

Proxy Discrimination and Legal Disruption

The disruptive power of reality

Robbe van Rossem*

Reality is mirrored in the many algorithmic systems that are increasingly embedded in our everyday life. When data that refers to real-world phenomena is used in algorithmic systems, an insightful reconstruction of reality is generated. This image of reality becomes more complete as greater amounts of data are involved and as this data is interpreted more intelligently. The trend of a greater mirroring of reality can, however, also trigger a legal disruption, as the law can be confronted with a reality alternative to the one it implies itself. This risk exists particularly in the context of discrimination. In its application to the algorithmic context, non-discrimination law has to examine the very systems that generate a mirroring of reality. This paper investigates the disruptive effects such a confrontation with reality can have for the law in the particular case of proxy discrimination. The features of discriminatory proxies are namely highly descriptive of the structural inequality and discrimination that characterizes society. When theories critical of the limits of non-discrimination law are subsequently confirmed by the reflections in the data, the law faces increased pressure to justify its current scope of and approach to illegal discrimination. While a true disruption depends on the willingness of the law to take on an position of self-reflection, it is argued that any distortion arising from the reflections in the data can hardly be called technological in nature.

* University of Copenhagen – University of Ghent [robbevanrossem@gmail.com]

1. Introduction

Reality is reflected in the data our world generates. In our so-called information society, little of this reality is free from being captured in digital form. Vast amounts of data are imported from analogue collections, captured in our online behaviour, collected through the observations of scientific research, etc. The recording and collection of all this data is, however, not without its purpose – and certainly not without its use. Big Data has proven that the consideration of great amounts of information can be extremely valuable for i.a. making decisions or predictions. Great interest thus exists to subject large proportions of data to processes of interpretation such as data mining. The results can be astounding as the data can reveal more than what we thought to know about our world. As human curiosity – or simply the desire for efficiency, knows little to no limit, also more intelligent technology like artificial intelligence has been put to the task to get the most out of our data. As a result, reality is increasingly being mirrored in the systems we use to parse it.

A ‘boxed-in’ overview of reality can be very enlightening. It is, however, the question whether the law is capable of dealing with the revelations that come with this increased understanding of the world we live in. The reflections of reality that can be found in systems subject to the law’s control could easily prove themselves to be overwhelming to the law, and as a result be disruptive. This is potentially the case in the context of discrimination. Algorithmic systems have been plagued by discriminatory results. While algorithmic discrimination always has caused a variety of difficulties for the application of non-discrimination law, the trend of an increased insight in the reality of discrimination could be especially problematic in this regard. After all, the legal frameworks that exist to protect the right to non-discrimination are often criticized for their blindness regarding reality and the discrimination that occurs in it. Now this reality is reflected in the systems that are subjected to the law’s examination, non-discrimination law’s claimed ignorance towards certain aspects of discrimination is again challenged.

The paper explores the possibility of such a disruption in the particular case of proxy discrimination. This particular form of algorithmic discrimination occurs when information on protected, discriminatory-sensitive characteristics is hidden in other, seemingly neutral data that is used by an algorithm. The

protected characteristics are themselves not included in the data, yet highly correlate with information that is fed to the algorithm – which are called their ‘proxies’. When this data is used as an input for the algorithm, the protected characteristics can indirectly influence the algorithm’s output, and as a result place the members of the protected group or class at a possibly illegal disadvantage. The discriminatory effect thus occurs because of the algorithm’s reliance on information that also happens to be indicative for a protected class. A postal code can, for example, function as a proxy for the protected characteristic race, considering neighbourhoods can have racial profiles due to the ethnicity of their inhabitants. Subsequently, when a decision is based on subjects’ postal code, inhabitants of historically racialized neighbourhoods can be discriminated, as the decision will indirectly be based on their race or ethnicity, even though the feature was not directly included in the decision-making process.

The paper commences with an explanation of the occurrence of proxies in datasets and the discrimination that can come from that (2). Next, the capability of an, at least partial, recreation of reality is demonstrated by the means of proxy discrimination’s features (3). After a concise look at certain critiques of non-discrimination law’s ignorance to the world it operates in (4), the paper discusses the disruption faced by the law and the unique nature that characterizes it (5).

2. Proxy discrimination

2.1 Proxies

Describing reality – for instance human beings – involves comparing corresponding features and adding significant values to as many as possible. Some of the features are independent and fully complementary (e.g. a first name and last name, a postal code and telephone number, ...). Others are more or less related to each other (birth date and age, body weight and clothing size, ...). When these features overlap to the extent that their correlation can cause them

to refer to the same information, they are regarded to be proxies to one another.¹ For example, your body mass index score (BMI) can indicate whether you are overweight. Similarly, the fact that your clothing size is XL or higher can bear the same information. As a result, a high BMI score and clothing size XXL are proxies for overweight but also for one another. From the perspective of the feature both information points relate to, their coexistence can thus be characterized by redundancy, as the pieces of data can easily substitute each other in a dataset while their mutual information remains intact. Whether this is favourable depends on the information that is reflected and the situation in which it is used.

2.2 Popularity of Proxies

The presence of proxies in data has increased significantly in the last years. In times when storing data was still cumbersome and expensive, redundant information such as proxies was always carefully avoided. In the age of Big Data such concerns are long gone. The capability to store vast amounts of data very cheaply has facilitated the trend of connecting and copying databases without any concern for identical information. The fact that these databases were developed from different perspectives actually adds information to the entire system, allowing for more patterns and conclusions to be found by Big Data tools. In a way, proxies have changed from being a nuisance to serving as a commodity. The popularity of proxies does, however, not necessarily solely relate to the coexistence of multiple substituting information points within the used datasets.

Alternatively, proxies can also be useful precisely when their counterpart is missing from a dataset. They are an efficient tool to include information in a dataset that itself is difficult to observe, unavailable or simply not allowed to be used. It can, for example, be very difficult and costly to determine someone's driving style. The observation would require multiple tests, interviews, field trials, etc. If, however, general test results would be available that indicate that

¹ Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104 California Law Review 671, 691.

male drivers predominately adopt an aggressive driving style, it is tempting to use the easily observable characteristic of gender as a proxy for someone's potential behaviour on the road.² Similarly, in the infamous practice of redlining financial institutions used postal codes in their decision to provide certain services such as granting loans.³ Although areas can coincide with particular levels of income, it has been established how this choice was based in racial animus and prejudice.⁴ In this way, geographic information functioned as a 'masked' replacement for an applicant's ethnicity or race, which is of course an illegal basis for differentiation.⁵

2.3 Proxy discrimination defined

The presence of proxies clearly cannot be considered to be desirable in all instances. When referring to certain sensitive characteristics, they can significantly add to the problem of algorithmic discrimination. Of course, algorithms are bound to discriminate in a technical sense; they are designed to tap into the vast amounts of data our 'scored society' generates for the exact purpose of evaluating, ranking, classifying, ... subjects in a manner that exceeds human cognition and fatigue, which naturally implies differentiation.⁶ While many of these differentiations are considered to be acceptable, illegal discrimination arises when they infringe the rules of non-discrimination law. For most frameworks of non-discrimination law, this implies that a differentiation was based on one of the societally important characteristics the law has rewarded

² Toon Calders and Indre Zliobaite, 'Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures' in Toon Calders and others (eds), *Discrimination and Privacy in the Information Society* (Springer 2013) 52–53.

³ Hunt Bradford, 'Redlining', *Encyclopedia of Chicago* (2005) <<http://www.encyclopedia.chicagohistory.org/pages/1050.html>> accessed 4 January 2020.

⁴ Barocas and Selbst (n 1) 689.

⁵ Andrea Romei and Salvatore Ruggieri, 'Discrimination Data Analysis: A Multi-Disciplinary Bibliography', in Calders and others (n 2) 121.

⁶ Claude Castellucia and Daniel Le Métayer, *Understanding Algorithmic Decision-Making: Opportunities and Challenges* (STOA 2019) 7.

a special legal protection.⁷ Attributes commonly included in these ‘protected characteristics’ are race, gender, sexuality, religion, etc.⁸ When one of these characteristics is used as a direct input or ground for a decision, the illegal discriminatory nature of the output is blatantly clear.⁹ Simply excluding these characteristics from the model does, however, not always suffice to prevent a discriminatory result. Proxies for protected characteristics may namely be lurking in the data, allowing the prohibited characteristics to have a continuing influence on the output of the algorithm. It is this indirect effect protected characteristics can have through their proxies, that causes proxy discrimination.

In its simplest form, proxy discrimination can be defined as a differentiation based on facially-neutral characteristics that significantly correlate with membership to a protected class.¹⁰ Although the protected characteristics are not directly involved in e.g. the decision-making process, they can have a similar discriminatory impact when they are represented by proxies that happen to be present in the data. The facially absent protected characteristics can thus be so-called ‘redundantly encoded’ in the dataset.¹¹ This is the case when ‘a particular piece of data or certain values for that piece of data are highly correlated with membership in specific protected classes.’¹² Present by representation, the legally-prohibited characteristics continue to impact the output of the algorithm,

⁷ Raphaële Xenidis and Linda Senden, ‘EU Non-Discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination’ in Ulf Bernitz and others (eds), *General Principles of EU law and the EU Digital Order* (Kluwer Law International 2020) 5.

⁸ Dagmar Schiek, Lisa Waddington and Mark Bell, *Cases, Materials and Text on National, Supranational and International Non-Discrimination Law* (Hart Publishing 2007) 510; Christopher McCrudden and Sacha Prechal, ‘The Concepts of Equality and Non-Discrimination in Europe: A Practical Approach’ (European Network of Legal Experts in the field of Gender Equality 2010) 60, 23.

⁹ Xenidis and Senden (n 7) 19.

¹⁰ Barocas and Selbst (n 1) 691–692. See also Anya Prince and Daniel Schwarcz, ‘Proxy Discrimination in the Age of Artificial Intelligence and Big Data’ (2019) 105 *Iowa Law Review* 1257, 1266 (who clarify that proxy discrimination relates more specifically to ‘scenarios in which an algorithm uses a variable whose predictive power derives from its correlation with membership in the suspect class’).

¹¹ Barocas and Selbst (n 1) 691.

¹² *ibid* 691–692.

and as a result place the members of a protected class at a possibly illegal disadvantage when they are subjected to the discretion of such an algorithm. In the classic example of redlining, for example, the decision to grant a loan based on a subject's postal code does not directly involve a protected characteristic. It can, however, indirectly amount to a proxy discrimination when areas and neighbourhoods highly correlate with racial profiles, as the subject's postal code would act as a proxy for their race or ethnicity. As proxy discrimination occurs in the form of a practice that facially appears to be neutral, yet disproportionately harms members of a protected class, it is often regarded as a specific subcategory of indirect discrimination.¹³

3. Mapping discrimination

Proxies add a great deal to the persistence of the problem of algorithmic discrimination. Their existence, however, also touches upon something more fundamental concerning the notion of discrimination itself. The redundant encodings offer a lens through which to observe discrimination not only as it appears in the algorithm, but also how it occurs in the real world. After all, it has to be reminded that data is a reflection of reality. In a way, an intelligent processing of data merely offers a cartography of the world we live in. The conclusions derived from the use of the data are only relevant given their analogy with what exists in the real world. Similarly, the information discriminatory proxies reflect and the relations they imply facilitate a 'mapping of discrimination'. This capability of proxy discriminations to map reality can be found in two of its features which coincidentally are of great importance in a judicial review on the illegal discriminatory nature of an algorithmic output.¹⁴ This paper discusses the trade-off between fairness and utility proxies impose on the designers of algorithmic models (3.1) and the endless amounts of proxies that are redundantly encoded in the data (3.2) to conclude on the harsh truth both features bring (3.3).

¹³ Prince and Schwarcz (n 10) 1260.

¹⁴ See *infra* 5.1. on the legal relevance of these features.

3.1 Trade-off

Redundant encodings have proven to be a difficult problem to solve. In case they could be detected, their deletion or exclusion from the model is not always a viable option. The information that doubles as a proxy for membership to a protected class is often ‘genuinely relevant in making rational and well-informed decisions’.¹⁵ As is mentioned above, the use of geographical information like postal codes can, for example, lead to an illegal discriminatory effect for certain groups as neighbourhoods can have different racial profiles.¹⁶ An individual’s address can, however, be highly relevant in a job related context, as the distance between home and workplace is a strong indicator for employee engagement.¹⁷ This confronts designers of algorithmic models with a difficult trade-off between fairness and utility.¹⁸ While withholding proxies from the data could seem beneficial in an attempt to secure a non-discriminatory result, their exclusion implies a high cost for the overall accuracy of the model as meaningful information would be missing from the decision-making or prediction process.¹⁹

Although a difficult balancing exercise for the designers of algorithmic systems, the utility-fairness trade-off also showcases how a deeper look into proxies can provide a meaningful addition to our perception of the discrimination faced by certain groups. Through the correlation between sensitive characteristics on the one hand and attributes that are relevant for a rational and well-informed decision on the other, the trade-off indicates how class membership can impactfully condition which traits an individual possesses. After all, one of the main reasons why members of certain classes are systematically discriminated against when ‘objective’ target variables are used, is

¹⁵ Barocas and Selbst (n 1) 691.

¹⁶ Romei and Ruggieri (n 5) 121.

¹⁷ Don Peck, ‘They’re Watching You at Work’ (*The Atlantic*, December 2013) 72 <<https://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/>> accessed 6 January 2020.

¹⁸ Philipp Hacker, ‘Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law’ (2018) 55 *Common Market Law Review* 1143, 1150.

¹⁹ Barocas and Selbst (n 1) 721.

that these relevant criteria happen to be possessed by classes at different rates.²⁰ This is of course no claim for superiority or inferiority of certain classes. Instead the phenomenon shines a light on the structural and systemic nature of discrimination.²¹ The trade-off originates from the wish to avoid a discriminatory output. What risks such an outcome, is the translation of existing inequality in the disposition connected to class which is reflected in the data. By revealing the disposition and the permeating effects class membership can have, proxy discrimination forces the observer to place instances of discrimination in a broader context. Notably, discrimination can not only be the cause of inequality, it can also very well be the result of it.

3.2 Lines of proxies

In the event that a proxy is detected and the designer indeed sacrifices predictive accuracy by excluding it from the model, this decision can still be futile as there may be many more proxies for the same protected characteristic encoded in the data.²² This possibility naturally increases as the amount of input data grows. In rich enough datasets, the chance for the redundant encoding of protected characteristics not only reaches near certainty, but often also presents itself in a way that the encodings are redundant to each other.²³ When a proxy is excluded for the purpose of a non-discriminatory output, other proxies for the same protected characteristic will simply continue the discriminatory effect.²⁴ In these instances ‘endless lines of proxies’ can be observed as the proxies can easily replace each other.²⁵ As a result, the attempt to exclude all proxies would have you block information at zero.²⁶ Even if it would be possible to design a system

²⁰ Sandra Mayson, ‘Bias In, Bias Out’ (2018) 128 *The Yale Law Journal* 2218, 2257–2259; Romei and Ruggieri (n 5) 130.

²¹ Barocas and Selbst (n 1) 691.

²² Ignacio Cofone, ‘Algorithmic Discrimination Is an Information Problem’ (2019) 70 *Hastings Law Journal* 1389, 1416.

²³ Barocas and Selbst (n 1) 695; *ibid* 1414.

²⁴ Cofone (n 22) 1414.

²⁵ Cofone (n 22) 1416.

²⁶ *ibid* 1414.

with such an objective, its purpose would quickly be defeated as the lack of remaining information would reduce the results to mere randomness.²⁷

An important contribution to this obstinacy of proxy occurrence is the fact that proxies do not always present themselves in the form of clear, single substitutes for the protected characteristic that is aimed to be excluded from the model.²⁸ A data particle might also only slightly correlate with a protected characteristic, to a degree that it seems to be completely neutral when observed individually.²⁹ In aggregation, however, the correlations of the different information points could cluster into a proper proxy.³⁰ As a data particle's potential to contribute to the formation of dispersed proxies may only be revealed in aggregation, each data point can theoretically be suspected to hold such a dormant potential for a discriminatory output when it would be combined with the corresponding data points. Illustrative for such dispersed proxy formations are the various kinds of personal traits and attributes that can be observed through someone's activity on social media. A single like on a social networking site such as Facebook is unlikely to reveal the user's sexuality or political views. The accumulation of likes, however, allows social media platforms to observe precisely such highly sensitive personal attributes of their users.³¹

The seemingly infinite chain of proxies that can be observed in large data collections builds on the previous feature of the trade-off to allow for a mapping of discrimination. Where the utility-fairness trade-off highlights how attributes can be distributed unequally between classes, the proxy lines show how many

²⁷ *ibid.*

²⁸ *ibid.*

²⁹ *ibid* 1413.

³⁰ *ibid* 1414.

³¹ Michal Kosinski, David Stillwell and Thore Graepel, 'Private Traits and Attributes Are Predictable from Digital Records of Human Behavior' (2013) 110 *Proceedings of the National Academy of Sciences* 5802, 5802; Solon Barocas, Moritz Hardt and Arvind Narayanan, *Fairness and Machine Learning: Limitations and Opportunities* (2019) ch 2 <<https://fairmlbook.org/classification.html>> accessed 26 October 2020 ('Several features that are slightly predictive of the sensitive attribute can be used to build high accuracy classifiers for that attribute').

attributes can actually be connected to the disadvantaged position membership to a certain class brings forth. Correlation per correlation, the inspection of a data collection for proxies of a protected characteristic shows how wide the impact of class membership can be. In a sense, the proxy lines unearth the branches of inequality and discrimination. Naturally the clarity of this image increases proportional to the data volume. The fact that the reflections of inequality are hardly inescapable in large data collections, that discriminatory potential can be lurking even in the smallest things in life, and that this is sometimes only observable when the respective data points are placed in the right constellation of data, is similarly to the trade-off revealing with regards to the nature of discrimination. Namely, rather than the consequence of a particular decision to discriminate, the proxy discrimination seems to be an expression of the systems and environment we live in. It is exactly these processes of disadvantage that can be observed through their exclusionary consequences that are recorded in the data in the form of proxies for protected characteristics.

3.3 An (in)convenient truth

The features of discriminatory proxies unveil that it is reality that produces discriminatory practices, not the machine. The discriminatory results that roll out of an algorithm are not to be reduced to purely virtual phenomena. Their discriminatory nature stems from the real world, whose inherent inequality resonates in the data the machine is being fed.³² The idea that the origins of algorithmic discrimination can also lay outside the algorithm is, however, not too shocking. Historical biases have been illustrative in this regard as they show that when data that reflects a discriminatory past is fed to an algorithm, the algorithm will reproduce similar discriminatory practices.³³ Sandra Mayson's play on the old computer-science adage 'garbage in, garbage out' wittily summarizes this as 'bias in, bias out.'³⁴ Thus, to the extent that the data actually

³² Xenidis and Senden (n 7) 7.

³³ Barocas and Selbst (n 1) 681.

³⁴ Mayson (n 20) 2224.

represents reality, the algorithm is bound to perpetuate that reality to the same, unequal image.³⁵

Proxies offer, however, a more impactful realisation than that inequality and discrimination are a product of the real world. The discussed features show that proxy discriminations also shine a light on the nature and the construction of discrimination. The use of its lessons thus exceeds the algorithmic context, as the revelations can also be insightful for the analysis of discrimination in general. Guided along the many proxies present in a model, the data reveals the effect membership to a class can have on the life of an individual and subsequently the decisions he or she is subjected to. In a way, proxy discriminations illustrate the interaction between inequality and discrimination and hint at the structural nature of both.³⁶ As a result, the exploration of discriminatory proxies places discriminatory practices in a broader context. It can, however, be questioned whether the occurring image of discrimination is compatible with the notion of discrimination held by the law. After all, after a walk along the ‘contours of inequality’, the current focus of non-discrimination law on individual cases of discrimination suddenly seems to be extremely narrow if not naïve.³⁷

4. Non-discrimination law

Proxy discrimination provides us with a broader picture of discrimination than the single discriminatory acts the law tends to focus on. Naturally this can be of great advantage for the fight against illegal discrimination. The insight offered by the use of algorithms and AI with regard to the construction of inequality and discrimination can be used for technological, socio-political and possibly even legal progress.³⁸ As regards the legal dimension, it can, however, be questioned

³⁵ Xenidis and Senden (n 7) 7.

³⁶ Cf *ibid* 7–9 (‘Structural discrimination, which is the product of past discrimination institutionalised over time and now reflected in many ways in the organisation of society, is mirrored in data’).

³⁷ For ‘contours of inequality’ see Barocas and Selbst (n 1) 721.

³⁸ Cf Mayson (n 20) 2284 (‘Because predictive algorithms transparently reflect inequality in the data from which they are built, they can also be deployed in reverse: as diagnostic tools to identify sites and causes of racial disparity in criminal justice’).

whether the law can parse the many revelations a ‘mapping of discrimination’ brings forth. Many critiques concerning the established limits of non-discrimination law are based on the claim that the law is blind to certain aspects of discrimination as presented in reality. When these theories are confirmed in the reflections of reality that are captured in the data, it can be argued that non-discrimination law experiences an increased pressure to justify its scope if it does not change its approach. After all, with the implications of opposing theories lurking in the very systems it has to assess, the law faces greater difficulty in maintaining its particular conception of discrimination. The following part discusses a number of critiques on the limits of non-discrimination law relevant for the revelations of i.a. proxy discrimination.

4.1 Intersectionality

An often called upon limit of non-discrimination law is its tendency to address a discriminatory act from the perspective of only one characteristic.³⁹ This single-axis approach to discrimination is central to the critique formulated in the literature on intersectionality.⁴⁰ Coined by Kimberlé Crenshaw in her 1989 feminist critique of the US antidiscrimination doctrine, the concept of intersectionality denotes the various ways in which personal characteristics interact with each other and as a result shape unique experiences for those residing in their overlap.⁴¹ When standing on an actual intersection, you could be hit by traffic coming not only from one direction, but from each direction, and possibly even at the same time.⁴² Similarly can a person be discriminated against simultaneously on the basis of his or her gender, race, religion, etc. Crenshaw explains, however, that this does not necessarily result in situations of ‘additive’ discrimination, where a differentiation is based on the combination of

³⁹ Anna Lauren Hoffmann, ‘Where Fairness Fails : Data , Algorithms , and the Limits of Antidiscrimination Discourse’ (2019) 22 *Information, Communication & Society* 900, 905.

⁴⁰ *ibid.*

⁴¹ Kimberle Crenshaw, ‘Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory, and Antiracist Politics’ [1989] *University of Chicago Legal Forum* 141.

⁴² *ibid* 149.

multiple, yet still distinguishable grounds.⁴³ Characteristics can also interact in a way that their combination can no longer be disentangled.⁴⁴ The discrimination experienced by a black woman, for example, is not per se based on her gender or her race, nor necessarily on the accumulation of both grounds, but can instead be focused on her black womanhood in particular.⁴⁵

Ignorance of these intragroup differences comes at a great risk. Crenshaw's critique of intersectionality sought more than a more accurate mapping of identity categories.⁴⁶ Central to her thinking was non-discrimination law's role in the reproduction of social hierarchy and inequality. She argued that by reducing experiences of discrimination to a single characteristic, the law banishes those whose experience cannot fully be grasped by one of the protected characteristics, to a permanent stay in the 'basement' of society.⁴⁷ This metaphorical basement will at one point host all disadvantaged people. Nevertheless, it reproduces the hierarchy that exists above ground.⁴⁸ A relative privilege is namely given to those whose experience can actually be fully addressed by one of the protected characteristics, as only they can claim their rise to the 'ground level'.⁴⁹ The other inhabitants of the basement can try to demand their own rise to equality using the same claims, and in this way strengthen the demands of the relatively privileged, but will at least partially be bound to stay in the basement.⁵⁰ For example, a black woman will support the fight against singular gender or race discrimination by using the corresponding characteristics to inaccurately address her own experience, yet cannot use these same handles to claim her own rise to a state of non-discrimination.⁵¹

This risk for reproduction of a social hierarchy by non-discrimination law through the mobilization of a socio-legal privilege remains existent today. Although extremely insightful for the experience of the discriminatee, the law

⁴³ Schiek, Waddington and Bell (n 8) 171.

⁴⁴ Crenshaw (n 41) 149.

⁴⁵ *ibid.*

⁴⁶ Anna Carastathis, 'Basements and Intersections' (2013) 28 *Hypatia* 698, 699.

⁴⁷ Crenshaw (n 41) 151.

⁴⁸ Carastathis (n 46) 710.

⁴⁹ Crenshaw (n 41) 151.

⁵⁰ *ibid.*

⁵¹ *ibid.*

has yet to adopt the theories of intersectionality.⁵² As a consequence, aspects of the discriminatory experience not addressed by the chosen characteristic are still rendered invisible. At the same time, many facets of discrimination that the law ignores are now recorded in the data used by the algorithm. Provided that the transparency of the algorithm is not obstructed by the complexity of the system, the use of algorithms allows, for example, for a more detailed determination of which grounds actually played a role in the result of a discriminatory output.⁵³ This also means that the intersectional nature of discriminatory practices becomes more visible, placing non-discrimination law's position under increased pressure. Moreover, as the amount of relevant information is increased, one could imagine a situation where the retainment of its single-axis approach could cause non-discrimination law to be inapplicable to any experience of discrimination, as none of the protected characteristics has enough of an impact on the output to amount to an illegal discrimination.

4.2 Protected characteristics

The conclusions of intersectionality are only more troubling when one looks at the narrow set of characteristics that are granted explicit legal protection. Most statutes within the framework of non-discrimination law operate on a limited list of grounds on which the discrimination has to be based in order to be considered illegal. Similar to how non-discrimination law can disadvantage victims of discrimination whose experience is only partially covered by one of the protected characteristics, the lack of recognition in any of the protected characteristics can render a discriminatory experience completely invisible to the law. This is not bizarre, as not every differentiation is a discrimination. Grounds commonly included are race, gender, religion, sexuality, disability and age.⁵⁴

⁵² Mieke Verloo, 'Multiple Inequalities, Intersectionality and the European Union' (2006) 13 *European Journal of Women's Studies* 211, 211.

⁵³ Talia B Gillis and Jann L Spiess, 'Big Data and Discrimination.' (2019) 86 *University of Chicago Law Review* 459, 474.

⁵⁴ American College of Emergency Physicians, 'Non-Discrimination,' (2006) 47 *Annals of Emergency Medicine* 510; McCrudden and Prechal (n 8) 1–60, 23.

Meanwhile differentiations on, for example, the basis of beauty⁵⁵, financial status⁵⁶ or vegan preference are currently not deemed discriminatory from the perspective of non-discrimination law.

Which characteristics are included in the list reveals to a certain degree the ruling definition of discrimination within the legal regime at hand. While many statutes share a considerable amount of characteristics, various theories exist on the rightful basis of the inclusion of these attributes. A popular foundation for legislators' reasoning of a list of protected characteristics is the idea that the grounds for illegal discrimination should track existing social categories worthy of protection.⁵⁷ This still leaves enormous room for discrepancy between legal frameworks, as concepts such as 'social category' or 'social group' are rather open and dynamic.⁵⁸ It can, for example, be debated what degree of saliency is required of the social group⁵⁹, whether its members must have experienced a form of subordination due to a power balance,⁶⁰ or whether the societal context should even play a role at all.⁶¹ To increase the potential diversity, each of these orientations allows for multiple perspectives. A grouping attribute might, for example, be considered to be defining by the broader public while it does not play a significant role in the subject's perspective on its own identity, and vice versa.⁶²

Whichever position is adopted with regard to the defining determinant for rewarding legal protection to characteristics, this choice will increasingly have to

⁵⁵ William R Corbett, 'Hotness Discrimination: Appearance Discrimination as a Mirror for Reflecting on the Body of Employment-Discrimination Law' (2011) 60 *Catholic University Law Review* 615.

⁵⁶ Frederik Zuiderveen Borgesius, 'Discrimination, Artificial Intelligence, and Algorithmic Decision-Making' (Council of Europe 2018) 35.

⁵⁷ Natalie Stoljar, 'Discrimination and Intersectionality' in Kasper Lippert-Rasmussen (ed), *The Routledge Handbook of the Ethics of Discrimination* (Routledge 2018) 72–78.

⁵⁸ *ibid* 68.

⁵⁹ *ibid*.

⁶⁰ Patrick Shin, 'Discrimination and Race' in Lippert-Rasmussen (n 57) 203.

⁶¹ Deborah Hellman, 'Discrimination and social meaning' in Lippert-Rasmussen (n 57) 97.

⁶² Tal Zarsky, 'An Analytic Challenge: Discrimination Theory in the Age of Predictive Analytics' (2017–18) 14 *I/S: A Journal of Law and Policy for the Information Society* 11, 16.

be justified as our image of the unfair differentiations made in society becomes more clear. Most of the positions discussed above are at least partly based on a moral appreciation of what *ought* to be, or more fitting what *should not* be. The relevance of reality for the adopted theory for the selection of a particular set of characteristics is, however, not to be underestimated. For example, when legislators embrace the idea that people suffer discrimination as a member of an identifiable social group, it can be argued that they should be aware of the social tags that dominate daily life.⁶³ Therefore, as long as the adopted theory is based on contingent social factors rather than purely on preconceived moral notions, it can be expected that when confronted with reality the protected characteristics indeed appear to be relevant, at least in the context of the chosen theory for legal protection. Now the use of algorithmic systems increasingly reveals the relevance of non-protected characteristics for unfair outcomes, the chosen theories are increasingly tested on their justification for the inclusion of only a few characteristics.

4.3 Bad actor frame

Ultimately, non-discrimination law's inability to address or perceive the 'full picture' of discrimination may be criticized in reference to the law's focus on the misaligned conduct of individual perpetrators.⁶⁴ With the neutralization of the actions of perpetrators as its main concern, the law seems to ignore important systemic and social issues.⁶⁵ From this perspective, discrimination is namely seen as being caused by atomistic, discrete events that operate outside a social fabric or historical continuity.⁶⁶ Important structural aspects may thus be overlooked as the discrimination is viewed as a particular wrongdoing rather than a social phenomenon.⁶⁷ This individualistic approach is most visible when non-

⁶³ Stoljar (n 57) 72, 78.

⁶⁴ Hoffmann (n 39) 904.

⁶⁵ *ibid*; Alan David Freeman, 'Legitimizing Racial Discrimination Through Antidiscrimination Law: A Critical Review of Supreme Court Doctrine' (1978) 62 *Minnesota Law Review* 1049, 1049.

⁶⁶ Neil Gotanda, 'A Critique of "Our Constitution is Color-Blind"' (1991) 44 *Stanford Law Review* 1, 44.

⁶⁷ Freeman (n 65) 1054.

discrimination law openly revolves around intent and a narrow conception of causation, as the requirement of such a fault easily reveals the hunt for ‘blameworthy’ perpetrators.⁶⁸ De-emphasizing these aspects, however, for example, through the incorporation of disparate impact or unintentional discrimination, does not seem to widen the law’s gaze too much as the focus remains firmly on discrete sources.⁶⁹

Much of non-discrimination law’s discrete source mentality can be traced back to the core mechanism of its design. Alan David Freeman, for example, explains how the core concept of ‘violation’ leads to such a narrow view on discrimination by inherently siding with the perspective of the perpetrator.⁷⁰ He points out that discrimination could instead be approached from the perspective of the victim.⁷¹ From this perspective, discrimination describes the conditions of social existence as a member of the particular group (e.g. employment, housing, education, the psychological effects of being perceived as a member of a group rather than as an individual, etc.). Here, the eradication of discrimination would imply the detection of all the contributing conditions associated with discrimination and consequently their elimination.⁷² From the perspective of the perpetrator, however, discrimination is conceived purely as the actions inflicted on the victim by that perpetrator.⁷³ Therefore, the remedy does not involve an overall improvement of the conditions of the victim’s life, but instead limits itself to the neutralization of the misaligned conduct.⁷⁴ It is on this basis that Freeman claims that by limiting its remedy to the ‘violation’ by the perpetrator, the law is hopelessly indifferent to the social, systemic context of discrimination as is reflected in the condition of the victim.⁷⁵ And let it be exactly the latter that is to be found in the data upon inspection of e.g. the features of proxy discrimination.

⁶⁸ Hoffmann (n 39) 905.

⁶⁹ *ibid.*

⁷⁰ Freeman (n 65).

⁷¹ *ibid* 1053.

⁷² *ibid.*

⁷³ *ibid.*

⁷⁴ *ibid.*

⁷⁵ *ibid* 1054.

5. Legal disruption

The presence of proxies of protected characteristics in datasets shows itself to be ambiguous. On the one hand, their presence in the data contributes significantly to the obstinacy of discriminatory outputs in algorithmic systems and prevents designers from finding a simple solution for these events. On the other hand, they allow for a more accurate mapping of inequality and discrimination itself. While the latter can be helpful in the fight against discrimination, it could also lead to uncomfortable conclusions for the law. Ultimately, supported by legal literature critical of modern non-discrimination law, the conception of discrimination implied by the reflections in the data appears to be incompatible with the current limits of non-discrimination law. This paper argues that although a confrontation with the image of discrimination evoked by proxies is inevitable for the law (5.1) and this potentially could be disruptive on a fundamental level (5.2) an actual legal disruption depends on the degree to which the law is willing to look itself in the mirror (5.3). In any case, the possible disruption resulting from non-discrimination law's confrontation with the reality reflected in the data is, notwithstanding its many technological requirements, not to be regarded as technological in nature (5.4).

5.1 An inevitable confrontation

An encounter with the 'reality of discrimination' seems unavoidable in the judicial examination of proxy discrimination. The application of non-discrimination law in cases of algorithmic discrimination is flawed in many ways, causing many to contemplate the optimal route to be taken in this context.⁷⁶ At all events, however, it has to be proven that the algorithmic system indeed is or is not discriminatory.⁷⁷ Whether this is established under direct or indirect discrimination, the confrontation with the implications of the features of proxy discrimination is bound to occur in the subsequent assessment of the justification of differentiation. After all, most non-discrimination statutes deem a differentiation as justified when i.a. requirements of necessity and

⁷⁶ See eg Barocas and Selbst (n 1); Hacker (n 18); Xenidis and Senden (n 7).

⁷⁷ Xenidis and Senden (n 7) 21.

proportionality are fulfilled.⁷⁸ In the case of algorithmic discrimination these tests are likely to come down to the evaluation of the trade-off between efficiency and non-discrimination made by the developers of the system.⁷⁹ In such instances, judges are not only confronted with the ground truth of the unequal distribution of goods, skills, etc., as implied by the trade-off itself, but they also have to interact with the second feature of proxy discrimination discussed in this paper. After all, in order to evaluate a particular balance in the exclusion and preservation of proxies, one should at least have a superficial idea of the amount of proxies present in the data. As a result, the adjudicating body is forced to follow the ‘lines of proxies’ to a point where their inconvenient truth can no longer be avoided.

5.2 A fundamental disruption

Once observed by a court in its analysis of a case of supposed algorithmic discrimination, the unveiled ‘reality of discrimination’ shows itself to be disruptive for non-discrimination law. The discriminatory proxies found in the data reflect an image of discrimination which is incompatible with the conception currently held by the law. Supported by legal theory critical of the current demarcations of non-discrimination law, the features of proxy discrimination imply the necessity to consider i.a. contextual, structural and systemic aspects of discrimination, and overall demand a broader and more nuanced approach to events of illegal differentiation.⁸⁰ Reminded that data merely reflects the reality it applies to, the discrepancy between law and what is mirrored quickly leads to an alarming conclusion: non-discrimination law is in

⁷⁸ See eg Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L180/22, art 2(b); Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services [2004] OJ L373/37, art 2(b); Directive of the European Parliament and of the Council 2006/54/EC of 5 July 2006 on the implementation of the principle of equal opportunity and equal treatment of men and women in matters of employment and occupation [2006] OJ L204/23, art 2(1)(b).

⁷⁹ Xenidis and Senden (n 7) 22.

⁸⁰ See supra part 3.

its current form incapable of fully considering discrimination as it presents itself in reality. This deficiency may hamper the doctrine's effectiveness for bringing about attempted positive change.⁸¹ At least, this is the case if one agrees with the popular opinion in academic literature that non-discrimination law finds its goal in directing 'social change to eliminate group-based status inequalities'.⁸²

The fundamental nature of this disruption follows from the adaptations to non-discrimination law required to accommodate the image in the mirror. Unfortunately, it is unlikely that it would suffice to simply broaden the scope of non-discrimination law. The law namely not only overlooks relevant social demarcations, intragroup differences or overall noteworthy experiences of discrimination, but it is also blind to the structural and systemic origins of many exclusionary practises.⁸³ A blindness that finds its significance in the unobstructed, if not re-entrenched continuation of these structures. An attempt to integrate these realisations in the law arguably implies a great intervention in its construction and its approach to discrimination. Take for example the critique that the law wrongly focusses on the misaligned conduct of faulty perpetrators, as illustrated by Freeman.⁸⁴ As this is a commentary on the law's core approach to discrimination, adapting its gaze to this conclusion would be fundamentally disruptive for the law's current shape and limits. The reflections of the data could thus not only require a calibration of non-discrimination law to the projected reality, they could also force it back to the drawing board.

⁸¹ Hoffmann (n 39) 901.

⁸² For discussions of the so-called antsubordination theory, see Ruth Colker, 'Anti-Subordination above All: Sex, Race, and Equal Protection' (1986) 61 *New York University Law Review* 1003; Kenneth Karst, 'Why Equality Matters' (1982) 48 *Sibley Lecture Series*. <https://digitalcommons.law.uga.edu/lectures_pre_arch_lectures_sibley/48/> accessed 15 September 2020; Jack M Balkin and Reva B Siegel, 'The American Civil Rights Tradition: Anticlassification or Antisubordination' (2003) 58 *University of Miami Law Review* 9; Abigail Nurse, 'Anti-Subordination in the Equal Protection Clause: A Case Study' (2014) 89 *New York University Law Review* 293; Cass Robert Sunstein, 'The Anticaste Principle' (1994) 92 *Michigan Law Review* 2410; Samuel R Bagenstos, 'The Structural Turn and the Limits of Antidiscrimination Law' (2006) 94 *California Law Review* 1.

⁸³ See *supra* part 4.

⁸⁴ Freeman (n 65).

5.3 Forced self-questioning

Nevertheless, any claim of a legal disruption by proxy discrimination ought to be nuanced by non-discrimination law's own influence on this matter. After all, proxies do not create an immediate obstacle for the application of non-discrimination law, however inescapable or infinite their presence may be. Instead, the disruption stems from the law's confrontation with an awkward image of reality. This image is, however, far from new.⁸⁵ The legal critiques mentioned in this paper have been well-established for decades, and undoubtedly must have come to all actors of the law's awareness.⁸⁶ Thus, nothing stops the law from continuing this alleged ignorance as before, regardless of the negative implications this may have for the eradication of discrimination. In the end, the law holds a factual monopoly on the decision of what it regards as discriminatory, and could turn a blind eye for the mere reason it does not wish to be disrupted. Furthermore, it has been addressed by others how a more structural approach might demand too much from non-discrimination law, and rather belongs to 'the realm of politics and social change...than to the narrow confines of legal doctrine'.⁸⁷ However strikingly diagnostic data's mirroring of reality thus may be, its image only proves to be disruptive where the law allows it to be.

However, the reflections in the data already make a compelling case for the law to embrace their implications. After all, to the degree that the law strives to base itself on the reality it tries to bring order to, it can be highly discreditable to disregard the reality which it is constantly confronted with in its application to e.g. proxy discrimination. Furthermore, the mirrored reality shows the law more than simply the structures and mechanisms of a socially stratified world. It may also confront the law with its own role in the continuation of inequality. Law's blindness to the reality of discrimination does namely not only allow discrimination and inequality to proceed at the same pace, but it can also

⁸⁵ Mayson (n 20) (who argues that algorithms merely shine a new light on the old problem of racial inequality in risk assessment).

⁸⁶ See eg Bagenstos (n 82) (describing a 'structural turn' in academic literature); Verloo (n 52) (documenting a growing body of studies and comments on multiple discrimination and intersectionality).

⁸⁷ Bagenstos (n 82) 45.

reproduce, entrench and exacerbate the disadvantage present in society.⁸⁸ Finally, the law inevitably observes the reflected image in the data, regardless of whether it later chooses to ignore it. It is thereby wise to make use of the diagnostic capabilities this ‘clear mirror’ offers, rather than to blindly proceed relying on the ‘cloudy mirror’ that is inherent to human decision.⁸⁹

5.4 A non-technological disruption

The disruption faced by non-discrimination law as a result of its confrontation with the reflections of reality lurking in the data, is not easily situated within the existing literature on legal disruption by technology. First of all, it can be questioned whether technology is directly responsible for the disruption discussed in this paper. The vast amounts of data, the computational power, the assistance of artificial intelligence etc. are of course necessary for the reflections to be shown to non-discrimination law in this particular way. Their role is, however, merely facilitative with regards to the disruption. Contrary to many other discussions concerning algorithmic discrimination, such as the difficulty of the opacity and complexity of certain algorithms, it are not the technical characteristics of the technology involved that create a difficulty for the application of the law.⁹⁰ Instead, the disruption is caused by the message these

⁸⁸ Barocas and Selbst (n 1) 674 (‘Approached without care, data mining can reproduce existing patterns of discrimination, . . . It can even have the perverse result of exacerbating existing inequalities by suggesting that historically disadvantaged groups actually deserve less favorable treatment.’); Elise Boddie, ‘Adaptive Discrimination’ (2016) 94 North Carolina Law Review 1235, 1266 (‘Time does not inevitably lead to improvement if we misunderstand the problem. In fact, if anything, time can exacerbate the problem if we leave the malady untreated’); Crenshaw (n 41) 151.

⁸⁹ See Mayson (n 20) 2224 (who explains that subjective prediction by humans reflects the past similarly to algorithmic prediction. Human prediction is, however, based on less reliable anecdotal data. The precise algorithmic mirror should thus not be discarded for the cloudy one).

⁹⁰ On the challenges of opaqueness and inexplainsibility of algorithms Borgesius (n 56) 34; Danielle Keats Citron and Frank Pasquale, ‘The Scored Society: Due Process For Automated Predictions’ (2014) 89 Washington Law Review 1; Tal Zarsky, ‘The Trouble

technological tools bring, not by the medium by which it is delivered. Secondly, it can similarly be questioned whether a need for change comes from a shift in the sociotechnical landscape. Both the disruption itself, as well as the necessary adaptations it requires from the law to overcome it, can hardly be based on the effects newly enhanced technological capabilities have on people's activities or environment.⁹¹ After all, the image reflected in proxy discriminations has not enlightened our society with a new, changed look on discrimination. The only novelty is that the law is now directly confronted with an old truth it was comfortable ignoring for a long time.

6. Conclusion

Reality is mirrored in the data that is used in the various algorithms that increasingly rule our lives. An intelligent processing of this data allows for a mapping of reality, which is accompanied by an increased understanding of the phenomena observed through the data. This development can disrupt non-discrimination law, as also existing inequalities and discriminations are reflected in the data. Observed through the lens of the many discriminatory proxies that lure in algorithmic systems, a broad notion of discrimination imposes itself on the law. As a result the law faces a potential disruption. Confronted with the reflections in the data, it can no longer ignore the world outside its scope, and thus, experiences an increased pressure to justify its limits. Non-discrimination law's position is only more problematized now that many theories and critical literature regarding the current state of non-discrimination law find basis in the data. The proxies in the data thus hold up a mirror to the law, challenging it to examine itself. While modern technology facilitates the mirrored image, there is nothing technological about the reality it depicts.

with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making' (2016) 41 Science, Technology & Human Values 118.

⁹¹ See for a more detailed discussion of sociotechnical change: Lyria Bennett Moses, 'Regulating in the Face of Sociotechnical Change' in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (2017).