BRUSSELS PRIVACY HUB
WORKING PAPER
VOL. 6 • N° 18 • JANUARY 2020

BRUSSELS PRIVACY HUB
Vrije Universiteit Brussel

CHALLENGING ALGORITHMIC PROFILING: THE LIMITS OF DATA PROTECTION AND ANTI-DISCRIMINATION IN RESPONDING TO EMERGENT DISCRIMINATION[1]

Dr Monique Mann[2] and Professor Tobias Matzner

The potential for biases being built into algorithms has been known for some time (e.g., Friedman and Nissenbaum, 1996), yet literature has only recently demonstrated the ways algorithmic profiling can result in social sorting and harm marginalized groups (e.g., Browne 2015; Eubanks 2018; Noble 2018). We contend that with increased algorithmic complexity, biases will become more sophisticated and difficult to identify, control for, or contest. Our argument has four steps: first, we show how harnessing algorithms means that data gathered at a particular place and time relating to specific persons, can be used to build group models applied in different contexts to different persons. Thus, privacy and data protection rights, with their focus on individuals (Coll, 2014; Parsons, 2015), do not protect from the discriminatory potential of algorithmic profiling. Second, we explore the idea that anti-discrimination regulation may be more promising, but acknowledge limitations. Third, we argue that in order to harness anti-discrimination regulation, it needs to confront emergent forms of discrimination or risk creating new invisibilities, including invisibility from existing safeguards. Finally, we outline suggestions to address emergent forms of discrimination and exclusionary invisibilities via intersectional and post-colonial analysis.

Keywords: Algorithms, profiling, GDPR, data protection, discrimination, intersectionality

Contents

The Brussels Privacy Hub Working Papers are intended to circulate research in progress for comment and discussion. Available at www.brusselsprivacyhub.org/publications.html
ISSN N° 2565-9979. This version is for academic use only.

Disclaimer

Algorithmic profiling

The data revolution has been driven by rapid innovation in "ubiquitous computing" which some claim has resulted in widespread "datafication" of the "surveillance society" or "information civilization" (Matzner, 2014; Dencik, et al., 2016; Lyon, 2001; Zuboff, 2015). Central to this is the exponential increase in data, expanded surveillance to gather more and more of it, and dynamic new ways to analyse it (Kitchin, 2014). Algorithmic profiling is a way of detecting patterns, and making predictions on the basis of them. This occurs in a range of contexts including insurance, finance, differential pricing, education, employment, marketing, governance, security, and policing (Stalder, 2002; O'Neil, 2016; Ferguson, 2017). More specifically, we understand algorithmic profiling[3] as a method of inferential analysis that identifies correlations or patterns within datasets, that can be used as an indicator to classify a subject as a member of a group (Hildebrandt, 2008; Schreurs et al., 2008).[4] These categories are formed from "probabilistic assumptions" (Leese, 2014: 502) that are de-individualised (Schermer, 2013). A decision for a loan application may not be made on the basis of individual risk of default, but on the basis of postcode or neighbourhood, that may operate as an indirect proxy of other indicators such as the socio-economic or racial composition of one's neighbours. This leads to concerns about social sorting and discrimination.

Social sorting and discriminatory potential

Algorithmic profiling may result in social sorting and other discriminatory outcomes (see e.g., Lyon, 2003; 2014; Parsons, 2015). Research in Australia (Mann and Daly, 2019) and North America (Browne 2015; Eubanks 2018; Peña Gangadharan 2012; Noble 2018; Sandvig et al 2016), demonstrates how algorithmic profiling targets marginalized groups, such as racial minorities, individuals of low socio-economic status, and women. Browne (2015) argues that algorithmic profiling perpetuates hierarchies predicated on the enmeshing of identity characteristics. Discriminatory practices become self-enforcing with feedback loops as datasets are constructed that disproportionately contain data about certain people, leading to over-monitoring and over-policing of those groups (see e.g., Ferguson, 2017). Importantly, discriminatory effects also occur if data on discriminatory features like gender, race, ethnicity etc. are not directly processed.[5] In fact, algorithmic profiling can easily identify "proxies", i.e., combinations of input data which are accurate predictors for the discriminatory categories (Harcourt, 2010; Kleinberg et al., 2016). This illustrates that the data which are used, as well as implicit assumptions while programming (and training in the case of machine learning algorithms), carry discriminatory potential (Campolo et al., 2017).

1 We acknowledge the excellent research assistance provided by Ms. Harley Williams, and QUT for financing Harley's contribution via the Vice-Chancellor's Research Fellowship awarded to Monique Mann. We would like to thank Associate Professor Peta Mitchell, Dr Ian Warren, Dr Angela Daly and the three anonymous reviewers for their helpful comments on previous versions of this article.
2 Corresponding author: monique.mann@deakin.edu.au
3 Profiling can be inductive or deductive, or a combination of both. Inductive profiling relates to the generation of profiles (as testable hypotheses), while deductive profiling is concerned with testing profiles on datasets to confirm hypotheses.
4 According to Article 4 of the General Data Protection Regulation (GDPR) 'profiling' means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements.
5 Direct processing is the processing of information for the specific purposes for which data were collected, according to the variables or categories of data that were collected. Indirect processing is processing for reasons other than for which data were collected, and can involve the identification of indirect or proxy relationships between variables or data categories collected, and other categories or variables for which data were not directly collected.

Data protection and the 'right' not to be subject to automated decisions

Currently, many challenges to algorithmic profiling refer to data protection law, especially the new General Data Protection Regulation (GDPR) of the European Union (EU), since it contains dedicated rules for algorithmic profiling that have resonated in academic critique internationally (see e.g. Vedder and Naudts, 2017; Selbst and Powles, 2017; Edwards and Veale, 2017). We focus specifically on the European Union (EU) due to the recent introduction of the GDPR[6], which replaced the Data Protection Directive (DPD)[7], and seeks to regulate the use of personal data,[8] and includes a 'right' not to be subject to automated decisions (Article 22 of the GDPR). The GDPR is designed to uphold data protection rights under Article 8 (Protection of Personal Data)[9] of the EU Charter of Fundamental Rights. Although this new regulatory regime is generally regarded as global 'gold standard' (see e.g., Buttarelli, 2016; Safari, 2017), there are limits to the application of data protection law in countering algorithmic profiling and the drawing of sensitive or discriminatory inferences (see also Wachter and Mittelstadt, 2019). Therefore, we argue that data protection law may not be a good resource to challenge the problems of algorithmic profiling introduced above. Instead, we show that anti-discrimination may offer a more promising outlook, however, existing protections should be amended or extended in order to cope with new forms of discrimination that emerge, or that do not pertain to known protected identities, but rather represent patterns that have little or no intuitive meaning to human practice.

There is a lack of consensus as to whether algorithmic profiles, or algorithmic inferences made about an individual, are considered as personal data. This is because according to Article 4 of the GDPR,[10] information must relate to an identified or identifiable natural person to be considered as personal data.[11] Koops (2014) argues that with ongoing technological innovations, what counts as personal data is becoming obscured, a point also made by Purtova (2018), who argues the distinction between personal and non-personal data should be suspended.[12] Purtova (2018) concludes all data processing with the potential to impact people should trig-

6 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).

7 Directive 95/46/EC of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data.

8 While the primary objective of the GDPR is to regulate personal data, it also seeks to facilitate the relationship between, and uphold other, fundamental rights. For example, Recital 4 of the GDPR states that the processing of personal data should be designed to serve mankind [***sic***]. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality. This Regulation respects all fundamental rights and observes the freedoms and principles recognized in the Charter as enshrined in the Treaties, in particular the respect for private and family life, home and communications, the protection of personal data, freedom of thought, conscience and religion, freedom of expression and information, freedom to conduct a business, the right to an effective remedy and to a fair trial, and cultural, religious and linguistic diversity.

9 Article 8 of the Charter of Fundamental Rights of the European Union affirms that (1) Everyone has the right to the protection of personal data concerning him or her. (2) Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified and (3) Compliance with these rules shall be subject to control by an independent authority.

10 According to Article 4 of the GDPR 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

11 See further Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136.'

12 This is because developments in the Internet of Things (IoT), hypoconnectivity and automated data processing are likely to render "absolute and irreversible anonymity" impossible, thus meaning binary differentiation between identifiable and non-identifiable data will become irrelevant (Purtova, 2018).

ger protection, and at the very least, should be assessed for the likely impact it will have upon them. Purtova's argument does not entail that data protection regulations suffice to deal with all kinds of data, but rather that the blurring of the distinction between personal data and other data underlines the need for new normative grounds to assess the impact of data processing. Wachter and Mittelstadt (2019) note the guidance provided by the Article 29 Working Party[13] provides support for inferences being considered as personal data, particularly if there is potential to impact an identifiable individual's rights and interests. Yet they also point to the conflicting decisions of the European Court of Justice that have a more constrained interpretation of personal data. The consequence is that those impacted by algorithmic profiling may enjoy limited data protection rights, such as access rights, that in turn may impede their ability to correct or rectify inaccurate inferences, or assess the lawfulness of data processing (Wachter & Mittelstadt, 2019).[14] Complicating matters further, anonymized data[15] can be used as a basis to construct profiles and draw sensitive inferences. Therefore, "by using data about people not linked to a particular individual, or by purposefully anonymising data prior to drawing inferences and constructing profiles, companies can thus avoid many of the restrictions of data protection law" (Wachter & Mittelstadt, 2019: 55).

A significant aspect of the GDPR is that it grants a 'right' "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her" (Article 22). This has been subject to debates. The Article 29 Working Party Guidelines on 'Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' provide further guidance on the specific provisions that establish the general prohibition for decision-making based solely on automated processing. They argue that "interpreting Article 22 as a prohibition rather than a right to be invoked means that individuals are automatically protected from the potential effects this type of processing may have." However, it is also argued that the scope of Article 22(1) is confined to decisions based ***solely*** on automated processing, and does not capture decisions that are ***not solely*** based on automated processing.[16] Further, the decision must ***have legal or similarity significant effects.*** There are also a number of exclusionary conditions for which Article 22 does not apply.[17]

Given the above, there have been calls for entirely new rights to be recognised under data protection law. Wachter and Mittelstadt (2019) argue for a 'right to reasonable inferences' to be incorporated into the GDPR. This would, in principle, require the data controller to establish whether an inference is reasonable. A key limitation of this proposal is that it does not specifically relate to managing differential treatment or discriminatory outcomes on the basis of sensitive inferences. Therefore, data protection, and suggested improvements such as a 'right to reasonable inferences', may not be an ideal framework for responding to the challenges

13 It should be noted that the guidance and opinions of the Article 29 Working Party are not legally binding.

14 There are also challenges for transparency and notification rights under the GDPR (Articles 13-14) as these rights concern data collected from the data subject or a third party, and not the creation of profiles, or inferences made by the data controller.

15 Anonymous data is any information that relates to an individual where the individual cannot be identified, see: Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' and Article 29 Working Party 'Opinion 5/2014 on Anonymization Techniques 0829/14/EN/WP216.'

16 For further guidance on provisions that cover profiling and automated individual decisions-making, including decisions-making processes that are solely and not solely automated see Article 29 Guidelines on 'Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679.'

17 The exclusionary conditions for Article 22 of the GDPR are set out in Article 22(1). These include necessity for entering into a contract between data subject and controller, authorisation by Union and Member state law to which the controller is subject to safeguards, and is based on explicit consent of the data subject. See also, Wachter et al., 2017 and Article 29 Guidelines on 'Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679.'

presented by algorithmic profiling. Questions about the suitability of data protection law intersect with the recent developments in the critique of algorithms that we discussed in the introduction: that is, the problems with algorithmic profiling are not limited to processing personal information or drawing sensitive inferences. Rather, the impact of algorithmic profiling on the actions, lives and personalities of profiled persons might derive from input that seems inconspicuous from the point of view of data protection, due to the fact that algorithmic profiling works on classes, aggregates and patterns. This of course falls within a long-discussed limit of privacy in general, and not just data protection, if understood in an individualising manner. Gilliom (2001: 122) argues:

"To the extent […] that the privacy paradigm relies on and maintains the idea of the autonomous individual and the idea of surveillance as mere visitation, it risks a massive misrepresentation of the full impact of surveillance in our lives. The positioning of extensive and ongoing surveillance in the modern state promises to recast the citizen into the frames and terms of bureaucratic analysis and translate our ongoing actions into tactics of compliance, evasion, and above all, calculation."

Anti-discrimination rules provide for a shift of focus away from privacy and data protection. As we argue below, anti-discrimination is a more promising candidate as the fundamental aim of algorithmic profiling is to ***discriminate*** (Hildebrandt, 2008; Gandy, 2010).

Anti-discrimination as an alternative

Given the potential for algorithmic profiling to facilitate discrimination, anti-discrimination law may provide a more promising avenue for responding. Gellert and colleagues (2013: 61) draw attention to the important distinction that data protection is concerned with certain actions (principally data '***processing***'), whereas anti-discrimination relates to an outcome irrespective of the action/s that led to it. Yet, anti-discrimination protections in contemporary (and future) data processing contexts may also have limited application due to the difficulties in identifying differential treatment on the basis of protected grounds, especially when they are abstracted, or intersectional, or ***emergent*** – a critique we extend below by diagnosing problems of applying anti-discrimination safeguards, and proposing ways of amending them in order to make them suitable to address the complicated forms of discrimination that arise in algorithmic profiling.

In the EU, both data protection[18] and non-discrimination[19] are fundamental rights enshrined in the Charter of Fundamental Rights of the EU as "regulatory human rights" (Gellert et al., 2013: 61). Article 14 (Protection from Discrimination) of the European Convention on Human Rights prohibits discrimination on the basis of demographic characteristics, including any possible

18 Article 8 (Protection of Personal Data) of the Charter of Fundamental Rights of the EU, specifically concerns the right to data protection (and it is upon this right which the GDPR is founded).

19 In the US there is a longer history of anti-discrimination law that emerged from the civil rights movement of the 1960s, although is at present, subject to restrictions, ideological challenge, and legal and political "backlash" (De Burca, 2012: 3). While in Europe anti-discrimination law is said to be developing and expanding. Despite different histories and cultural understandings of equality, individual freedom, and social welfare, there are similarities in the development and trajectory of anti-discrimination approaches between the EU and the US, with the EU approach being modeled closely on UK anti-discrimination laws, which were influenced by US law (De Burca, 2012). However, there are distinctions in language between these major jurisdictions that should be explained: Direct discrimination in the US is known as ***disparate treatment discrimination***, which is discrimination based on differential treatment according to a protected ground. This type of discrimination differs from the US model of ***disparate impact treatment*** which corresponds to the EU model of ***indirect discrimination***, where an "apparently neutral rule or practice has a discriminatory effect on a prohibited group" (Marcat-Bruns, 2018: 44).

'other status'.[20] Further, Title III of the Charter of Fundamental Rights of the EU is dedicated specifically to equality, and composed of a general provision on anti-discrimination and equality (Article 21), and provisions for specific demographics such as cultural, religious and linguistic diversity, gender, rights of child, elderly, and persons with disabilities (Articles 22-26).[21] Gellert et al. (2013: 65) argue the specific types of discrimination (i.e., Articles 22-26) represent "a more conceptually refined notion of discrimination" in comparison to the general principle of equality. Yet, we contend that with respect to the abstracted nature of profiling and the drawing of inferences, it may not be possible to identify grounds for discrimination as per specific protected grounds, and that a broader and more diversified approach to anti-discrimination may be an avenue to explore. This is a point we return to, but first a brief comment on direct and indirect discrimination is required.

Direct and indirect discrimination

Direct discrimination focuses on situations whereby an individual has been treated unfairly on the basis of protected grounds, whereas indirect discrimination refers to practices that may ***inadvertently or indirectly*** discriminate (Gellert et al., 2013). Indirect discrimination is difficult to detect, indeed, it has been argued that "most of the time, persons who have been victimized [...] will not know precisely if, when, or how they have been discriminated against" (Gandy, 2010: 40). In the context of algorithmic profiling, especially where machine learning is used to create new inferential categories, this problem exceeds even the issue of indirect discrimination. Leese (2014: 504) argues that as "data-driven profiles produce artificial and non-representational categories rather than actual real-life social groups, the individual is unlikely to notice when he or she becomes part of a 'risky' category."

Thus, algorithmic profiling complicates the notion that a discriminatory outcome can be linked to a protected identity in a two-fold manner: first by enabling proxies, and second, by using new categories that have no clear meaning to human interpretation. This is significant in the context of arguments made by Leese (2014: 505) in relation to "deep-seated epistemological conflict between an anti-discrimination framework that conceives of knowledge as the establishment of causality and data-driven analytics that build fluid hypotheses on the basis of correlation patterns in dynamic databases." This means that "discrimination will not concern any of the protected grounds, but rather attributes such as income, postal code, browsing behaviour, type of car, etc., or complex algorithmic combinations of several attributes" (Gellert et al., 2013: 80). Therefore, it is necessary to identify "whether attributes, and complex algorithmic combinations of attributes, which do ***not*** belong to any of the specifically protected grounds" may create discrimination (Gellert et al., 2013: 81, emphasis in original). Leese (2014: 500)

20 Article 14 of the European Convention on Human Rights prohibits discrimination through the following: "The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, color, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status."

21 Article 9 of the GDPR contains restrictions on the processing of special categories of personal data (sensitive data) defined as: "processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited." It should be noted that special categories of personal data do not include age, financial situation or personal profiles which often form the basis of discrimination, however non-sensitive data can become sensitive if it is used to make inferences about sensitive data (as per Wachter and Mittelstadt, 2019). Significantly neither sex nor gender are considered sensitive categories of information although discrimination frequently occurs along these lines (see Article 4 and Article 9 of the GDPR; Bano, 2018).

argues that "data-driven forms of profiling produce a distinct form of knowledge that appears dynamic and implicit, and thus continually escapes the scope of the regulatory legal regime." We term these forms of indirect discrimination based on complex combinations and correlations as ***emergent discrimination***; the focus of the final section of this article.

Reconceptualising discrimination

On a very fundamental level, every form of algorithmic judgment could be treated as discriminatory. Profiling aims at making predictions, that is, uses statistical methods or machine learning to predict pieces of information which are not directly available – otherwise profiling would not be necessary. Very generally, it looks for differences among people that entail that they are treated differently. Anti-discrimination asks us to treat people equally despite their differences.

There are challenges for individuals to even identify that they have been subject to differential treatment on the basis of a protected ground. This is because, for example, an individual denied a mortgage on the basis of their neighbourhood is "not a member of a protected group. She is a victim, not because of her race, but because of the race of the people that live in, and help determine the profile of her neighborhood" (Danna and Gandy, 2002: 382). Moreover, when the aim of the algorithmic system is to identify and manage risks, some outcomes may never arise: "we must also consider that fact that as system objectives more routinely come to be framed in terms of the identification, minimization, or management of risks, rather than the achievement of objectively measured goals or achievements, the consequences of systematic error will be more difficult to observe and control" (Gandy, 2010: 39). Our main contribution is the development of new concepts that capture a more precise notion of discrimination and that enables emergent classifications to be recognised as discriminatory. To that aim, there is a need to connect with intersectional perspectives, that also do not map so easily onto existing protected identities.

Intersectional discrimination

Intersectional forms of discrimination have been shown as important expansions to existing views on discrimination on the basis of one protected ground (Crenshaw, 1989). Intersectional theory argues that the specific combination of identities, e.g. a woman of color, cannot be understood in terms of the discrimination that people of color, or women, experience. That is, intersectionality highlights the entanglement of protected identities. In one of the first texts that defines intersectional analysis, Crenshaw argues that the combined effects of discrimination (in her case, on the basis of both sex and race) are particular forms of discrimination that are experienced by people with a specific combination of (protected) identities – and cannot be reduced to one of its "elements" (Crenshaw 1989: 149). Thus, intersectional theory highlights the specificity of discrimination: it may be more specific than one protected ground.

With the promise of personalization through algorithms that tie in many more features than human judgement does, the results become much more specific than just the intersection of two or three prominent markers. The same is true of emergent discrimination: it might be much more specific than the intersection of two or more identities. Crenshaw illustrated that such forms of discrimination are hard to prove statistically, even if they are "just" concerning

race and gender, as the necessary data might not be available or the statistical populations too small (Crenshaw 1989: 146). This sensitivity for experiences of discrimination that are hard to prove statistically, or to be objectified in another manner, are an important insight from intersectional thought, that can be carried over to the analysis of emergent discrimination. The fact that a large proportion of the populace with a protected feature are processed in a 'fair' manner is no guarantee that discrimination does not take place.

Significantly, algorithmic profiling that facilitates the inclusion of different sources and types of data is likely to contribute to increasing entanglements of protected identities, thus creating new categories and groups of people that experience forms of intersectional discrimination. Intersectional theory has shown that safeguards against discrimination wrongly assume that all forms of discrimination function similarly or independently. Although intersectional theory was conceived against the backdrop of US anti-discrimination law, a similar disregard of the specifics of particular social positions and intersectional identities has been diagnosed for the EU (see e.g. Verloo, 2006). Algorithmic systems, for which everything is yet another potential input feature or proxy variable for correlative analysis, might increase that assumption. These aspects highlight the problems of new forms of discrimination that emerge from complex intersectional combinations of protected grounds, or correlative abstractions from them.

Emergent discrimination

In addition to these more complex intersectional forms of discrimination, completely new forms of discrimination may emerge. Leese (2014: 504) calls these "non-representational" to express that these new classes of discriminated people might be formed by combinations of input features that do not even make sense as the intersectional combination of identities. That is, both the input and output of algorithmic systems may not have a direct relation to a protected ground but it might still be the case that an algorithmic system systematically disadvantages persons with, say, a specific combination of browsing history, make of the computer, and favourite bands (for example, Facebook likes). However, it is not clear that this would count as discrimination. A first rebuttal could be that such outcomes are not discriminatory at all. Given that the system works well, it tracks existing statistical differences – and if one does not want to call that discriminatory as explained above, that is just how the system works. After all, if it could not find any differences, the system would not work. Following this line of thought, if someone is incorrectly profiled by such a system, and as a consequence suffers from some form of disadvantage, that would be an individual error, not discrimination. However, discrimination is not an issue of wrong classifications. In fact, that a system used for algorithmic profiling should be as error-free as possible, is a matter of course. Anti-discrimination safeguards carry a stronger intuition than protection against erroneous treatment: even if there are differences in the world, we might better not differentiate along them. Thus, even if the algorithm in this case was not 'wrong' in a narrowly conceived epistemic understanding, applying the principles of anti-discrimination might mean refraining from using this information to make discriminatory decisions.

In consequence, the problem is how to single out results that count as discriminatory and should be avoided. One approach would be the perspective of 'data justice' by Dencik et al. who advocate for a critique that scrutinises "***interests*** and ***power relations*** at play in 'datafied'

societies that enfranchise some and disenfranchise others, highlighting also forms of exclusion and discrimination" (Dencik et al., 2016: 9, emphasis in original). This shares affinity with arguments made by Gilliom (2001: 136) who argues that new forms of surveillance, as new technologies of power, can never be "removed from the ongoing dynamics of political struggle." Perhaps then what is actually required is further attention to the landscapes of power and social ***(in)***justice in this new mode of algorithmic governance (Coll, 2014; Leese, 2014; Amoore, 2011). In these landscapes of power, the logic of anti-discrimination corresponds to our insight that algorithms contribute to the creation of social inequalities. Such changes and modulations can also be observed regarding emergent categorizations, for example in the application of profiling for security measures, personal features and relations that are available as input data (preferences, friendships, family ties) are invested with forms of threat and suspicion (Matzner, 2016). One could imagine that a group of people are regularly singled out by algorithmic profiling for additional screening during travel. That group of people would then be discriminated not just because they need to spend additional time and resources, but because their identity as traveller is invested with suspicion. Therefore, new forms of discrimination that emerge with algorithmic profiling can still be addressed in the spirit of anti-discrimination principles and protections. But they also ask us to divert our gaze from the algorithmic process and towards society and to ask: what does it mean that the categories created by the algorithms exist? This requires a way in which the social and political situation of the people that belong to groups that are ***created*** by algorithms can be assessed. Here, intersectional and post-colonial theories can provide valuable insights for dealing with such emergent forms of discrimination.

Making exclusionary invisibilities visible

There are differences between the newly emerging forms of discrimination and intersectional forms, as intersectional theory focuses on identities that are already recognized as a source of discrimination. Emergent forms of algorithmic discrimination stem from features and indirect proxies that themselves, on face value, seem harmless. However, it is a combination of such seemingly harmless features that might lead to emergent forms of discrimination. In this regard, however, parallels become visible. Intersectionality has put the focus on people with complex identities that suffer discrimination that is not visible from the perspective of simple singular protected grounds. This is repeated structurally with emergent forms of discrimination in a more complex way. Thus, remedies for intersectional analysis can point towards possible approaches that bring greater attention to emergent forms of discrimination.

There is a need for new strategies or methods for showing discrimination that do not rely on direct comparisons, as it may be so specific – or personalized - that any comparisons become meaningless (see e.g. Marcat-Bruns, 2018). In relation to intersectional discrimination in the EU, Marcat-Bruns (2018: 49) argues that "more efficient institutional monitoring" is required, and we agree that this is the case in relation to emergent forms of discrimination also. Fredman (2016: 8) argues that intersecting relationships of power can be analyzed and counteracted by four dimensions: "(i) the need to redress disadvantage, (ii) the need to address stigma, stereotyping, prejudice and violence, (iii) the need to facilitate voice and participation; and (iv) the need to accommodate difference and change structures of discrimination." We argue that these arguments for improvements based on intersectional theories of equality can inspire countermeasures for emergent forms of algorithmic discrimination. As above, the second

point may be difficult in the case of complexly intersectional or emergent forms of algorithmic discrimination since they are so hard to identify because they do not provoke a socially recognizable form of stigmatization. However, the other dimensions can be readily extended to emergent algorithmic discrimination. This starts by ensuring that anti-discrimination institutions and officers are attentive to the possibilities of emergent discrimination. Thus, possibilities for challenging algorithmic verdicts and demands for redress need to be available to all, regardless of belonging to a specific protected group. Yet, this will only be possible via broad anti-discrimination logics and protections, that operate independently of specific protected grounds, for example by embracing provisions such as Article 14 of the ECHR that prohibits discrimination on the basis of any possible '***other status***.' Learning from arguments raised about amending anti-discrimination protections to encompass intersectional discrimination, there should be recognition of "the risks of compartmentalization generated by the existence of [specific] grounds for discrimination" (Marcat-Bruns, 2018: 48). In turn, this will contribute to the third dimension, to increase participation and voice, not only for representatives of certain groups, but for all who may be impacted by discriminatory processes. Anti-discrimination officers working in the field of algorithmic profiling should work less in the name of particular groups but towards broader dimensions of equality. Apart from legally institutionalized forms of voice, ideally practices like participatory design can raise awareness of complex forms of discrimination.

This would also conform to Fredman's (2016: 80) suggestion "that in designing proactive measures, groups should be defined not merely in terms of their status markers, but with reference to the particular aims of equality." Fredman (2016: 66) continues that "new intersectional groups should be recognized in their own right," and argues for entirely new grounds for discrimination – an argument that could also be applied to emergent forms of discrimination, provided they can be identified. Following Crenshaw's (1989) seminal piece there was wide recognition and acceptance, including within the judiciary, of intersectional forms of discrimination (see e.g. Marcat-Bruns, 2018). Drawing attention to the possibilities of emergent forms of discrimination in algorithmic profiling in this way may also contribute towards a rethinking of anti-discrimination approaches, particularly when they connect to exclusion and marginalization: "this recognition might narrow the focus on those who are most often disenfranchised at the intersection of multiple forms of subordination" (Marcat-Bruns, 2018: 47, citing Crenshaw). Moreover, the existing safeguards for the legally encoded protected groups need to be ameliorated. Crenshaw writes that such simple lists are not grounded in a bottom-up commitment to improve the substantive conditions for those who are victimized by the interplay of numerous factors. Instead, the dominant message of anti-discrimination law is that it will regulate only the limited extent to which race or sex interferes with the process of determining outcomes (Crenshaw 1998, 151). This insight can be carried over to algorithmic judgment as well. It might help to construct safeguards for protected grounds in a way that reflects that discrimination might not be experienced by all members (e.g. all women), but only some. This would help wherever new emergent forms of discrimination include members of protected groups or categories.

Another important approach to diagnosing and protecting against emergent forms of discrimination comes from the post-colonial view that practices of control and power have been developed in complex back-and-forth traffic between the West and its colonies, and that have always included data gathering and processing (Foucault, 2003; Legg, 2007; Thatcher et al.,

2016). Increasing global flows of data, and the relative ease to tap into them, has made algorithmic profiling an important tool that extends the reach of states' institutions beyond their national borders. The 'Five-Eyes' spying collaboration of the US and four Commonwealth states including the UK, imports the British colonial legacy into the very structure of the internet (Mann and Daly, 2019). Thus, when new criteria are formed through algorithms, they have to be assessed against the backdrop of a global surveillance system that transports its own norms and processes of suspicion. Therefore, post-colonial attention to (in)visibilities and marginalization is important, especially regarding algorithmic profiling that is used to protect borders, migration and other 'outsides' (Adey, 2012; Monahan, 2017). As Mann and Daly (2019) show, algorithmic profiling continues many of the colonial practices of creating margins, outsides, and invisibilities of excluded subjects. For example, data-based border controls have become decisive in the processing of migration, asylum requests, and the ensuing actions like moving people to detention camps (Mann and Daly, 2019). Here, the verdicts of algorithmic profiling are directly related to exertions of power. These, then, are further instances where seemingly harmless data are directly invested with powerful measures that are hard to challenge. Further, as Monahan (2017: 202) argues, these types of marginalizing surveillance produce new forms of "exclusionary invisibility" where algorithmic profiling is aimed at persons who are hardly visible, and who often do not fall under the scope of existing protections. This social invisibility is mirrored and augmented if the emergent categories are also 'invisible' from the point of view of existing anti-discrimination protection. It becomes an invisible production of invisibilities.

An important cue to analyse newly created categories is the question of whether they enforce, facilitate, or legitimise, such exclusionary invisibilities. Thus, the fight against emergent forms of discrimination through anti-discrimination protections runs the risk of continuing the colonial practices of providing safeguards by creating exclusions. Here we are making the dynamics of exclusionary invisibility, in both algorithmic profiling, and anti-discrimination logics, more visible. There is a need for further research in order to make such invisibilities visible. This may include algorithmic accountability and auditing initiatives that seek to identify when, why, and how, emergent discrimination is occurring, yet opening the 'black-box' (Pasquale, 2015; Amoore, 2011) is likely to be challenging. Perhaps one way of doing so, and aligned with current movements in the field of Artificial Intelligence (AI), is incorporating such post-colonial sensibilities for power structures into the development of ethical frameworks for AI, and specifically into measures of 'fairness, accountability and transparency', although we acknowledge critiques of "ethics washing" as a way to side-step hard law and regulation (Wagner, 2018). However, a more successful route for implementing such attention to invisible shifts of power might be in social and political forms of oversight and potential new legislation that needs to address which forms of data should be allowed for algorithmic profiling and thus be scrutinized, and which actions should or should not be invested with the power that algorithmic profiling creates.

Conclusion

In this article, we have analysed algorithmic profiling as a process of knowledge construction from large sets of data that often bear no direct relation to the protected grounds of anti-discrimination laws. Still, they form complex intersectional and non-representative categories that may bring about systematic disadvantage for a hitherto unnoticed group of people. We term this process as emergent discrimination. There are limits to the applicability of both data protection and anti-discrimination law in responding to new forms of discrimination that emerge, or that do not pertain directly to protected identities, but rather represent patterns that have little or no intuitive meaning to human practice. However, we have shown that the intuition of anti-discrimination law can be carried over to these new forms of discrimination. Inspired by intersectional reconceptualizations of justice and the ensuing proposals for institutional amendments, we have shown potential remedies. Furthermore, post-colonial attention to (in) visibilities is required to counter the risk of continuing marginalization. These insights should inform ethical assessments, design processes, and other proactive protective measures in creating and applying algorithmic profiling.

References

Adey P (2012) Borders, identification and surveillance. In: Lyon D, Ball K, and Haggerty KD (eds) ***Routledge Handbook of Surveillance Studies***. London: Routledge, pp. 193–201.

Amoore L (2011) Data derivatives: On the emergence of a security risk calculus for our times. ***Theory, Culture and Society*** 28(6): 24-43.

Article 29 Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136'. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/index_en.htm (accessed 25 November 2019).

Article 29 Working Party 'Opinion 5/2014 on Anonymization Techniques 0829/14/EN/WP216.' Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/index_en.htm (accessed 25 November 2019).

Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679.' Available at: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 (accessed 25 November 2019).

Bano M (2018) Artificial intelligence is demonstrating gender bias – and it's our fault. ***Kings College London News Centre.*** Available at: https://www.kcl.ac.uk/news/news-article.aspx?id=c97f7c12-ae02-4394-8f84-31ba4d56ddf7 (accessed 25 November 2019).

Browne S (2015) ***Dark matters: On the surveillance of blackness.*** Durham: Duke University Press.

Buttarelli G (2016) The EU GDPR as a clarion call for a new global digital standard. ***International Data Privacy Law*** 6(2): 77-78.

Campolo A, Sanfilippo M, Whittaker M and Crawford K (2017) ***AI Now 2017 Report***. ***Technical report***. New York: AI Now Institute.

Coll S (2014) Power, knowledge, and the subjects of privacy: Understanding privacy as the ally of surveillance. ***Information, Communication and Society*** 17: 1250–1263.

Crenshaw K (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. ***The University of Chicago Legal Forum*** 1(8): 139-176.

Danna A and Gandy OH (2002) All that glitters is not gold: Digging beneath the surface of data mining. ***Journal of Business Ethics*** 40: 373-386.

Dencik L, Hintz A, and Cable J (2016) Towards data justice? The ambiguity of anti-surveillance resistance in political activism. ***Big Data and Society*** 3(2): 1-12.

Edwards L and Veale M (2017) Slave to the algorithm: Why a right to explanation is probably not the remedy you are looking for. ***Duke Law and Technology Review*** 16(1): 18-84.

Eubanks V (2018) ***Automating inequality: How high-tech tools profile, police and punish the poor***. London: St Martin's Press.

Ferguson AG (2017) ***The rise of big data policing: Surveillance, race and the future of law enforcement.*** New York: NYU Press.

Foucault M (2003) ***Society Must Be Defended: Lectures at the Collège de France, 1975-76***. In: Bertani M, Fontana A and Ewald F (eds) New York: Picador.

Fredman S (2016) ***Intersectional discrimination in EU gender equality and non-discrimination law.*** Brussels: European Commission. Available at: https://publications.europa.eu/en/publication-detail/-/publication/d73a9221-b7c3-40f6-8414-8a48a2157a2f (accessed 25 November 2019).

Friedman B and Nissenbaum H (1996) Bias in computer systems. ***ACM Transactions on Infor-***

mation Systems 14(3): 330-347.
Gandy OH (2010) Engaging rational discrimination: Exploring reasons for placing regulatory constraints on decision support systems. ***Ethics and Information Technology*** 12: 29-42.
Gellert R and Gutwirth S (2013) The legal construction of privacy and data protection. ***Computer Law and Security Review*** 29(5): 522-530.
Gellert R, de Vries K, de Hert P and Gutwirth S (2013) A comparative analysis of anti-discrimination and data protection legislation. In: Custers B, Calders T, Schermer B and Zarsky T (eds) ***Discrimination and Privacy in the Information Society.*** Berlin: Springer-Verlag, pp. 61-89.
Gilliom J (2001) ***Overseers of the poor: Surveillance, resistance and the limits of privacy.*** Chicago: University of Chicago Press.
Harcourt BE (2010) ***Risk as a Proxy for Race***. ID 1677654, SSRN Scholarly Paper, 16 September. Rochester, NY: Social Science Research Network. Available at: https://papers.ssrn.com/abstract=1677654 (accessed 18 July 2018).
Hildebrandt M (2008) Defining profiling: A new type of knowledge? In: Hildebrandt M and Gutwirth S (eds) ***Profiling the European Citizen: Cross-Disciplinary Perspectives.*** Dordrecht: Springer, pp. 17-45.
Kitchin R (2014) ***The data revolution: Big data, open data, data infrastructures and their consequences***. Los Angeles: Sage.
Kleinberg J, Mullainathan S and Raghavan M (2016) Inherent trade-offs in the fair determination of risk scores. In: ***Proceedings of Innovations in Theoretical Computer Science*** (ITCS), 2017.
Koops B (2014) The trouble with European data protection law. ***International Data Privacy Law 4***(4): 250-261.
Leese M (2014) The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. ***Security Dialogue*** 45(5): 494-511.
Legg S (2007) Beyond the European Province: Foucault and Postcolonialism. In: Crampton JW and Elden S (eds) ***Space, Knowledge and Power: Foucault and Geography***. Aldershot, England; Burlington, VT: Ashgate, pp. 265–289.
Lyon D (2014) Surveillance, Snowden, and big data: Capacities, consequences, critique. ***Big Data & Society*** 1(2): 1-13.
Lyon D (2003) Surveillance as social sorting: Computer codes and mobile bodies. In: ***Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination.*** New York: Routledge.
Lyon D (2001) ***Surveillance society: Monitoring everyday life***. Buckingham: Open University Press.
Mann M and Daly A (2019) (Big) data and the North-in-South: Australia's informational imperialism and digital colonialism. ***Television and New Media*** 20(4): 379-395.
Marcat-Bruns M (2018) Multiple discrimination and intersectionality: Issues of equality and liberty. ***International Social Science Journal*** 67(223-224): 43-54.
Matzner T (2014) Why privacy is not enough in the context of "ubiquitous computing" and "big data." ***Journal of Information, Communication and Ethics in Society*** 12(2): 93-106.
Matzner T (2016) Beyond data as representation: The performativity of big data in surveillance. ***Surveillance & Society*** 14(2): 197-210
Monahan T (2017) Regulating belonging: Surveillance, inequality, and the cultural production of abjection. ***Journal of Cultural Economy*** 10(2): 191-206.
Noble SU (2018) ***Algorithms of oppression: How search engines reinforce racism.*** New York: NYU Press.
O'Neil C (2016) ***Weapons of math destruction: How big data increases inequality and threatens***

democracy. UK: Penguin, Random House.
Parsons C (2015) Beyond privacy: articulating the broader harms of pervasive mass surveillance. *Media and Communication* 3(3): 1-11.
Pasquale F (2015). *The blackbox society: The secret algorithms that control money and information.* Cambridge, Massachusetts; London, England: Harvard University Press.
Peña Gangadharan S (2012) Digital inclusion and data profiling. *First Monday* 17(5).
Purtova N (2018) The law of everything: Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology* 10(1): 40-81.
Safari B (2017) Intangible privacy rights: How Europe's GDPR will set a new global standard for personal data protection. *Seton Hall Law Review* 47*:* 809-848.
Sandvig C, Hamilton K, Karahalios K and Cedric Langbort C (2016) When the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software. *International Journal of Communication* 10: 4972-4990.
Schermer B (2013) Risks of profiling and the limits of data protection law. In: Custers B, Calders T, Schermer B and Zarsky T (eds) *Discrimination and Privacy in the Information Society.* Berlin: Springer, pp. 137-152.
Schreurs W, Hildebrandt M, Kindt E and Vanfleteren M (2008). Cogitas, ergo sum. The role of data protection law and non-discrimination law in group profiling in the private sector. In: Hildebrandt M and Gutwirth S (eds) *Profiling the European Citizen: Cross-Disciplinary Perspectives.* Dordrecht: Springer, pp. 241-270.
Selbst AD and Powels J (2017) Meaningful information and the right to explanation. *International Data Privacy Law* 7(4): 233-242.
Stalder F (2002) Privacy is not the antidote to surveillance. *Surveillance & Society* 1: 120–124.
Thatcher J, O'Sullivan D and Mahmoudi D (2016) Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space* 34(6): 990–1006.
Vedder A and Naudts L (2017) Accountability for the use of algorithms in a big data environment. *International Review of Law, Computers & Technology* 31(2): 206–224.
Verloo M (2006). Multiple inequalities, intersectionality and the European Union. *European Journal of Women's Studies* 13(3): 211-28.
Wachter S, Mittelstadt B and Floridi L (2017) Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation, *International Data Privacy Law* 7(2): 76-99.
Wachter S and Mittelstadt B (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review* 2: 1-130.
Wagner B (2018) Ethics as an escape from regulation: From ethics-washing to ethics-shopping? In M Hildebrandt (ed) *Being Profiled: Cogitas Ergo Sum.* Amsterdam: Amsterdam University Press, pp. 84-44.
Zuboff S (2015) Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30(1): 75-89.

The Brussels Privacy Hub Working Papers series

N°1 "The data protection regime applying to the inter-agency cooperation and future architecture of the EU criminal justice and law enforcement area" (November 2014) by Paul De Hert and Vagelis Papakonstantinou (35 pages)

N°2 "The new cloud computing ISO/IEC 27018 standard through the lens of the EU legislation on data protection" (November 2014) by Paul de Hert, Vagelis Papakonstantinou, Irene Kamara (25 pages)

N°3 "Towards efficient cooperation between supervisory authorities in the area of data privacy law" (October 2015) by Dariusz Kloza, Antonella Galetta (24 pages)

N°4 "The data protection regime in China" (November 2015) by Paul De Hert and Vagelis Papakonstantinou (30 pages)

N°5 "The right to privacy and personal data protection in Brazil: time for internet privacy rights?" (February 2016) by Vinícius Borges Fortes (23 pages)

N°6 "Permissions and Prohibitions in Data Protection Jurisdiction" (May 2016) by Mistale Taylor (25 pages)

N°7 "Structure and Enforcement of Data Privacy Law in South Korea" (October 2016) by Haksoo Ko, John Leitner, Eunsoo Kim and Jong-Gu Jung (20 pages)

N°8 "The "Right to be Forgotten" and Search Engine Liability" (December 2016) by Hiroshi Miyashita (15 pages)

N°9 "European Human Rights, Criminal Surveillance, and Intelligence Surveillance: Towards "Good Enough" Oversight, Preferably but Not Necessarily by Judges" (March 2017) by Gianclaudio Malgieri & Paul De Hert (25 pages)

N°10 "Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw" (July 2017) by Meg Leta Jones, JD, PhD (31 pages)

N°11 "The Microsoft Ireland case and the cyberspace sovereignty trilemma. Post-territorial technologies and companies question territorial state sovereignty and regulatory state monopolies" (July 2018) by Paul De Hert and Johannes Thumfart (27 pages)

N°12 "Understanding the balancing act behind the legitimate interest of the controller ground: a pragmatic approach" (August 2018) by Irene Kamara and Paul De Hert (35 pages)

N°13 "Big data analytics by telecommuni-cations operators and the draft ePrivacy Regulation" (September 2018) by Vagelis Papakonstantinou & Paul de Hert (14 pages)

N°14 "Enforcement in Indonesia Data Privacy Laws: The 2018 Facebook-Cambridge Analytica scandal as a case study" (October 2018) by Anbar Jayadi (21 pages)

N°15 "Belgium, Courts, Privacy and Data Protection. An inventory of Belgian case law from the pre-GDPR regime (1995-2015)." (January 2019) by Paul De Hert (34 pages)

N°16 Big data analytics in electronic communications: A reality in need of granular regulation (even if this includes an *interim* period of no regulation at all) (June 2019) by Vagelis Papakonstantinou & Paul de Hert (25 pages)

N°17 Data Localisation: Deconstructing myths and suggesting a workable model for the future. The cases of China and the EU (September 2019) by Author: Yanqing Hong, Senior Fellow, Law and Development Institute, Peking University of China, Edited by Vagelis Papakonstantinou, Brussels Privacy Hub (31 pages)

N°18 Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination (January 2020) by Dr Monique Mann and Professor Tobias Matzner (18 pages)

The Brussels Privacy Hub Working Papers series

The Brussels Privacy Hub Working Papers are intended to circulate research in progress for comment and discussion. The Working Papers focus on all areas of data protection and privacy research and can contain empirical research on privacy issues and analytical work on privacy governance and regulation in the EU; global flows of data; reconciling law enforcement and privacy interests; privacy challenges posed by new technologies; comparative research on privacy in different regions; jurisprudential issues of privacy protection; and many others

Available at www.brusselsprivacyhub.org/publications.html

Contact: info@brusselsprivacyhub.eu

BRUSSELS PRIVACY HUB