifiable" indicate the legal and ethical limits of classification that have so far been made publicly manifest in well-known incidents like the NSA scandal,[23] Facebook tracking,[24] the Domain Awareness System (DAS)[25] and TrapWire software.[26] Yet, the virtual number of classifications and biases is actually infinite [38].

The verb "misclassify" and its related adjectives ("misclassified," "misclassifying") denote instances of error in classification. As paradoxical as it might sound, these algorithms' "mistakes" (false positives and false negatives) reinforce the strength of the classifying society, while simultaneously opening a Pandora's box of new and often unknown/undiscoverable patterns of discrimination.

The verbs "overclassify" and "pre-classify" entail, for instance, risks of excessive and anticipatory classification capable of limiting autonomy, and can certainly be attached to the use of predictive coding in any capacity, be it automatic or human controlled/readable.

Indeed, since literature has clearly illustrated that in the classifying society full personal identification[27] is not necessary to manipulate the environment in which we are operating [29], it is paramount to focus on the tracked/traceable[28] algorithmic "anonymous" identities upon which businesses and governments[29] rely to deal with us—that is, to focus on the classes, those various sets of data that pigeonhole us (these sets of data, are, to use the more appealing technological term, the "models" upon which algorithms act). After all, if a model is already built on data available to our counterpart and only very few of "our" data, even a pseudo-anonymized or fully anonymized model [42] is sufficient to pigeonhole; the classifying society is, across the continents, altogether bypassing data protection as we know it because it is targeting subsets of data fitting specific models rather than individuals expressly. Moreover, as anticipated, these clusters are mostly related to things rather than to individuals. Hence, and for instance, no warning (notice) is needed if the model does not need (therefore it is not even interested) to know it is targeting a given individual in real life; it only needs to identify a cluster of data fitting the model, related to one or more things, regardless of their correspondence with an identified or identifiable

[23] The NSA spying story is nothing new [for the timeline, 33, 34].

[24] FaceBook tracks micro-actions such as mouse movements as well [35].

[25] See Privacy SOS [36].

[26] "TrapWire is a unique, predictive software system designed to detect patterns indicative of terrorist attacks or criminal operations. Utilizing a proprietary, rules-based engine, TrapWire detects, analyzes and alerts on suspicious events as they are collected over periods of time and across multiple locations" [37].

[27] See, on the risk of re-identification of anonymized data, Ohm [39].

[28] "We are constantly tracked by companies and governments; think of smart energy meters, biometric information on passports, number plate tracking, medical record sharing, etc." [7, p. 256].

[29] Often acting synergistically: see Hoofnagle [40]; Singer [41].

individual. Yet, if behind the anonymous subset of data there is an undiscovered real person, that person is effectively being targeted—even if the law as currently formulated does not regard it as such.[30]

These technologies are deployed in the name of a "better user experience," and of fulfilling the future promise of augmented reality and enhanced human ability. Yet, living immersed in the classifying society, we wonder whether the reality that "better matches" our (even anonymous) profiles (shared characteristics in a model) is also a distorted one that reduces our choices.

Such a troubling doubt calls into question the role that law has to play in facing the power of algorithms. It requires exploring both the adapted applications of existing rules and the design of new ones. It also suggests that ethical and legal principles should be shared and embedded in the development of technology [44]. The power to classify in a world that is ruled by classification should entail duties along with rights.

Nevertheless, even the newest regulation on data protection, the EU GDPR,[31] does not address these concerns. To the contrary, it might amplify them, legitimizing the entire classifying society model. But we cannot deal with such issues here.[32]

Finally, the emergence of the classifying society is sustained by an economic argument that is as widespread as it is false: massive data collection rewards companies that provide "free" services and access where an alternate business model (for payment) would fail. First of all, this argument itself makes clear that since the service is rewarded with data collection, it is not actually free at all.[33] Secondly, markets based entirely on a pay-for-service model do exist: the Chinese internet services system is a clear example.

Against this general framework we need to now re-frame a number of the legal issues—already stressed by the literature—generated by the widespread use of algorithms in the classifying society.

2.1 (Re)sorting Out the Legal Issues

A large portion of the legal and ethical issues arising from the use of algorithms to read big (or small) data have been already identified by both legal and information

[30] See Article 29 Data Protection Working Party [43].

[31] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, or "GDPR").

[32] See art. 6 of the GDPR on subsequent processing and pseudo-anonymous data.

[33] The argument that "free" services actually command a price (in data) for their services and the suggestion that "free users should be treated as consumers for the purposes of consumer protection law" has been already advanced [45, pp. 661–662]; on the economic value of data see Aziz and Telang [46].

technology experts. Similarly, there are a variety of taxonomies for algorithms related to their ability to look for hypotheses and explanations, to infer novel knowledge (deductive algorithms) or transform data into models (inductive algorithms), and to use third-party data to colour information (so-called socialized searches such as the ones used by Amazon's book recommendations).[34] All of these taxonomies can be related to the "classify" vocabulary mentioned above, but this is not the task we have here. The following brief overview will explicate the main issues, demonstrating how the two identified deficiencies in the literature call for a different approach and a multilayer regulatory strategy.

A growing literature[35] has already illustrated the need for more transparency in the design and use of data mining, despite the fact that transparency as such can undermine the very goal of predictive practices in a manner that is disruptive to the public interest (for instance by making public the random inspection criteria used to select taxpayers).[36] Nevertheless, what we are questioning here is not the use of classification and algorithms for public goals and by public authorities. Instead, we focus on the pervasiveness of these processes in daily life and at the horizontal level among private entities with strong asymmetries of knowledge and power. Such asymmetries in the use of descriptive and predictive algorithms can generate micro-stigmas which have not been fully explored, let alone uncovered. These micro-stigmas or classifications are so dispersed and undisclosed that mobilization and reaction appear unrealistic. Indeed, and for instance, it is a new stereotype that Apple products users are more affluent than PC users; yet it is a stereotype that escapes existing legal and ethical rules and can lead to higher prices without even triggering data protection rules.

In their reply to Professor Richards, Professors Citron and Gray [54, p. 262] recall the various forms of surveillance leading to "total-information awareness": "coveillance, sousveillance, bureaucratic surveillance, surveillance-industrial complex, panvasive searches, or business intelligence". In stressing the role of fusion centers[37] as a key to this shift to total surveillance,[38] they emphasize the fall of the public/private surveillance divide.

[34] These algorithms use the model built on other people's similar behavioral patterns to make suggestions for us if they think we fit the model (i.e. the classification) they have produced [47].

[35] As beautifully described by Pasquale and Citron [48, p. 1421]: "Unexplained and unchallengeable, Big Data becomes a star chamber… secrecy is a discriminator's best friend: unknown unfairness can never be challenged, let alone corrected". On the importance of transparency and accountability in algorithms of powerful internet intermediaries see also Pasquale [49, 50]. But see, on the role of transparency and the various levels of anonymity, Zarsky [51, 52]; Cohen [53].

[36] The point is clearly illustrated by Zarsky [52].

[37] In their description [54, pp. 264–265, footnotes omitted]: "Fusion centers access specially designed data-broker data-bases containing dossiers on hundreds of millions of individuals, including their Social Security numbers, property records, car rentals, credit reports, postal and shipping records, utility bills, gaming, insurance claims, social network activity, and drug- and food-store records. Some gather biometric data and utilize facial-recognition software."

[38] See the official description [55].

Yet, this is not the Orwellian nightmare of 1984 but a distributed mode of living and "manipulating" in which we are normally targeted indirectly—that is, by matching a subset of our data (that do not need to be personal identifying information in the eyes of law) with the models generated by algorithms during the analysis of large quantities of data (data not necessarily related to us either), all without the need to identify us as individuals.

When we read the actual landscape of the information society in this way, continuous surveillance, as it is strikingly put by Julie E. Cohen [53], alters the way people experience public life.[39] However, the implications for human daily life and activities are deeper and more subversive—and yet mostly unnoticed—in the relationships we develop under private law [29].[40]

Predictive algorithms are applied widely [58][41] and extensively resort to data mining [62, pp. 60–61]. They have intensively and deeply changed the notion of surveillance and have introduced a novel world of covered interaction between individuals and those who sense them (private or public entities or their joint alliance). As Balkin [63, p. 12] puts it, "Government's most important technique of control is no longer watching or threatening to watch. It is analyzing and drawing connections between data.... [D]ata mining technologies allow the state and business enterprises to record perfectly innocent behavior that no one is particularly ashamed of and draw surprisingly powerful inferences about people's behavior, beliefs, and attitudes."[42]

A significant American literature has tackled the problem by examining potential threats to the Fourth Amendment to that country's Constitution [64–67]. Yet, from a global perspective wherein the American constitution does not play a role, it is

[39] See also Cohen [56]. For the psychological impact of surveillance see Karabenick and Knapp [57].

[40] "More unsettling still is the potential combination of surveillance technologies with neuroanalytics to reveal, predict, and manipulate instinctual behavioral patterns of which we are not even aware" [54, p. 265]. Up to the fear that "Based on the technology available, the emergence of a 'Walden 3.0′ with control using positive reinforcements and behavioral engineering seems a natural development." [7, p. 265]. Walden 3.0 would be the manifestation of "Walden Two," the utopian novel written by behavioral psychologist B. F. Skinner (first published in 1948) embracing the proposition that even human behaviour is determined by environmental variables; thus, systematically altering environmental variables can generate a sociocultural system driven by behavioral engineering.

[41] See also Citron [59]; Coleman [60]; Marwick [61, p. 22].

[42] This phenomenon is particularly problematic for jurists since "[o]ne of the great accomplishments of the legal order was holding the sovereign accountable for decisionmaking and giving subjects basic rights, in breakthroughs stretching from Runnymede to the Glorious Revolution of 1688 to the American Revolution. New algorithmic decisionmakers are sovereign over important aspects of individual lives. If law and due process are absent from this field, we are essentially paving the way to a new feudal order of unaccountable reputational intermediaries" [63, p.19].

important to investigate legal and ethical rules that can be applied to private entities as well, beyond their potential involvement in governmental operations.[43]

Despite the fact that scholars have extensively stressed the potential chilling effect[44] of the phenomenon, we claim that the "risk" accompanying the opportunities transcends a disruption of the public arena and casts on individual and collective lives the shadow of an unseen conformation pressure that relentlessly removes the right to be divergent from the available modes of acceptable behaviour, action and reaction, in our commercial and personal dealings. The individual and social impact goes beyond the risks posed by continuous surveillance technology and reduces "the acceptable spectrum of belief and behavior". As it has been stressed, continuous surveillance may result in a "subtle yet fundamental shift in the content of our character". It arrives "not only to chill the expression of eccentric individuality, but also, gradually, to dampen the force of our aspirations to it" [15].[45]

Yet, what we are describing in the classifying society is more than, and different from, surveillance. While surveillance might impair individuals' and groups' ability to "come together to exchange information, share feelings, make plans and act in concert to attain their objectives" [74, p. 125], the classifying society can prevent us from thinking and behaving in our private relationships in a way that diverges from the various models we are pigeonholed into.

Platform neutrality is also in danger in the classifying society since large players and intermediaries have the ability to distort both commerce and the public sphere by leveraging their size or network power or big data.[46] The issue becomes more problematic once we look at the legal protections claimed by intermediaries such as Google.[47] Indeed, the claim of such intermediaries to being merely neutral collectors of preferences[48] is often undermined by the parallel claim they make, as corporations, they enjoy a free speech defence [81] that would allow them to manipulate results in favour of (or contrary to), for example, a political campaign, or a competitor, or a cultural issue. Such a result is somehow acknowledged in the American legal system by Zhang v. Baidu.com Inc.,[49] which affirms that a for-profit platform's selection and arrangement of information is not merely copyrightable, but also represents a form of free speech.[50]

[43] Government actions have triggered and driven a critical debate. See for instance Ramasastry [68]; Slobogin [69]; Solove [70].

[44] On the issue see also Solove [71]; Cate [72]; Strandburg [73].

[45] See also Schwartz [13].

[46] A serious concern shared both in Europe [75] and in the USA [76], stressing the systematic suppression of conservative news.

[47] Of course, it is not only Google that is the problem [77].

[48] It was the case in the Google Anti-defamation league [78]; see also Woan [79]; Wu [80].

[49] 2014 WL 1282730, at 6 (SDNY 2014) ("[A]llowing Plaintiffs to sue Baidu for what are in essence editorial judgments about which political ideas to promote would run afoul of the First Amendment.").

[50] *Contra e.g.*, Case C-131/12.

This latter example of corporate strategic behaviour, switching from one to another self-characterization, illustrates very clearly the need for a transnational legal intervention,[51] or, at least, the need to overcome the stand-alone nature of legal responses that allows a company to use one approach under competition law and another one under constitutional law, for instance.[52]

Yet, this very opportunistic approach should make us wonder if it could be claimed as well by the individuals sensed by algorithms. Since sensed and inferred data are mostly protected as trade or business secrets, it would be worth exploring the possibility of protecting individual anonymities as individual trade secrets in shaping their own bargaining power.

2.2 *On Discrimination, (Dis) Integration, Self-Chilling, and the Need to Protect Anonymities*

Analytical results may have a discriminatory—yet sometimes positive [85]—impact based on unacceptable social factors (such as gender, race, or nationality), on partial information (which being by definition incomplete is typically thereby wrong), on immutable factors over which individuals have no control (genetics, psychological biases, etc.), or on information the impact of which is unknown to the individual (e.g. a specific purchasing pattern). With reference to the latter it can also be the case of a plain and tacit application of a generalization to a specific individual [86, p. 40],[53] or of plain errors in the data mining process [68] amplifying the risks of misclassification or over-classification we have anticipated above.

It is also rather obvious that data mining casts data processing outside of "contextual integrity" [88], magnifying the possibility of a classification based on clusters of data that are taken literally out of context and are therefore misleading. For instance, assume that, for personal reasons unrelated to sexual habits, someone goes through an area where prostitution is exercised. This individual does it at regular times, every week, and at night. The "traffic" in the area forces a slow stop-and-go movement, compatible with attempts to pick up a prostitute, and such a movement is tracked by the geo-localization of one or more of the devices the individual is carrying along. Accordingly, once connected to the web with one of those devices (recognized by one or more of their components' unique identification numbers, such as their Wi-Fi or Bluetooth antenna) the device is targeted by advertising of pornographic materials, dating websites, sexual enhancement drugs, or remedies for sexually transmitted diseases because it fits the model of a sex addict prone to mercenary love.

[51] A need already signalled in the literature [82, 83].

[52] The high risks of enabling such a free speech approach have been highlighted [84; 7, p. 269].

[53] See also Pasquale and Citron [48]; Zarsky [87].

Note that we intentionally switched from referring to the human individual to referring to his/her devices and to the fact data collected, which are directly and exclusively related to the device. There is no need to even relate the cluster of data to personal identifying information such as ownership of the car or contracts for the services provided to the device. Nor is there any need to investigate the gender of the individual since to apply the model it is sufficient that a component in the device is characterized by features "normally" referable to a specific gender (e.g. the pink colour of the smartphone back cover). Of course, the more our society moves towards the Internet of Things, the more the classifications will be related to individual things and groups of related things (related, for instance, because they are regularly physically close to one another—such as a smartphone, a wallet, a laptop, . . .). Targeting groups of related things can be even more accurate than targeting the human individual directly. Indeed, some common sharing of characteristics among things increases the salience of each bit of individual information; for instance their "place of residence" (home) is detected by the fact that they all regularly stay together at the same geo-localized point at night, but variations in their uses offer the possibility of fine tuning the targeting by, for instance, noting that while at "home" the group shows a different behavioural pattern (e.g. related to a different gender/age because a companion or family member is using the device). Accordingly, targeting the device at different hours of the day may expand the reach of the classifying society, again without the need to resort to personal identifying information and with a higher granularity. To the contrary, the classifying society avoids personal identifying information because such information reduces the value of the different clusters of data referable in the various models to the device or to the group of things.

This example clearly illustrates how and why a change of paradigm is required: to consider and regulate anonymities—not only identities—in data protection.

This rather simple example of analytical mismatch also triggers the issue of the harm caused to the undiscovered individual behind the devices. It might be problematic to claim his/her privacy has been intruded. Arguably, the new (erroneous) knowledge produced to sell her/his devices' data for targeted advertising could even be protected as a trade secret. Apparently the model is acting on an error of "perception" not very different from the one a patrolling police officer might have in seeing an individual regularly in a given area, week after week, and deciding to stop him for a "routine control". Yet, while the impact of a regular and lawful search by police is trivial and might end up correcting the misperception, the erroneous fitting to the model of mercenary love user would lead to a variety of problems—such as unease in displaying in public or with relatives, children, spouse, . . . ones' own web searches due to the advertising they trigger and that cannot be corrected by traditional privacy preserving actions (cookies removal, cache cleaning, anonymous web searches, . . .) and reduced ability to exploit searches since "personalized" results pollute the results of such searches on the given device [39].

These are all forms of privacy harm that are difficult to uncover and even more problematic to prove and remedy. In the words of Ryan Calo [89]: "The objective category of privacy harm is the unanticipated or coerced use of information

concerning a person against that person. These are negative, external actions justified by reference to personal information. Examples include the unanticipated sale of a user's contact information that results in spam and the leaking of classified information that exposes an undercover intelligence agent." In ordinary people's lives this might involve, for instance, "the government [leverage of] data mining of sensitive personal information to block a citizen from air travel, or when one neighbor forms a negative judgment about another based on gossip" [89, p. 1143].[54]

We claim here that the expanding use of algorithms amplifies the "loss of control over information about oneself or one's own attributes" [89, p. 1134] to a level beyond personal identification. Indeed, in several legal systems data protection is only triggered when personal identifiable information is at stake.[55] In the classifying society vulnerability does not require personal identification.[56]

Data mining is already contributing to a change in the meaning of privacy, expanding the traditional Warren and Brandeis legacy of a "right to be let alone" to privacy as "unwanted sensing" by a third party [90, pp. 225–226]. However, in the classifying society, the interplay among the various techniques for gathering and enriching data for analysis and during the mining process[57] (from self tracking [91] to direct interaction or company logs [11], or the intermediary logging of data such as google searches or cookies or the purchase of data by a broker to integrate and augment existing data bases)[58] has reached a point in which data protection as we know it is powerless and often effectively inapplicable to modern data mining technologies.

Data mining, the "nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data" [93],[59] is producing a deep change in our thinking paradigm. We used to consider data in contexts, to track them to individuals. Algorithms, big data, and the large computing ability that connect them do not necessarily follow causal patterns and are even able to identify information unknown to the human individual they refer to.[60] Pattern (or "event-based") and subject-based data mining [69, p. 438] search for patterns that describe events and relate them. They result in both descriptive and predictive results. In the latter case, data analysis generates new information based on previous data. This information

[54] According to R. Calo [89] harm must be "*unanticipated* or, if known to the victim, *coerced*".

[55] This is the case for both the EU and the USA. See for instance California Online Privacy Protection Act, CAL. BUS & PROF. CODE §§ 22575–22579 (West 2004) (privacy policy requirement for websites on pages where they collect personally identifiable information); CAL. CIV. CODE §§ 1785.11.2, 1798.29, 1798.82 (West 2009); CONN. GEN. STAT. ANN. § 36a-701b (West 2009 & Supp. 2010); GA. CODE ANN. § 10-1-910, 911 (2009).

[56] See footnotes 35 and 40 and accompanying text.

[57] Mining itself generates new data that change the model and the reading of the clusters.

[58] Clarke [92] defines dataveillance as "the systematic use of personal data systems in the investigation or monitoring of the actions or communications of one or more persons".

[59] However, there are several technical definitions of data mining available but they all refer to the discovery of previously unknown, valid patterns and relationships.

[60] This is the case for a recent study on pancreatic cancer [94].

should be able to predict outcomes (e.g. events or behaviours) by combining previous patterns and new information. The underlying assumption is that results found in older data apply to new ones as well, although no causal explanation is provided even for the first set of results.

Datasets are actively constructed and data do not necessarily come from the same sources, increasing the risks related to de-contextualization. Moreover, since an analyst must define the parameters of the search (and of the dataset to be searched), human biases can be "built in" to the analytical tools (even unwillingly),[61] with the additional effect of their progressive replication and expansion once machine learning is applied. Indeed, if the machine learning process is classified as "non-interpretable" (by humans, sic!), for instance because the machine learned scheme is assessing thousands of variables, there will not be human intervention or a meaningful explanation of why a specific outcome is reached (e.g. a person, object, event, ... is singled out and a group of devices is targeted as a sex addict).

In the case of interpretable processes (translatable in human understandable language) a layer of human intervention is possible (although not necessary), as in, for example, the police search. Again this possibility is a double-edged sword since human intervention can either correct biases or insert new ones by interfering with the code, setting aside or inserting factors [95].

The lack of (legal and ethical) protocols to drive human action in designing and revising algorithms clearly calls for their creation, but requires a common setting of values and rules, since algorithms are basically enjoying a-territoriality in the sense that they are not necessarily used in one given physical jurisdiction. Moreover, the expansion of the autonomous generation of algorithms calls for building similar legal and ethical protocols to drive the machine generation of models. In both cases, there is also a need for technological verifiability of the effectiveness of the legal and ethical protocols making human readable at least the results of the application of the model when the model is not readable by humans.

Interpretable processes are certainly more accountable[62] and transparent[63] (although costlier),[64] but they still present risks. Moreover, by using (or translating into) interpretable processes, data mining will increase the possibility of getting back to searching for causality in the results instead of accepting a mere statistical association[65]: the group of devices passing through a city's solicitation block could be cleared of the social stigma the model has attached to them with all of its ensuing consequences. This is an important note since data mining results are NOT based

[61] For an explanation of the actual mechanisms see Solove [70].

[62] Meaning agents have an ethical and sometimes legal obligation to answer for their actions, wrongdoing, or mistakes.

[63] Transparency is intended as the enabling tool for actual accountability.

[64] The different cost-impact of the level of transparency required is analysed by Zarsky [52].

[65] Efforts to generate Transparency Enhancing Tools (TETs) is producing an expanding body of research at the crossroad between law and technology [96]. But on the side effects and risks of an excess of transparent information see Shkabatur [97].

on causality in the traditional sense but instead on mere probability.[66] Human readability would also enable a confidence check on the level of false positives in the rule produced by the algorithm, and, *ex post*, on the number of false negatives that were missed by the predictive model.[67]

What comes out of the preceding analysis is also a more nuanced notion of anonymity for the classifying society. After all, as clearly illustrated in the literature, anonymous profiles are "quantified identities imposed on individuals, often without their consultation, consent, or even awareness" [48, p. 1414].[68] The concept of anonymity[69] has influenced the very notion of personal data that is still defined in the EU GDPR 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 one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person".[70] Meanwhile a "nothing-to-hide" cultural approach has been blooming with the expansion of the so-called web 2.0,[71] generating a less prudent approach to the data our devices generate and share. Most (not to say all) apps and software we use "require" access to (and also generate) an enormous amount of data that are unrelated to the purpose of the software or app. Those data are widely shared and subsequently fed in to a myriad of models that are then applied to the same and other sets of data-generating "things" (not necessarily to the individual owner of them).

Thus the technological promise of data protection through anonymization has been defeated by technology itself. Moreover, "... where anonymity for the sake of eliminating biases is desirable, one cannot assume that technical anonymity by itself guarantees the removal of bias. Rather, if technical anonymity allows for biased misattribution, then at least in some contexts, there may need to be additional steps introduced to mitigate this possibility" [105, pp. 178–179].

Finally, it is the very notion of pseudo-anonymous data[72] that provides the key to bypass entirely personal data protection in the classifying society.

---

[66] On this issue see in general Zarsky [22, 98].

[67] Literature concentrates on the potential harms of predictive algorithms [67].

[68] The "undiscovered observer represents the *quintessential* privacy harm because of the unfairness of his actions and the asymmetry between his and his victim's perspective" [89, p. 1160].

[69] The literature on the obstacles to obtaining acceptable levels of anonymity on the web is immense [39, 99–101]. See also Datalogix [102] privacy policy.

[70] Art. 4 GDPR.

[71] See the seminal work of Solove [103]; see also Zarsky [104].

[72] According to the EU GDPR (art. 4) "pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person".

2.3 *Data and Markets: A Mismatch Between Law, Technology, and Businesses Practices*

A modern clash of rights seems to emerge in the classifying society—between, on the one hand, the right of individuals to control their own data, and, on the other, the interest of business in continuously harnessing that data[73] as an asset. The latter are increasingly protected by IP or quasi IP as trade secrets.[74]

This clash is echoed among economists. Some support business arguing that more (customers') data disclosure reduces information asymmetries [108, 109] while others argue it is data protection that generates social welfare [110]. In this debate legal scholars mostly advocate more individuals' control over their data at least in terms of an extended propertization [111, 112].[75] However, theories [113, 116, 117] about how a market for personal data, acknowledging data subjects' property rights, do not seem to consider the actual technological state of art [21] and the corresponding legal landscape.

In this landscape the legal framework remains dominated by the Privacy Policy Terms and Conditions whose legal enforceability remains at least doubtful in many legal orders. Meanwhile, *de facto* PPTC are enforced by the lack of market alternatives, by the widespread lack of data subjects' awareness both of the existence of tracking procedures and of the role of data aggregators and data brokers.[76] The hurdles and costs of a lawsuit close this vicious loop in which, although doubtful in legal terms, PPTC actually govern the data processing beyond statutory constraints in any given legal system.

In addition, the composite nature of most transactions makes opaque, to say the least, information collection and sharing processes. Very often, data processing is externalized [119] making it even more opaque and difficult to track data once they are shared.

Note also that some key players such as Google, Amazon, and e-Bay do not identify themselves as data brokers even though they regularly transfer data[77] and generate models.

Further and subsequent processing, even if data are anonymized or pseudo-anonymized, generates models that are later applied to the same (or other) groups of shared characteristics, a process which impacts individuals whose data, although technically non-personal, matches the model. In other words, the production cycle

[73] An average of 56 parties track activities on a website [106]. On the evolution of personal data trade see World Economic Forum [107].

[74] "No one can challenge the process of scoring and the results because the algorithms are zealously guarded trade secrets" [10, p. 5]. As illustrated by Richards and King [66, p. 42], "[w]hile Big Data pervasively collects all manner of private information, the operations of Big Data itself are almost entirely shrouded in legal and commercial secrecy".

[75] But see for concerns over propertization Noam [113]; Cohen [114]; Bergelson [115].

[76] The role of data aggregation and data brokers is vividly illustrated by Kroft [118].

[77] See e.g. Kosner [120].

we are describing (data collection of personal/things data, their enrichment with other data, and their pseudo-anonymization, the generation of the model and its continuous refinement) allows for application of the model to individuals without the need to actually involve further personal data (in legal technical terms), entirely bypassing data protection laws—even the most stringent ones.

Paradoxically, as it might seem that such a process could easily pay formal homage to a very strict reading of the necessity principle and the ensuing data minimization. Incidentally, the design of privacy policy terms that require (better yet: impose) consent in order to access the required services in a technological ecosystem that does not provide alternatives[78] trumps altogether those principles, while a business model based on the classifying society and on things' data (that is pseudo-anonymized data) disregards those principles altogether. In such a scenario, the data minimization rule normally required under the EU data protection directive[79] to distinguish between information needed to actually deliver the good or service and additional data becomes irrelevant—in any event, it has never acquired actual grip in business practice due to the low enforcement rate and despite its potential ability to increase data salience [119].

Technology experts have developed various concepts to enhance privacy protection—e.g. k-anonymity [123], l-diversity [124] and t-closeness [125]—but these concepts have revealed themselves as non-conclusive solutions. A similar fate is common to other privacy-enhancing technologies (PETs) [126, 127] that rely on the assumed preference for anonymity, clearly (yet problematically) challenged by the use of social networks,[80] while Do-Not-Track initiatives have not obtained enough success either [133]. And yet, the importance of all these attempts is becoming progressively less relevant in the classifying society. Accordingly, users must keep relying on regulators to protect them and on businesses to follow some form of ethical behaviour[81] calling for more appropriate answers from both stakeholders.

As mentioned, device fingerprints enable identification of individuals and devices themselves even when, for instance, cookies are disabled. If we consider that each device is composed of several other devices with their own digital fingerprints (e.g. a smart phone has a unique identification number while its Bluetooth and Wi-Fi antennas and other components also have their own), and that each of them continuously exchange information that can be compiled, the shift from personal

[78] Other authors have already pointed out that one key reading of privacy in the digital age is the lack of choice about the processes that involve us and the impossibility of understanding them [121, p. 133].

[79] See now the GDPR; for a technical analysis see Borcea-Pfitzmann et al. [122].

[80] See Gritzalis [128]. Indeed several authors have already highlighted the risks to privacy and autonomy posed by the expanding use of social networks: see, for instance the consequent call for a "Social Network Constitution" by Andrews [129] or the proposal principles of network governance by Mackinnon [82] or the worries expressed by Irani et al. [130, 131]; see also Sweeney [123]; Spiekermann et al. [132].

[81] See Fujitsu Res. Inst. [134].

data to things' data and to clusters of things' (non-personal) data as the subject of data processing begins to emerge in all its clarity and pervasiveness. When the GSM of a mobile phone allows geo-localization, we are used to thinking and behaving as if the system is locating the individual owner of the mobile phone. Yet technically it is not; it is locating the device and the data can be easily pseudo-anonymized. Once it becomes possible to target the device and the cluster(s) of data contingently related to it in a model, without the need to expressly target a specifically identified individual, the switch from personal data to things' data is complete, requiring an alternative approach that has not yet fully emerged.

The situation described above leads to a deadly stalemate of expanding asymmetry among players in markets and in society overall.

The overarching concept behind the classifying society is that individuals are no longer targetable as individuals but are instead selectively addressed for the way in which some clusters of data that they (technically, one or more of their devices) share with a given model fit in to the model itself. For instance, if many of our friends on a social network have bought a given book, chances are that we will do it as well; the larger the number of people in the network the more likely it is we will conform. What results is that individuals (or, rather, the given cluster of information) might be associated with this classification (e.g. high buying interest) and be treated accordingly—for instance, by being advertised a higher price (as happened for OS users versus Windows users).

Note that we normally do not even know such a classification model exists, let alone that it is used. Moreover, if engineered well from a legal standpoint, a data processing might not even amount to a *personal data* processing, preventing any meaningful application of data protection rules. On the other hand, the construction of models on a (very) large array of data that might or might not include ours is beyond our control and is normally protected as a business asset (e.g. as a trade secret).

In addition, since the matching between individuals and the model just requires a reduced amount of data, the need for further enriching the cluster of data with personal identifying information is progressively reduced. Thus, it becomes increasingly easy for the model to be uninterested in "identifying" us—instead contenting itself to target a small cluster of data related to a given device/thing (or ensemble of devices)—because all of the relevant statistical information is already incorporated in the model. The model need not identify us nor search for truths or clear causality as long as a given data cluster statistically shares a given amount of data with the model. Moreover, even if the model gets it wrong, it can learn (automatically with the help of AI and machine learning techniques) from the mistake, refining itself by adding/substituting/removing a subset of data or giving it a different weight that adjusts the statistical matching results: (big) data generate models and the use of models generates new data that feed the models and change them in an infinite creative loop whose human (and especially data subject) control is lost [7].

Such a deep change is unsettling for data protection because it undermines any legal and technological attempt (e.g. of anonymization techniques) to intervene.

Also, it can lead to unpleasant—even if unsought—results. The societal impact and risks for freedom of speech have been highlighted elsewhere. Yet, the selection of information and the tailoring of the virtual (and real) world in which we interact, even if not concerted among businesses,[82] reduce the chances of producing divergent behaviours with the result of actually reinforcing the model, on the one hand, and reducing the possibility of divergence even further on the other.[83]

Accordingly, technical solutions against privacy loss prove unhelpful if actions are based upon the correspondence of a small number of shared characteristics to a model. Indeed, privacy-preserving techniques, privacy-by-design, and other forms of technical protection for privacy all aim at reducing the amount of information available to prevent identification or re-identification of data [135][84] while personal identification as such is fading away in the encounter between technology and businesses models driven by algorithms.

This state of affairs calls for further investigation and a different research approach.

3 A Four-Layer Approach Revolving Around Privacy as a Framework

As anticipated, the data economy offers many advantages for enterprises, users, consumers, citizens, and public entities offering unparalleled scope for innovation, social connection, original problem-solving, and problem discovery using algorithms and machine learning (e.g. on emerging risks or their migration patterns such as in pandemic emergencies).[85] The actual unfolding of these possibilities requires the gathering and processing of data on both devices and individuals that raises divergent concerns [140] related to, on one hand, the will to share and participate (in social networks, for instance, or in public alerting systems in the case of natural disasters or pharmacovigilance), and, on the other hand, to the reliability of the organisations involved, regardless of their private or public nature.[86]

The complete pervasion of interactive, context-sensitive mobile devices along with the exponential growth in (supposedly) user-aware interactions with service providers in the domains of transportation and health-care or personal fitness,

[82] We are not discussing a science fiction conspiracy to control human beings but the actual side effects of the embrace of specific technological advancements with specific business models and their surrounding legal constraints.

[83] This holds true also when the code is verified or programmed by humans with the risk of embedding in it, even unintentionally, the biases of the programmer: "Because human beings program predictive algorithms, their biases and values are embedded into the software's instructions, known as the source code and predictive algorithms" [10, p. 4].

[84] See also Danezis [136].

[85] See also Bengtsson et al. [137]; Wesolowski et al. [138, 139].

[86] See also Wood and Neal [141]; Buttle and Burton [142]; Pew Research Centre [143]; Reinfelder [144].

for instance, dynamically generates new sources of information about individual behaviour or, in our framework, the behaviour of (their) devices—including personal preferences, location, historical data, and health-related behaviours.[87] It is already well known that in the context of mobile cells, Wi-Fi connections, GPS sensors, and Bluetooth, et cetera, several apps use location data to provide 'spatial aware' information. These apps allow users to check in at specific locations or venues, to track other users' movements and outdoor activities, and to share this kind of information [151]. Health-state measurement and monitoring capabilities are being incorporated into modern devices themselves, or provided through external devices in the form of smart watches, wearable clips, and bands.

The information provided can be employed for health care, for personal fitness, or in general for obtaining measurable information leading to marvellous potentials both in public and private use. Nevertheless, once cast in light of the blossoming role of algorithms in the classifying society this phenomenon contributes to the above-mentioned concerns, and calls for a clear, technologically-neutral regulatory framework that moves from privacy and relates to all legal fields involved [152], empowering balanced dealing between individual data subjects/users and organised data controllers.

The legal landscape of algorithm regulation requires that traditional approaches to privacy, conceived as one component of a multifaceted and unrelated puzzle of regulations, are transcended, regardless of the diversity of approaches actually taken in any given legal order (opt-in versus opt-out; market versus privacy as a fundamental right, ...). Only if data protection features at the centre of the regulatory process can it offer a comprehensive approach that does not overlook the link between social and economic needs, law and the technological implementation of legal rules, without underestimating the actual functioning of specific devices and algorithms-based business models.

The general trends of technology and business modeling described so far demonstrate the need to embed effective protection in existing legal data control tools (privacy as a fundamental right, for instance, or privacy by default/design) and the eventual need/opportunity to introduce *sui generis* forms of proprietary protection (personal data as commodities) to counterbalance the expanding protection of businesses in terms of both trade secrets and recognition of fundamental freedoms.[88]

In this framework, predictive algorithms are the first to raise concerns—especially in terms of consent, awareness (well beyond the notice model), and salience. Our privacy as readers and our ability to actually select what we want to read is an important issue that promptly comes to the fore when predictive algorithms are used [154].[89]

[87] See also Elkin-Koren and Weinstock [145]; FTC Staff Report [146–148]; Canadian Offices of the Privacy Commissioners [149]; Harris [150].

[88] See for instance Zhang v. Baidu.com, Inc (2014). But see, *e.g.*, Case C-131/12; Pasquale [153].

[89] On the privacy concerns and their social impact see Latar and Norsfors [155]; Ombelet and Morozov [156].

Nevertheless, using privacy as a framework reference for other areas of law in dealing with the classifying society would require an integrated strategy of legal innovation and technical solutions.

3.1 Revise Interpretation of Existing Rules

The first layer of this strategy would leverage the existing sets of rules.

In the actual context of algorithms-centrality in our societies at any relevant level, we claim that it is necessary to explore the potential interplay between the fundamental principles of privacy law and other legal rules such as unfair practices, competition, and consumer protection. For instance, although it may be based on individual consent, a practice that in reality makes it difficult for users to be aware of what data processing they are actually consenting to might be suspicious under any of the previously mentioned sets of legal rules. Business models and practices that segregate customers in a technically unnecessary way (e.g. by offering asymmetric internet connections to business and non-business clients), directing some categories to use the services in certain ways that make data collection easier and more comprehensive (e.g. through concentrated cloud-based services or asymmetric pricing) might appear suspicious under another set of regulations once privacy as a framework illustrates their factual and legal implications.

Furthermore, a business model that targets data clusters directly but refers to individuals only indirectly, using models generated by algorithms without clearly acknowledging this, can be questionable under several legal points of view.

Within this landscape dominated by opacity and warranting transparency and accountability, a 'privacy as a framework' approach is required to overcome a siloed legal description of the legal rules in order to cross-supplement the application of legal tools otherwise unrelated among each other. This layer of the approach envisions not unsettling regulatory changes by the rule makers but the use of traditional hermeneutic tools enlightened by the uptake of the algorithm-technological revolution.

In this vein some authors have begun to explore the application of the unconscionability doctrine to contracts that unreasonably favour data collecting companies [157].[90] To provide another example, once we tackle the topic of predictive algorithms, it is unavoidable to delve into the unfair commercial practices legal

[90]The authors also propose: "mandatory active choice between payment with money and payment with data, ex post evaluation of privacy notices, democratized data collection, and wealth or income-responsive fines". Their proposals enrich an already expanding host of regulatory suggestions. See Hajian and Domingo-Ferrer [158]; Mayer-Schonberger and Cukier [159]; Barocas and Selbst [19]. For a more technical account on fostering discrimination-free classifications, see Calders and Verwer [160]; Kamiran et al. [161]. Recently, the establishment of an ad hoc authority has also been advocated [162]. On market manipulation through the use of predictive and descriptive algorithms see the seminal work of Calo [38].

domain. All western societies devote a specific regulation to them.[91] In a March 2014 panel, the Federal Trade Commission [163] identified some topics relevant to our discussion[92]:

- "How are companies utilizing these predictive scores?
- How accurate are these scores and the underlying data used to create them?
- How can consumers benefit from the availability and use of these scores?
- What are the privacy concerns surrounding the use of predictive scoring?
- What consumer protections should be provided; for example, should consumers have access to these scores and the underlying data used to create them?"

Predictive algorithms decide on issues relevant to individuals "not because of what they've done, or what they will do in the future, but because inferences or correlations drawn by algorithms suggest they may behave in ways that make them poor credit or insurance risks, unsuitable candidates for employment or admission to schools or other institutions, or unlikely to carry out certain functions." The panel concluded by insisting on "transparency, meaningful oversight and procedures to remediate decisions that adversely affect individuals who have been wrongly categorized by correlation." The need to "ensure that by using big data algorithms they are not accidently classifying people based on categories that society has decided—by law or ethics—not to use, such as race, ethnic background, gender, and sexual orientation" was considered urgent [164, pp. 7–8].

Yet, if we are concerned about de-biasing the algorithms in order to avoid "already" non-permitted discriminations[93] and to promote transparency (intended both as accountability and as disclosure to customers),[94] we should be even more concerned about the risks posed by "arbitrariness-by-algorithm", which are more pressing for the million facets of our characteristics, either permanent or temporary—such as our momentary mood, or our movement through a specific neighbourhood—that the classifying society makes possible.

[91] See the EU Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2000, concerning unfair business-to-consumer commercial practices in the internal market.

[92] We do not report here those topics that are exclusively related to the USA on which the FTC has authority, such as the relevance of the Fair Credit Reporting Act.

[93] For a recent account of algorithms' disparate impact see Barocas and Selbst [19, p. 671].

[94] See an analysis of some techniques potentially available to promote transparency and accountability [165]. See also Moss [24, p. 24], quoting relevant American statutes. Yet if action is not taken at a global level, online auditing can be run in countries where it is not forbidden and results transferred as information in other places. Analogously, a technical attempt to create auditing by using volunteers profiling in a sort of crowdsourcing empowering exercise might even make permissible online auditing in those mentioned jurisdictions forbidding the violation of PPT of web sites by using bots. There is an ongoing debate on this issue. See Benkler [166]; Citron [167]. But see *contra* Barnett [168]. For a critical analysis urging differentiation of the approach targeting the specific or general public see Zarsky [52].

Once we begin to ask ourselves whether it is fair is to take advantage of the vulnerabilities [169], of health-related[95] or other highly sensitive information [171], the need for a legal and ethical paradigm emerges vigorously and calls for reinterpreting the existing rules to expand their scope.

Indeed, once it is acknowledged that ToS and PPTCs are normally not read before accepting them [172],[96] and that the actual flow of data is often not even perceived by users because it runs in the background, the only apparent safeguard would seem to be an application of the necessity principle that not only limits data gathering and processing but that is normatively prescriptive well beyond authorizing a data processing.

For example, everybody realises that pre-installed apps[97] on ICT devices do not even allow activation of the apps themselves before consenting to the ToS and PPTCs, let alone the kind and extent of personal data collection and processing that they involve.[98] A sound application of the necessity principle (and of the privacy by design and by default approach to regulation) would impose a deep change in the design of web sites, of distributed products, of the structure of licence agreements, and of patterns of iterative bargaining even before dealing with the content of clauses on data processing and further uses of collected or given data.

For this reason, both the re-interpretation of existing legal rules and the eventual enactment of new ones are both necessary components of the overall strategy.

Such a process, which reverses the ongoing deregulation output via private law embedded in the content of the terms and in the design of their deployment in the progress of the business relationship (the pattern of alluring customers to buy), is rather intrusive on the part of the State and might be better pursued by reinterpreting existing private law rules governing the relationship between data subjects and data processors and/or service/product providers. This means enabling an interpretation of existing remedies (for instance against unfair business practices or of tort law rules, consumer protection statutes, and mistake and errors doctrines, ...) enlightened by the deep and subterranean change in the technological and legal landscape we have described so far.

For instance, if data protection rules are insufficient to prevent every installed app from claiming access to the full range of personal data present on a device in order to work, perhaps the (unknown or not read) terms requesting such an extension of access can be considered illicit—as unfair terms, unfair practices, or as anticompetitive, thereby triggering traditional contractual or liability remedies.

Reinterpretation, however, presents its own limits and certainly cannot cover every facet of the classifying society.

[95] On the potential for discriminatory and other misuses of health data regularly "protected" by professional secrecy see Orentlicher [170].

[96] Indeed, it has been estimated that on average we would need 244 h per year to read every privacy policy we encounter [173].

[97] Here, app is used as a synonym for software.

[98] See also Lipford et al. [174]; Passera and Haapio [175].

3.2 *Promote Changes in Regulation and Empower Data Subjects Technologically*

On a different level, the approach could lead to: (1) a reclassification of legal rules applicable when algorithms play a key role related to privacy along innovative lines; (2) the production of new rules in the same instances.

Such an approach would influence and inform the entire regulatory framework for both public and private activities, aiming at overcoming the pitfalls of an approach that maintains separate rules, for instance, on data protection, competition, unfair practices, civil liability, and consumer protection. However, to date, such an unsettling move is not realistically foreseeable.

Innovative regulation would be required, for example, in rearranging the protection of the algorithms as trade secrets in order to serve an emerging trend of considering personal data as (at least partially) property of the data subject, thereby enabling some forms of more explicit economic exchange. However, this innovation would also require the development and deployment of appropriate technology measures.

Data generation is not treated as an economic activity for the data subject while data gathering from data processors is characterized as such. This is mostly due to the fact that technologies to harness and exploit data from the business side are largely available while technology on the data subject side has to date failed to empower data subjects with workable tools for maintaining economic control over the information related to them.

The need to strengthen the legal and technological tools of data subjects due to the scale and pervasive use of algorithms in our societies is clearly manifest in the literature across any scientific field.

From our point of view, however, the background premise of data subject empowerment is similar to the one we can make on consumer protection, for instance. Both consumer protection and data protection laws endow consumers/data subjects with a large set of rights with the aim of reducing information asymmetries and equalizing bargaining power. The effective implementation of these consumer rights is mainly based on information that businesses are required to provide at an important cost for them. Indeed, both the relevant legal framework and the EUCJ consider consumers to be rational individuals capable of becoming "reasonably well-informed and reasonably observant and circumspect" (Case C-210/96, par. 31), for instance by routinely reading food labels and privacy policies (Case C–51/94),[99] and of making efficient decisions so long as the relevant (although copious) information is presented to them in "a comprehensible manner".[100]

[99] See Case C–51/94, para 34, holding that consumers who care about ingredients (contained in a sauce) read labels (*sic*); see also Phillips [176]; Gardner [177]; Ayres and Schwartz [178].

[100] E.g. artt. 5, 6 and 10 of Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce').

However, proposing to provide information in order to remedy bargaining asymmetries and sustain a proper decision-making process does not seem to effectively fill the gaps between formal rules and their impact on the daily life of consumers/users.[101] Once information is technically provided[102] and contract/TOS/PPTCs[103] are formally agreed upon, customers remain trapped in them as if they had actually profited from mandated information transparency by having freely and rationally decided upon such terms. Despite all the costs for industries, the attempt to reduce asymmetries and to empower efficient consumer rights often remains a form of wishful thinking[104] with a significant impact on consumer trust and market growth.

Indeed, very few people read contracts when buying products or services online [172]. Actually, the expectation that they would do so is possibly unrealistic, as a typical individual would need, on average, 244 h per year just to read every privacy policy they encounter [173], let alone the overall set of information that is actually part of the contract and the contract itself [172].

Moreover, it is likely that users are not encouraged to read either terms and conditions or information related to the product/service due to systematic patterns in language and typographic style that contribute to the perception of such terms as mere document formalities [187] and/or make them uneasily accessible (see, for example, actual difficulties in reading labels and information displayed on packages).

The difficulty of reading terms and technical information stems out not only from their surface structure but also from readers' lack of familiarity with the underlying concepts [188] and the fact that they figure in a format totally unfamiliar and not plainly understandable to the majority of consumers. Moreover, the vagueness and complex language of clauses, along with their length, stimulate inappropriate and inconsistent decision patterns on behalf of users.

As a result, lack of effective information, unawareness, and perception of a loose control over the process of setting preferences have a decisive impact on consumer trust, leading to weakening consumption patterns and to expanding asymmetries of power between individual users/customers/consumers and businesses.

[101] See also Bar-Gill and Ben-Shahar [179]; Luzak [180, 181]; Purnhagen and Van Herpen [182].

[102] See for an information mandate approach: Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts; Directive on electronic commerce; Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market; Directive 2008/48/EC of the European Parliament and of the Council of 23 April 2008 on credit agreements for consumers and repealing Council Directive 87/102/EEC; Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights and Regulation (Eu) No 531/2012 of the European Parliament and of the Council of 13 June 2012 on roaming on public mobile communications networks within the Union.

[103] See McDonald et al. [183].

[104] See in the economics literature: Stigler [184]; Akerlof [185]; Macho-Stadler and Perez-Castrillo [186].

Reversing this state of affairs to actually empower users/data subjects to select and manage the information they want (and want to share), and to expand their bargaining powers as a consequence of their ability to govern the flow of information, requires a deep change in approach. A change as deep as it is simple. Since businesses are not allowed to change the notice and consent approach[105] (although they largely manipulate it), a careful selection of useful information can be construed only from the other end (data subject/users).

The concept is rather simple: law forces the revelation of a large amount of information useful to users/data subjects, but the latter need an application tool to efficiently manage such information in order to actually become empowered and free in their choices.[106] While businesses cannot lawfully manage to reduce the amount of information they owe to users as consumers or data subjects, users can reduce and select that information for their improved use and for a more meaningful comparison in a functioning market.

A key feature of this layer in our approach is to embrace, technologically and in regulatory terms, a methodology pursuing an increased technological control over the flow of information, also with the assistance of legal rules that sustain such a control.

Yet, we are aware of the difficulties this layer entails. Indeed, it is difficult to regain a degree of control over a large part of data used by algorithms or by the models they create. Indeed, as off-line customers, our data are collected and processed either by relating them to devices we wear (e.g. e-objects), carry (e.g. smart phones), or use (e.g. fidelity cards, credit cards, apps, ...) when buying products or services (e.g. paying the highway or the parking with electronic means).

It is important to note that a personal data-safe technological approach (e.g. data are maintained in the control of the data subjects that licence access to them) would require that businesses actually consent to opt-in to a legal system that allows and enforces the potential for data subjects to avoid surrendering data beyond a clear and strict application of the necessity principle. Similarly, and where algorithms' functioning remains unveiled by the shadow of trade secrets protection, only a technological and cooperative approach can be of help. For instance, collective sharing of experiences along with shared (high) computational abilities can "reverse engineer" algorithms' outputs to detect old and new biased results.

In other words, without a strict enforcement (that requires both a technological and a legal grip) of the necessity principle in delivering goods and services, we will never see a different market for data emerge—and certainly not one in which data subjects get most of the economic benefits from data (even pseudo-anonymous) related to them.

[105] On the failure of the disclosure model see in general Ben-Shahar and Schneider [189], Radin [190], and with reference to privacy Ben-Shahar and Chilton [191].

[106] The issue of actual market alternatives is not addressed here.

3.3 Embed Ethical Principles in Designing Algorithms and Their Related Services

There are several reasons to attempt to embed ethical principles in the design of algorithms and their related services [192, 193].

First of all, to date any technological and legal strategy to rebalance the allocation of information, knowledge, and power between users/data subjects and businesses/data processors and brokers has proved unsuccessful.[107]

The reinterpretation of the solutions in any given legal system in light of technological developments and a sound privacy-as-a-framework approach would require a strong commitment by scholars, courts, and administrative authorities. In any event, it would require a certain amount of time to deploy its effects.

To the contrary, ethical principles and social responsibility are already emerging as drivers of businesses' decision-making. Ethical principles, once technologically incorporable, in designing and testing algorithms and their related services can use exactly the same technology used by algorithms to monitor algorithms' and business models' actual operation.

Also, the adoption of such principles by the community of code developers, for instance, does not in principle need to receive the approval of businesses since these would be ethical rules of the "profession".

4 A Summary as a Conclusion

In the classifying society a new target for protection emerges: personal anonymities, the clusters of characteristics shared with an algorithm-generated model that are not necessarily related to personal identifiable information and thus are not personal data in the strict meaning even of EU law.

The centrality of personal anonymities is blossoming with the maturing of the Internet of Things that exponentially increases the potential of algorithms and of their use unrelated to personal identifying information.

Personal anonymities stem from innovative business models and applied new technologies. In turn, they require a combined regulatory approach that blends together (1) the reinterpretation of existing legal rules in light of the central role of privacy in the classifying society; (2) the promotion of disruptive technologies for disruptive new business models enabling more market control by data subjects over their own data; and, eventually, (3) new rules aiming, among other things, to provide to data generated by individuals some form of property protection similar to that enjoyed by the generation of data and models by businesses (e.g. trade secrets).

[107] See above footnotes 80–81 and accompanying text.

The blend would be completed by (4) the timely insertion of ethical principles in the very generation of the algorithms sustaining the classifying society.

Different stakeholders are called to intervene in each of the above mentioned layers of innovation.

None of these layers seems to have a prevalent leading role, to date. However, perhaps the technical solutions enabling a market for personal data led by data subjects, if established, would catalyse the possibility of generating alternate business models more in line with the values reflected in personal data protection rules. Yet, the first step remains to put privacy at the centre of the wide web of legal rules and principles in the classifying society.

References

1. European Data Protection Supervisor, Opinion No 4/2015: Towards a new digital ethics: Data, dignity and technology. https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-09-11_Data_Ethics_EN.pdf. Accessed 24 Oct 2016
2. Angwin, J.: The web's new gold mine: your secrets. Wall Street J. http://online.wsj.com/article/SB10001424052748703940904575395073512989404.html (2010). Accessed 24 Oct 2016
3. Bain & Company: Using Data as a Hidden Asset. http://www.bain.com/publications/articles/using-data-as-a-hidden-asset.aspx (2010). Accessed 24 Oct 2016
4. Pariser, E.: The Filter Bubble. Penguin Press, New York (2011)
5. Tene, O., Polonetsky, J.: Big data for all: privacy and user control in the age of analytics. Northwest. J. Technol. Intellect. Prop. **11**(5), 239–273 (2013)
6. Ohm, P.: Response, the underwhelming benefits of big data. Univ. Pa. Law Rev. Online. **161**, 339–346 (2013)
7. Van Otterlo, M.: Automated experimentation in Walden 3.0: the next step in profiling, predicting, control and surveillance. Surveill. Soc. **12**(2), 255–272 (2014)
8. Lomas, N.: Amazon patents "anticipatory" shipping—to start sending stuff before you've bought it. https://techcrunch.com/2014/01/18/amazon-pre-ships/ (2014). Accessed 24 Oct 2016
9. Vaidhyanathan, S.: The Googlization of Everything. University of California Press, Berkeley (2011)
10. Citron, D.K., Pasquale, F.: The scored society: due process for automated predictions. Wash. Law Rev. **89**(1), 1–33 (2014)
11. Duhigg, C.: How Companies Learn Your Secrets. The New York Times, New York (2012). http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html. Accessed 24 Oct 2016.
12. Gray, D., Citron, D.K.: The right to quantitative privacy. Minn. Law Rev. **98**, 62–144 (2013)
13. Schwartz, P.M.: Privacy and democracy in cyberspace. Vanderbilt Law Rev. **52**, 1609–1701 (1999)
14. Schwartz, P.M.: Internet privacy and the state. Conn. Law Rev. **32**, 815–859 (2000)
15. Cohen, J.E.: Examined lives: informational privacy and the subject as object. Stanford Law Rev. **52**, 1373–1438 (2000)
16. Cohen, J.E.: Cyberspace as/and space. Columbia Law Rev. **107**(1), 210–256 (2007)
17. Solove, D.J.: The Digital Person. New York University Press, New York (2004)

18. Mattioli, D.: On Orbitz, Mac users steered to pricier hotels. Wall Street J.. **23**, 2012 (2012). http://www.wsj.com/articles/SB10001424052702304458604577488822667325882. Accessed 24 Oct 2016.
19. Barocas, S., Selbst, A.D.: Big data's disparate impact. Calif. Law Rev. **104**, 671–732 (2016)
20. Colb, S.F.: Innocence, privacy, and targeting in fourth amendment jurisprudence. Columbia Law Rev. **56**, 1456–1525 (1996)
21. Korff, D.: Data protection laws in the EU: the difficulties in meeting the challenges posed by global social and technical developments. In: European Commission Directorate-General Justice, Freedom and Security, Working Paper No. 2. http://ec.europa.eu/justice/policies/privacy/docs/studies/new_privacy_challenges/final_report_working_paper_2_en.pdf (2010). Accessed 24 Oct 2016
22. Zarsky, T.Z.: Governmental data mining and its alternatives. Penn State Law Rev. **116**(2), 285–330 (2011)
23. Bamberger, K.A.: Technologies of compliance: risk and regulation in a digital age. Tex. Law Rev. **88**(4), 669–739 (2010)
24. Moss, R.D.: Civil rights enforcement in the era of big data: algorithmic discrimination and the computer fraud and abuse act. Columbia Hum. Rights Law Rev. **48**(1) (2016).
25. Exec. Office of The President: Big data: seizing opportunities, preserving values. http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf (2014). Accessed 24 Oct 2016
26. Turow, J.: Niche Envy. MIT Press, Cambridge, MA (2006)
27. Al-Khouri, A.M.: Data ownership: who owns "my data"? Int. J. Manag. Inf. Technol. **2**(1), 1–8 (2012)
28. Rajagopal, S.: Customer data clustering using data mining technique. Int. J. Database Manag. Syst. **3**(4), 1–11 (2011)
29. Frischmann, B.M., Selinger, E.: Engineering humans with contracts. Cardozo Legal Studies Research Paper No. 493. https://ssrn.com/abstract=2834011 (2016). Accessed 24 Oct 2016
30. Perzanowski, A., Hoofnagle, C.J.: What we buy when we 'buy now'. Univ. Pa. Law Rev. **165**, 317 (2017). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2778072 (forthcoming 2017). Accessed 24 Oct 2014
31. Apple: Terms and Conditions—Game Center. http://www.apple.com/legal/internet-services/itunes/gamecenter/us/terms.html (2013). Accessed 24 Oct 2016
32. Behm, R.: What are the issues? Employment testing: failing to make the grade. http://employmentassessment.blogspot.com/2013/07/what-are-issues.html (2013). Accessed 24 Oct 2016
33. EFF: Timeline of NSA domestic spying. https://www.eff.org/nsa-spying/timeline (2015). Accessed 24 Oct 2016
34. Schneier, B.: Want to evade NSA spying? Don't connect to the internet. Wired Magazine. http://www.wired.com/opinion/2013/10/149481 (2013). Accessed 24 Oct 2016
35. Rosenbush, S.: Facebook tests software to track your cursor on screen. CIO J. http://blogs.wsj.com/cio/2013/10/30/facebook-considers-vast-increase-in-data-collection (2013). Accessed 24 Oct 2016
36. PrivacySOS: NYPD's domain awareness system raises privacy, ethics issues. https://privacysos.org/blog/nypds-domain-awareness-system-raises-privacy-ethics-issues/ (2012). Accessed 24 Oct 2016
37. TrapWire: The intelligent security method. http://www.trapwire.com/trapwire.html (2016). Accessed 24 Oct 2016
38. Calo, M.R.: Digital market manipulation. George Wash. Law Rev. **82**(4), 95–1051 (2014)
39. Ohm, P.: Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Rev. **57**, 1701–1777 (2010)
40. Hoofnagle, C.Y.: Big brother's little helpers: how choicepoint and other commercial data brokers collect and package your data for law enforcement. N. C. J. Int. Law Commer. Regul. **29**, 595–637 (2004)

41. Singer, N.: Mapping, and sharing, the consumer genome. The New York Times. http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html (2012). Accessed 24 Oct 2016
42. Tucker, P.: Has big data made anonymity impossible? MIT Technology Review. http://www.technologyreview.com/news/514351/has-big-data-madeanonymity-impossible/ (2013). Accessed 24 Oct 2016
43. Article 29 Data Protection Working Party: Opinion 5/2014 on anonymization techniques. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf (2014). Accessed 24 Oct 2016
44. Ruggieri, S., Pedreschi, D., Turini, F.: Data mining for discrimination discovery. ACM Trans. Knowl. Discov. Data. **4**(2), 1–40 (2010)
45. Hoofnagle, C.Y., Whittington, J.: "Free": accounting for the costs of the Internet's most popular price. UCLA Law Rev. **61**, 606–670 (2014)
46. Aziz, A., Telang, R.: What is a digital cookie worth? https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757325 (2016). Accessed 24 Oct 2016
47. Bozdag, E.: Bias in algorithmic filtering and personalization. Ethics Inf. Technol. **15**(3), 209–227 (2013)
48. Pasquale, F., Citron, D.K.: Promoting innovation while preventing discrimination: policy goals for the scored society. Wash. Law Rev. **89**(4), 1413–1424 (2014)
49. Pasquale, F.: Beyond innovation and competition: the need for qualified transparency in Internet intermediaries. Northwest. Univ. Law Rev. **104**(1), 105–174 (2010)
50. Pasquale, F.: Restoring transparency to automated authority. J. Telecommun. High Technol. Law. **9**, 235–256 (2011)
51. Zarsky, T.Z.: Thinking outside the box: considering transparency, anonymity, and pseudonymity as overall solutions to the problems in information privacy in the Internet society. Univ. Miami Law Rev. **58**, 1301–1354 (2004)
52. Zarsky, T.Z.: Transparent predictions. Univ. Ill. Law Rev. **2013**(4), 1503–1570 (2013)
53. Cohen, J.E.: Configuring the Networked Self: Law, Code, and the Play of Everyday Practice. Yale University Press, New Haven, CT (2012)
54. Citron, D.K., Gray, D.: Addressing the harm of total surveillance: a reply to professor Neil Richards. Harv. L. Rev. F. **126**, 262 (2013)
55. DHS: National network of fusion centers fact sheet. https://www.dhs.gov/national-network-fusion-centers-fact-sheet (2008). Accessed 24 Oct 2016
56. Cohen, J.E.: Privacy, visibility, transparency, and exposure. Univ. Chicago Law Rev. **75**(1), 181–201 (2008)
57. Karabenick, S.A., Knapp, J.R.: Effects of computer privacy on help-seeking. J. Appl. Soc. Psychol. **18**(6), 461–472 (1988)
58. Peck, D.: They're watching you at work. The Atlantic. http://www.osaunion.org/articles/Theyre%20Watching%20You%20At%20Work.pdf (2013). Accessed 24 Oct 2016
59. Citron, D.K.: Data mining for juvenile offenders. Concurring Opinions. http://www.concurringopinions.com/archives/2010/04/data-mining-for-juvenile-offenders.html (2010). Accessed 24 Oct 2016
60. Coleman, E.G.: Coding Freedom. Princeton University Press, Princeton (2013)
61. Marwick, A.E.: How your data are being deeply mined. The New York Review of Books. http://www.nybooks.com/articles/2014/01/09/how-your-data-are-being-deeply-mined/ (2014). Accessed 24 Oct 2016
62. Abdou, H.A., Pointon, J.: Credit scoring, statistical techniques and evaluation criteria: a review of the literature. Intell. Syst. Account. Finance Manag. **18**(2–3), 59–88 (2011)
63. Balkin, J.M.: The constitution in the national surveillance state. Minn. Law Rev. **93**(1), 1–25 (2008)
64. Kerr, O.S.: Searches and seizures in a digital world. Harv. Law Rev. **119**(2), 531–585 (2005)
65. Citron, D.K.: Technological due process. Wash. Univ. Law Rev. **85**(6), 1249–1313 (2008)
66. Richards, N.M., King, J.H.: Three paradoxes of big data. Stanford Law Rev. **66**, 41–46 (2013)

67. Crawford, K., Schultz, J.: Big data and due process: toward a framework to redress predictive privacy harms. Boston Coll. Law Rev. **55**(1), 93–128 (2014)
68. Ramasastry, A.: Lost in translation? Data mining, national security and the "adverse inference" problem. Santa Clara Comput. High Technol. Law J. **22**(4), 757–796 (2004)
69. Slobogin, C.: Government data mining and the fourth amendment. Univ. Chicago Law Rev. **75**(1), 317–341 (2008)
70. Solove, D.J.: Data mining and the security-liberty debate. Univ. Chicago Law Rev. **75**, 343–362 (2008)
71. Solove, D.J.: Privacy and power: computer databases and metaphors for information privacy. Stanford Law Rev. **53**(6), 1393–1462 (2001)
72. Cate, F.H.: Data mining: the need for a legal framework. Harv. Civil Rights Civil Liberties Law Rev. **43**, 435 (2008)
73. Strandburg, K.J.: Freedom of association in a networked world: first amendment regulation of relational surveillance. Boston Coll. Law Rev. **49**(3), 741–821 (2008)
74. Bloustein, E.J.: Individual and Group Privacy. Transaction Books, New Brunswick, NJ (1978)
75. Conseil National Numerique, Platform Neutrality: Building an open and sustainable digital environment. http://www.cnnumerique.fr/wp-content/uploads/2014/06/ PlatformNeutrality_VA.pdf (2014). Accessed 24 Oct 2016
76. Nunez, M.: Senate GOP launches inquiry into Facebook's news curation. http://gizmodo.com/senate-gop-launches-inquiry-into-facebook-s-news-curati-1775767018 (2016). Accessed 24 Oct 2016
77. Chan, C.: When one app rules them all: the case of WeChat and mobile in China. Andreessen Horowitz. http://a16z.com/2015/08/06/wechat-china-mobile-first/ (2015). Accessed 24 Oct 2016
78. ADL: Google search ranking of hate sites not intentional. http://archive.adl.org/rumors/google_search_rumors.html (2004). Accessed 24 Oct 2016
79. Woan, T.: Searching for an answer: can Google legally manipulate search engine results? Univ. Pa. J. Bus. Law. **16**(1), 294–331 (2013)
80. Wu, T.: Machine speech. Univ. Pa. Law Rev. **161**, 1495–1533 (2013)
81. Volokh, E., Falk, D.: First amendment protection for search engine search results. http://volokh.com/wp-content/uploads/2012/05/SearchEngineFirstAmendment.pdf (2012). Accessed 24 Oct 2016
82. MacKinnon, R.: Consent of the Networked. Basic Books, New York (2012)
83. Chander, A.: Facebookistan. N. C. Law Rev. **90**, 1807 (2012)
84. Pasquale, F.: Search, speech, and secrecy: corporate strategies for inverting net neutrality debates. Yale Law and Policy Review. Inter Alia. http://ylpr.yale.edu/inter_alia/search-speech-and-secrecy-corporate-strategies-inverting-net-neutrality-debates (2010). Accessed 24 Oct 2016
85. Richtel, M.: I was discovered by an algorithm. The New York Times. http://archive. indianexpress.com/news/i-was-discovered-by-an-algorithm/1111552/ (2013). Accessed 24 Oct 2016
86. Slobogin, C.: Privacy at Risk. University of Chicago Press, Chicago (2007)
87. Zarsky, T.Z.: Understanding discrimination in the scored society. Wash. Law Rev. **89**, 1375–1412 (2014)
88. Nissenbaum, H.F.: Privacy in Context. Stanford Law Books, Stanford, CA (2010)
89. Calo, M.R.: The boundaries of privacy harm. Indiana Law J. **86**(3), 1131–1162 (2011)
90. Goldman, E.: Data mining and attention consumption. In: Strandburg, K., Raicu, D. (eds.) Privacy and Technologies of Identity. Springer Science + Business Media, New York (2005)
91. Pasquale, F.: The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press, Cambridge, MA (2015)
92. Clarke, R.: Profiling: a hidden challenge to the regulation of data surveillance. J. Law Inf. Sci. **4**(2), 403 (1993)

93. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery: an overview. In: Fayyad, U. (ed.) Advances in Knowledge Discovery and Data Mining. AAAI Press, Menlo Park, CA (1996)
94. Paparrizos, J., White, R.W., Horvitz, E.: Screening for pancreatic adenocarcinoma using signals from web search logs: feasibility study and results. J. Oncol. Pract. **12**(8), 737–744 (2016)
95. Friedman, B., Nissenbaum, H.: Bias in computer systems. ACM Trans. Inf. Syst. **14**(3), 330–347 (1996). In: Friedman, B. (ed.). Human Values and the Design of Computer Technology. CSLI Publications, Stanford, CA (1997)
96. Hildebrant, M.: Profiling and the rule of law. Identity Inf. Soc. **1**(1), 55–70 (2008)
97. Shkabatur, J.: Cities @ crossroads: digital technology and local democracy in America. Brooklin Law Rev. **76**(4), 1413–1485 (2011)
98. Zarsky, T.Z.: "Mine your own business!": making the case for the implications of the data mining of personal information in the forum of public opinion. Yale J. Law Technol. **5**(1), 1–56 (2003)
99. Mayer, J: Tracking the trackers: where everybody knows your username. http://cyberlaw.stanford.edu/node/6740 (2011). Accessed 24 Oct 2016
100. Narayanan, A: There is no such thing as anonymous online tracking. http://cyberlaw.stanford.edu/node/6701 (2011). Accessed 24 Oct 2016
101. Perito, D., Castelluccia, C., Kaafar, M.A., Manilsr, P.: How unique and traceable are usernames? In: Fischer-Hübner, S., Hopper, N. (eds.) Privacy Enhancing Technologies. Springer, Berlin (2011)
102. Datalogix: Privacy policy. https://www.datalogix.com/privacy/ (2016). Accessed 24 Oct 2016
103. Solove, D.J.: Nothing to Hide. Yale University Press, New Haven, CT (2011)
104. Zarsky, T.Z.: Law and online social networks: mapping the challenges and promises of user-generated information flows. Fordham Intell. Prop. Media Entertainment Law J. **18**(3), 741–783 (2008)
105. Himma, K.E., Tavani, H.T.: The Handbook of Information and Computer Ethics. Wiley, Hoboken, NJ (2008)
106. Angwin, J.: Online tracking ramps up—popularity of user-tailored advertising fuels data gathering on browsing habits. Wall Street J. http://www.wsj.com/articles/SB10001424052702303836404577472491637833420 (2012). Accessed 24 Oct 2016
107. World Economic Forum: Rethinking personal data: strengthening trust. http://www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf (2012). Accessed 24 Oct 2016
108. Posner, R.A.: The economics of privacy. Am. Econ. Rev. **71**(2), 405–409 (1981)
109. Calzolari, G., Pavan, A.: On the optimality of privacy in sequential contracting. J. Econ. Theory. **130**(1), 168–204 (2006)
110. Acquisti, A., Varian, H.R.: Conditioning prices on purchase history. Mark. Sci. **24**(3), 367–381 (2005)
111. Schwartz, P.M.: Property, privacy, and personal data. Harv. Law Rev. **117**, 2056–2128 (2003)
112. Purtova, N.: Property rights in personal data: an European perspective. Dissertation, Uitgeverij BOXPress, Oistervijk (2011)
113. Noam, E.M.: Privacy and self-regulation: markets for electronic privacy. In: Wellbery, B.S. (ed.) Privacy and Self-Regulation in the Information Age. U.S. Dept. of Commerce, National Telecommunications and Information Administration, Washington, D.C. (1997)
114. Cohen, J.E.: Examined lives: informational privacy and the subject as object. Stanford Law Rev. **52**, 1373–1437 (1999)
115. Bergelson, V.: It's personal but is it mine? Toward property rights in personal information. U.C. Davis Law Review. **37**, 379–451 (2003)
116. Laudon, K.C.: Markets and privacy. Commun. ACM. **39**(9), 92–104 (1996)
117. Aperjis, C., Huberman, B.: A market for unbiased private data: paying individuals according to their privacy attitudes. First Monday **17**(5) (2012)

118. Kroft, S.: The data brokers: selling your personal information. 60 Minutes. http://www.cbsnews.com/news/data-brokers-selling-personal-information-60-minutes/ (2014). Accessed 24 Oct 2016
119. Jentzsch, N., Preibusch, S., Harasser, A.: Study on monetizing privacy: an economic model for pricing personal information. ENISA Publications. https://www.enisa.europa.eu/publications/monetising-privacy (2012). Accessed 24 Oct 2016
120. Kosner, A.W.: New Facebook policies sell your face and whatever it infers. Forbes. http://www.forbes.com/sites/anthonykosner/2013/08/31/new-facebook-policies-sell-your-faceand-whatever-it-infers/ (2013). Accessed 24 Oct 2016
121. Solove, D.J.: Understanding Privacy. Harvard University Press, Cambridge, MA (2008)
122. Borcea-Pfitzmann, K., Pfitzmann, A., Berg, M.: Privacy 3.0: = data minimization + user control + contextual integrity. Inf. Technol. **53**(1), 34–40 (2011)
123. Sweeney, L.: K-anonymity: a model for protecting privacy. Int. J. Uncertain Fuzziness Knowl Based Syst. **10**(5), 557–570 (2002)
124. Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M.: L-diversity: privacy beyond k-anonymity. ACM Trans. Knowl. Discov. Data **1**(1), 1–52, Art. 3 (2007)
125. Li, N., Li, T., Venkatasubramanian, S.: t-closeness: privacy beyond k-anonymity and l-diversity. In: IEEE 23rd International Conference on Data Engineering, pp. 106–115. IEEE, Istanbul (2007)
126. Karjoth, G., Schunter, M., Waidner, M.: Privacy-enabled services for enterprises. http://www.semper.org/sirene/publ/KaSW_02.IBMreport-rz3391.pdf (2002). Accessed 24 Oct 2016
127. Cranor, L.F., Guduru, P., Arjula, M.: User interfaces for privacy agents. ACM Trans. Comput. Hum. Interact. **13**(2), 135–178 (2006)
128. Gritzalis, S.: Enhancing web privacy and anonymity in the digital era. Inf. Manag. Comput. Secur. **12**(3), 255–288 (2004)
129. Andrews, L.: I Know Who You Are and I Saw What You Did: Social Networks and The Death of Privacy. Free Press, New York (2012)
130. Irani, D., Webb, S., Li, K., Pu, C.: Large online social footprints—an emerging threat. http://cobweb.cs.uga.edu/~kangli/src/SecureCom09.pdf (2009). Accessed 24 Oct 2016
131. Irani, D., Webb, S., Pu, C., Li, K.: Modeling unintended personal-information leakage from multiple online social networks. IEEE Internet Comput. **15**(3), 13–19 (2011)
132. Spiekermann, S., Dickinson, I., Günther, O., Reynolds, D.: User agents in e-commerce environments: industry vs. consumer perspectives on data exchange. In: Eder, J., Missikoff, M. (eds.) Advanced Information Systems Engineering. Springer, Berlin (2003)
133. Bott, E.: The do not track standard has crossed into crazy territory. http://www.zdnet.com/the-do-not-track-standard-has-crossed-into-crazy-territory-7000005502/ (2012). Accessed 24 Oct 2016
134. Fujitsu Res. Inst.: Personal data in the cloud: a global survey of consumer attitudes. http://www.fujitsu.com/downloads/SOL/fai/reports/fujitsu_personal-data-in-the-cloud.pdf (2010). Accessed 24 Oct 2016
135. Brunton, F., Nissenbaum, H.: Vernacular resistance to data collection and analysis: a political theory of obfuscation. First Monday. **16**(5), 1–16 (2011)
136. Danezis, G.: Privacy technology options for smart metering. http://research.microsoft.com/enus/projects/privacy_in_metering/privacytechnologyoptionsforsmartmetering.pdf (2011). Accessed 24 Oct 2016
137. Bengtsson, L., Lu, X., Thorson, A., Garfield, R., von Schreeb, J.: Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. PLoS Med. **8**(8), e1001083 (2011)
138. Wesolowski, A., Eagle, N., Tatem, A.J., Smith, D.L., Noor, A.M., Snow, R.W., Buckee, C.O.: Quantifying the impact of human mobility on malaria. Science. **338**(6104), 267–270 (2012)

139. Wesolowski, A., Buckee, C., Bengtsson, L., Wetter, E., Lu, X., Tatem, A.J.: Commentary: containing the ebola outbreak—the potential and challenge of mobile network data. http://currents.plos.org/outbreaks/article/containing-the-ebola-outbreak-the-potential-and-challenge-of-mobile-network-data/ (2014). Accessed 24 Oct 2016
140. Phelps, J., Nowak, G., Ferrell, E.: Privacy concerns and consumer willingness to provide personal information. J. Public Policy Mark. **19**(1), 27–41 (2000)
141. Wood, W., Neal, D.T.: The habitual consumer. J. Consum. Psychol. **19**(4), 579–592 (2009)
142. Buttle, F., Burton, J.: Does service failure influence customer loyalty? J. Consum. Behav. **1**(3), 217–227 (2012)
143. Pew Research Centre: Mobile health 2012. http://www.pewinternet.org/2012/11/08/mobile-health-2012 (2012). Accessed 24 Oct 2016
144. Reinfelder, L., Benenson, Z., Gassmann, F.: Android and iOS users' differences concerning security and privacy. In: Mackay, W. (ed.) CHI '13 Extended Abstracts on Human Factors in Computing Systems. ACM, New York, NY (2013)
145. Elkin-Koren, N., Weinstock Netanel, N. (eds.): The Commodification of Information. Kluwer Law International, The Hague (2002)
146. FTC Staff Report: Mobile apps for kids: current privacy disclosures are disappointing. http://www.ftc.gov/os/2012/02/120216mobile_apps_kids.pdf (2012). Accessed 24 Oct 2016
147. FTC Staff Report: Mobile apps for kids: disclosures still not making the grade. http://www.ftc.gov/os/2012/12/121210mobilekidsappreport.pdf (2012). Accessed 24 Oct 2016
148. FTC Staff Report: Mobile privacy disclosures: building trust through transparency. http://www.ftc.gov/os/2013/02/130201mobileprivacyreport.pdf (2013). Accessed 24 Oct 2016
149. Canadian Offices of the Privacy Commissioners: Seizing opportunity: good privacy practices for developing mobile apps. http://www.priv.gc.ca/information/pub/gd_app_201210_e.pdf (2012). Accessed 24 Oct 2016
150. Harris, K.D.: Privacy on the go: recommendations for the mobile ecosystem. http://oag.ca.gov/sites/all/files/pdfs/privacy/privacy_on_the_go.pdf (2013). Accessed 24 Oct 2016
151. GSMA: User perspectives on mobile privacy. http://www.gsma.com/publicpolicy/wpcontent/uploads/2012/03/futuresightuserperspectivesonuserprivacy.pdf (2011). Accessed 24 Oct 2016
152. Sundsøy, P., Bjelland, J., Iqbal, A.M., Pentland, A.S., De Montjoye, Y.A.: Big data-driven marketing: how machine learning outperforms marketers' gut-feeling. In: Greenberg, A.M., Kennedy, W.G., Bos, N. (eds.) Social Computing, Behavioral-Cultural Modeling and Prediction. Springer, Berlin (2013)
153. Pasquale, F.: Reforming the law of reputation. Loyola Univ. Chicago Law J. **47**, 515–539 (2015)
154. Ombelet, P.J., Kuczerawy, A., Valcke, P.: Supervising automated journalists in the newsroom: liability for algorithmically produced news stories. Revue du Droit des Technologies de l'Information. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2768646 (forthcoming 2016). Accessed 24 Oct 2016
155. Latar, N.L., Norsfors, D.: Digital identities and journalism content—how artificial intelligence and journalism may co-develop and why society should care. Innov. Journalism. **6**(7), 1–47 (2006)
156. Ombelet, P.J., Morozov, E.: A robot stole my Pulitzer! How automated journalism and loss of reading privacy may hurt civil discourse. http://www. slate.com/articles/technology/future_tense/2012/03/narrative_science_robot_journalists_customized_news_and_the_danger_to_civil_ discourse_.single.html (2012). Accessed 24 Oct 2016
157. Hacker, P., Petkova, B.: Reining in the big promise of big data: transparency, inequality, and new regulatory frontiers. Northwest. J. Technol. Intellect. Prop. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2773527 (forthcoming 2016). Accessed 24 Oct 2016
158. Hajian, S., Domingo-Ferrer, J.: Direct and indirect discrimination prevention methods. In: Custers, B., Calders, T., Schermer, B., Zarsky, T. (eds.) Discrimination and Privacy in the Information Society. Springer, New York (2013)

159. Mayer-Schonberger, V., Cukier, K.: Big Data. A Revolution That Will Transform How We Live, Work, And Think. Eamon Dolan/Houghton Mifflin Harcourt, Boston, MA (2014)
160. Calders, T., Verwer, S.: Three naïve Bayes approaches for discrimination-free classification. Data Min. Knowl. Disc. **21**(2), 277–292 (2010)
161. Kamiran, F., Calders, T., Pechenizkiy, M.: Techniques for discrimination-free predictive models. In: Custers, B., Calders, T., Schermer, B., Zarsky, T. (eds.) Discrimination and Privacy in the Information Society. Springer, New York (2013)
162. Tutt, A.: An FDA for algorithms. Adm. Law Rev. **67**, 1–26 (2016)
163. FTC: Spring privacy series: alternative scoring products. http://www.ftc.gov/news-events/events-calendar/2014/03/spring-privacy-series-alternative-scoring-products (2014). Accessed 24 Oct 2016
164. Ramirez, E.: The privacy challenges of big data: a view from the lifeguard's chair. https://www.ftc.gov/public-statements/2013/08/privacy-challenges-big-data-view-lifeguard%E2%80%99s-chair (2013). Accessed 24 Oct 2016
165. Sandvig, C., Hamilton, K., Karahalios, K., Langbort, C.: Auditing algorithms: research methods for detecting discrimination on internet platforms. Data and discrimination: converting critical concerns into productive inquiry. http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf (2014). Accessed 24 Oct 2016
166. Benkler, Y.: The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press, New Haven, CT (2006)
167. Citron, D.K.: Open code governance. Univ. Chicago Legal Forum. **2008**(1), 355–387 (2008)
168. Barnett, J.M.: The host's dilemma: strategic forfeiture in platform markets for informational goods. Harv. Law Rev. **124**(8), 1861–1938 (2011)
169. Moses, L.: Marketers should take note of when women feel least attractive: what messages to convey and when to send them. ADWEEK. http://www.adweek.com/news/advertising-branding/marketers-should-take-note-when-women-feel-least-attractive-152753 (2013). Accessed 24 Oct 2016
170. Orentlicher, D.: Prescription data mining and the protection of patients' interests. J. Law Med. Ethics. **38**(1), 74–84 (2010)
171. WPF: Data broker testimony results in new congressional letters to data brokers about vulnerability-based marketing. http://www.worldprivacyforum.org/2014/02/wpfs-data-broker-testimony-results-in-new-congressional-letters-to-data-brokers-regarding-vulnerability-based-marketing/ (2014). Accessed 24 Oct 2016
172. Bakos, Y., Marotta-Wurgler, F., Trossen, D.R.: Does anyone read the fine print? Consumer attention to standard-form contracts. J. Leg. Stud. **43**(1), 1–35 (2014)
173. MacDonald, A.M., Cranor, L.F.: The cost of reading privacy policies. J. Law Policy Inf. Soc. **4**(3), 540–565 (2008)
174. Lipford, H.R, Watson, J., Whitney, M., Froiland, K., Reeder, R.W.: Visual vs compact: a comparison of privacy policy interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1111–1114 (2010)
175. Passera, S., Haapio, H.: Transforming contracts from legal rules to user-centered communication tools: a human-information interaction challenge. Commun. Des. Q. Rev. **1**(3), 38–45 (2013)
176. Phillips, E.D.: The Software License Unveiled. Oxford University Press, Oxford (2009)
177. Gardner, T.: To read, or not to read... the terms and conditions. The Daily Mail. http://www.dailymail.co.uk/news/article-2118688/PayPalagreement-longer-Hamlet-iTunes-beats-Macbeth.html (2012). Accessed 24 Oct 2016
178. Ayres, I., Schwartz, A.: The no-reading problem in consumer contract law. Stanford Law Rev. **66**, 545 (2014)
179. Bar-Gill, O., Ben-Shahar, O.: Regulatory techniques in consumer protection: a critique of European consumer contract law. Common Mark. Law Rev. **50**, 109–126 (2013)

180. Luzak, J.: Passive consumers vs. the new online disclosure rules of the consumer rights. Directive. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2553877 (2014). Accessed 24 Oct 2016
181. Luzak, J.: To withdraw or not to withdraw? Evaluation of the mandatory right of withdrawal in consumer distance selling contracts taking into account its behavioral effects on consumers. J. Consum. Policy. **37**(1), 91–111 (2014)
182. Purnhagen, K., Van Herpen, E.: Can bonus packs mislead consumers? An empirical assessment of the ECJ's mars judgment and its potential impact on EU marketing regulation. In: Wageningen Working Papers Series in Law and Governance 2014/07, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2503342 (2014)
183. MacDonald, A.M., Reeder, R.W., Kelley, P.G., Cranor, L.F.: A comparative study of online privacy policies and formats. In: Goldberg, I., Atallah, M.J. (eds.) Privacy Enhancing Technologies. Springer, Berlin (2009)
184. Stigler, G.: The Economics of information. J. Polit. Econ. **69**(3), 213–225 (1961)
185. Akerlof, G.A.: The Market for "lemons": quality uncertainty and the market mechanisms. Q. J. Econ. **84**(3), 488 (1970)
186. Macho-Stadler, I., Pérez-Castrillo, J.D.: An Introduction to the Economics of Information. Oxford University Press, Oxford (2001)
187. Evans, M.B., McBride, A.A., Queen, M., Thayer, A., Spyridakis, J.H.: The effect of style and typography on perceptions of document tone. http://faculty.washington.edu/jansp/Publications/Document_Tone_IEEE_Proceedings_2004.pdf (2004). Accessed 24 Oct 2016
188. Masson, M.E.J., Waldron, M.A.: Comprehension of legal contracts by non-experts: effectiveness of plain language redrafting. Appl. Cogn. Psychol. **8**, 67–85 (1994)
189. Ben-Shahar, O., Schneider, C.E.: More Than You Wanted to Know: The Failure of Mandated Disclosure. Princeton University Press, Princeton (2014)
190. Radin, M.: Boilerplate. Princeton University Press, Princeton, NJ (2013)
191. Ben-Shahar, O., Chilton, A.S.: "Best practices" in the design of privacy disclosures: an experimental test. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2670115 (2015). Accessed 24 Oct 2016
192. Miller, A.A.: What do we worry about when we worry about price discrimination? The law and ethics of using personal information for pricing. J. Technol. Law Policy. **19**, 41–104 (2014)
193. Mittlestadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L.: The ethics of algorithms: mapping the debate. Big Data Soc. 1–21 (2016)