Big data, AI, and proxy discrimination: a review of Professor Margarida Lima Rego's 'dissent' on insurance discrimination bans

Madison Stevens*

Abstract

In July 2021, Professor Margarida Lima Rego delivered a presentation on her dissent from discrimination bans in insurance. Since the 2011 ruling of the Court of Justice of the European Union in *Test Achats,* it has been settled that the use of statistical discrimination within the European Union is prohibited. Dissenting from this approach, Professor Lima Rego proposes a more nuanced approach to statistical discrimination. Instead, she advances a twofold test to distinguish between admissible and inadmissible forms of discrimination. Within her dissent, Professor Lima Rego notes how the rise of AI and big data is changing insurance practice and shifting questions about the acceptability of statistical discrimination. This paper will provide a review of Professor Lima Rego's dissent and elaborate on her claims that AI and big data are changing the kinds of discrimination we see in insurance. Specifically, it will explain the concept of "proxy discrimination" which is likely to occur when machine-learning algorithms are used. This kind of proxy discrimination is exceedingly difficult to detect and manage under discrimination law. In recognition of this, this paper will explore whether the balanced approach proposed by Professor Lima Rego would be more appropriate to regulate the changing nature of insurance discrimination.

Keywords: statistical discrimination; proxy discrimination; artificial intelligence; big data; consumer insurance

1. Introduction

As part of the theme of the Special Issue, Professor Lima Rego's dissent centred on the blanket bans on discrimination in insurance imposed within the European Union (EU). These bans mean that all discrimination based on protected characteristics, such as sex or race, is prohibited, even where the differentiation of treatment is based on accurate statistical data. Lima Rego dissents from the view that statistical discrimination should be banned outright. Rather, she suggests that a more nuanced approach to insurance discrimination should be adopted. Instead of blanket bans, she proposes that differentiation based on protected characteristics should be allowed where a twofold test is met. The first requirement of the test is that the actuarial data which supports the discrimination must not be based on past discrimination. The second requirement is that there must be no other suitable (or even slightly less suitable) factors which can be used to instead predict the risk. Lima Rego contends that arguments about statistical discrimination will become more prevalent as the insurance industry adopts technological advancements such as artificial intelligence (AI) and big data. As AI and big data are used in insurance underwriting, the kinds of discrimination that we may see will differ from the insurance discrimination which led to the decision to effectively ban statistical discrimination in *Test Achats*.[1] As such, Lima Rego suggests

* PhD Researcher at the University of Southampton. This paper is based on the presentation delivered by Professor Margarida Lima Rego at the 'Dissent in Insurance Law' webinar on 06 July 2021, organised by the *British Insurance Law Association* and the *Corporate and Financial Law Research Group* of the University of Glasgow. Professor Lima Rego's presentation can be viewed at www.gla.ac.uk/schools/law/research/groups/corporate-and-financial/news/headline_801711_en.html.

that the questions around the acceptability of insurance discrimination will change. This paper is a review of the presentation delivered by Lima Rego and seeks to unpack her submissions regarding AI, big data, and insurance discrimination.

Within the past few years, the insurance industry has entered the beginning of a technological revolution. Insurance-focused AI (often termed "Insurtech") is rapidly evolving and creating numerous opportunities to add value to the insurance business chain.[2] Though still at the early stages of adoption, the insurance industry has started to see "meaningful experimentation" with AI which, when implemented, could change the way in which insurance has traditionally operated.[3]

One of the key areas which stands to benefit from the use of AI is risk classification. Risk classification refers to the process of grouping different insured individuals together according to the risk they represent. Traditionally, these groups have been determined based on a limited number of risk factors or characteristics, which have been found to be predictive of risk using actuarial and statistical analysis. AI, when fuelled by big data, provides opportunities to enhance this process. By automating the risk classification process and supplying more data, AI and big data is promising to provide more precise risk assessments and to allow insurers to predict risk more accurately.

Yet, despite the promised benefits of the use of AI and big data, its adoption for risk classification is concerning because of its ability to discriminate. Under traditional risk classification processes, easily available and simple characteristics were used to ascertain the separate classes. These have, in the past, included protected characteristics such as sex. This has led to debates about how 'fairness' can be achieved in insurance. On the one hand, insurers have argued that the use of protected characteristics advances "actuarial fairness" and ensures individuals are charged premiums which reflect their individual risk. On the other hand, it has been argued that the use of protected characteristics constitutes a disproportionate interference with an individual's rights and freedoms. The adoption and AI and big data for risk classification adds an additional complexity to this debate. Commentators have highlighted the potential for AI and big data to result in a new kind of discrimination: unintentional proxy discrimination by machine-learning algorithms. Where machine-learning algorithms are prohibited from using protected characteristics, proxy discrimination will occur where an algorithm finds new and difficult to detect correlates for the protected characteristic. For example, an algorithm may find that whether an individual drinks tea or not correlates with an individual's sex.[4] Where an algorithm is forbidden from using "sex" within its risk classification process, but sex proves to be an actuarial accurate predictor of risk, the algorithm may use "whether the individual drinks tea" as a proxy risk factor. The resulting scenario is that it appears that there has been no discrimination of protected characteristics, yet the use of the consumption of tea indicates an individual's sex and differentiates on that basis. This practice allows bias and discrimination to discreetly make

[1] Judgment of the Court (Grand Chamber) of 1 March 2011, *Association Belge des Consommateurs Test-Achats ASBL and Others v Conseil des ministers,* C-236/09, EU:C:2011:100.

[2] Arrk Group, 'An Introduction to Insurtech: What You Should Know About This New Industry' (*Arrk Group,* 12 March 2019) <www.arrkgroup.com/thought-leadership/an-introduction-to-insurtech-what-you-should-know-about-this-new-industry/> accessed 07 December 2021.

[3] Shanique Hall, 'How Artificial Intelligence is Changing the Insurance Industry' (*Centre of Insurance Policy and Research,* August 2017) <https://www.naic.org/cipr_newsletter_archive/vol22_ai.pdf> accessed 22 December 2021; Annmarie Geddes Baribeau, 'The Insurtech Revolution' (2018) 45 The Actuarial Review 32.

[4] Say, for example, a woman is more likely to drink tea than a man.

its way into the algorithm's pricing model and it can unfairly burden protected groups. It is predicted that proxy discrimination will be inevitable where protected characteristics have predictive value. As highlighted by Lima Rego, therefore, big data and AI have shifted old debates about the 'fairness' of direct statistical discrimination to focus on the indirect and opaque nature of proxy discrimination. This paper aims to describe how big data and machine-learning algorithms will change perceptions of fairness in risk classification and consider whether equality law is equipped to deal with these changes. Addressing the potential contribution that the ban on statistical discrimination, the paper will also consider whether an alternative approach to insurance discrimination, such as Lima Rego's twofold test, can be used to lessen the risk of new forms of proxy discrimination.

Section Two of this paper will provide a brief background to the statistical discrimination debate and the law on its prohibition. Section Two will also set out alternative approaches to statistical discrimination, specifically Lima Rego's twofold test. Section Three will then move to discuss how the use of big data and machine-learning algorithms will transform the risk classification process and introduce worrying new forms of discrimination that may be difficult to deal with under discrimination law. Section four will explore whether an approach proposed by Lima Rego will be better equipped to regulate unintentional proxy discrimination.

2. Risk classification and discrimination

The insurance business relies on the uncertainty of future events. For risk adverse individuals, insurance is a means by which they can redistribute the financial responsibility of uncertain events onto insurers in exchange for premiums.[5] This uncertainty fuels the demand for insurance products. If an individual could tell with 100% certainty when and how events could happen, the desire for transfer risks onto insurers would simply cease to exist.

Yet, some degree of certainty is required for insurance companies to operate profitably. To manage their total risk exposure, insurers must know the likelihood that the risks they insure will occur and the losses they will internalise when they do. Risk classification has long been the method insurers use to manage their risks. Risk classification refers to the process of separating insured individuals into homogenous groups based on the risk they present. Using the law of large numbers, insurers can then predict how much loss will occur within each group.[6] The groups are differentiated according to certain factors, known as risk characteristics. Risk characteristics are identified by actuaries by statistically analysing past claims data and can be used to predict an insured's behaviour and the likelihood that the risk will materialise. For example, an individual's occupation may be used to predict their care levels while driving. At the underwriting stage, an insurer compared the information an individual has supplied with the risk characteristics to place the individual in the correct risk class. The insurer charges them premiums which reflect the risk exposure of the class they have been placed in.

[5] Tom Baker, 'Risk, insurance, and the social construct of responsibility' in Tom Baker and Jonathon Simon (eds), *Embracing Risk: The Changing Culture of Insurance and Responsibility* (University of Chicago Press 2002).

[6] The law of large numbers is a theorem that holds that the average of a large number of independently identical variables mirrors the expected value. Applied to insurance, as an insurer's exposure to losses increases, the predicted loss will be closer to the actual loss. See Michael L Smith and Stephen A Kane, 'The Law of Large Numbers and the Strength of Insurance' in Sandra G Gustavson and Scott E Harrington, *Insurance, Risk Management, and Public Policy* (Kluwer Academic Publishers 1994).

A core part of insurance therefore involves differentiating between classes using certain characteristics.[7] In this sense, it is "necessarily discriminating".[8] Whilst this is generally accepted, a debate has arisen as to what characteristics are appropriate to be used for risk classification. In particular, the use of sex has been a subject of controversy. Sex is, to use the terminology of American jurisprudence, a suspect classification. Suspect classifications are groups of people who are more likely to be discriminated against due to historic injustice against them. Because of this, their use as a means of differentiation will trigger a presumption of wrongness. Within the UK and EU, suspect classifications have received legal protection against discrimination through equality law, earning them their title as "protected characteristics". Protected characteristics include the likes of sex, race, religion, and sexual orientation. Historically, protected characteristics have been a popular factor used for risk classification. Whilst risk classification involves the statistical analysis of historic data to determine precise predictors of risk, the selection of which characteristics to use in risk classification is motivated by economic efficiency. The aim of insurers is not necessarily to find the most accurate predictors of risk, but to find the predictors which are as accurate only so far as they are profitable.[9] It will often be the case that the information that would tell the insurer the most about the risk is not readily available or quantifiable. Because of this, insurers will select easily obtainable and verifiable surrogates for this information. For instance, the less care that a motorist takes whilst driving, the more likely that the motorist will be in a motor accident. Accordingly, levels of care are a directly predictive risk factor. A motor insurer, then, would ideally like to know the care an individual typically takes when driving to determine the risk they present. However, care levels are not easily quantifiable or accessible using the resources available to insurers. Instead, the insurer will use a surrogate factor which can correlate with the care taken by motorists. Within the fields of personal insurance, sex has emerged as a popular risk characteristic because of its potential to proxy for a range of information. As a data point commonly collected at the underwriting stage, actuaries have an abundance of data on sex which can be used to determine how it correlates with certain risks. As a risk factor, sex is "comparatively available, reliable and verifiable", making it an economically efficient and effective risk characteristic.[10] For instance, for the purposes of life insurance, sex was used as a proxy for longevity after statistics showed that women live longer than men. On this basis, men and women were split into separate risk classifications and men were charged more for their insurance.[11]

2.1 Two opposing views of fairness: actuarial fairness v anti-discrimination

As equality has become an increasingly important legislative aim, anti-discrimination legislation has been enacted across the world to prohibit discrimination against protected groups. Alongside this, an important question has emerged: can insurers discriminate using protected characteristics where such differentiation is based on actuarial findings?

---

[7] Ronen Avraham, Kyle D Logue and Daniel Schwarz, 'Understanding Insurance Antidiscrimination Law' (2014) 87 S. Cal. L. Rev 195, 195.
[8] Margarida Lima Rego, 'Discrimination bans and insurance law' in Carmine di Nioa, Matteo Gargantini and Marco Lamandini, *Fundamental Rights and Financial Law* (Edward Elgar Publishing, forthcoming) 4.
[9] James Davey, 'Genetic Discrimination in Insurance: Lessons from Test Achats' in Gerard Quinn, Aisling de Poar and Peter Blanck, *Genetic Discrimination: Transatlantic Perspectives on the Case for a European Level Legal Response* (1st edn, Routledge 2014) 146.
[10] Lima Rego (n 8) 4.
[11] Richard A Booth, 'Sex, Lies and Life Insurance' (2012) 6 Va L & Bus Rev 403, 405.

For years insurers continued to use protected characteristics on the assumption that risk classification was a form of fair discrimination because of its actuarial links. Fair discrimination advances views of "actuarial fairness", a term which was first coined by Kenneth Arrow in 1963.[12] Actuarial fairness is the principle that fair premiums are those which match as closely as possible to the insured risk. It is a core rationale for the risk classification process and progresses the view that all differentiation is justified when based on "objective, relevant and reliable" actuarial findings.[13] The objective and scientific actuarial methods "represents people in terms of risk, that is, in terms of the contingent future".[14] In this sense, the differentiation in risk classification is simply a means to an end. By treating alike risks alike, insurers can charge individuals the premiums which most accurately reflect their risk. Placing limitations on which factors can be used for risk classification would undermine the accuracy and precision of the process and would result in individuals in lower risk individuals subsidising the risk of higher risk individuals. According to the conception of fairness as "equal treatment for equal risks",[15] this practice would be *morally* unfair.[16] In order for insurance to be fair, insurers should therefore be allowed the freedom to underwrite based on any factor which best predicts the risk.

Arguments for actuarial fairness claim that in addition to achieving fair and just outcomes, actuarial discrimination promotes economic efficiency and the proper functioning of the market. This is framed through the twin issues of insurance: adverse selection and moral hazard.[17] Adverse selection refers to the "theoretical tendency for low-risk individuals to avoid or drop out of insurance pools" where they have to subsidise for higher-risk individuals, leading to a "disproportionate percentage of high-risk individuals".[18] The idea is simple: where an insurer decides not to use a risk factor such as sex in risk classification, they will have to raise the average price of premiums to reflect the change of classification. The lowest risk insureds will then leave that insurer and seek insurance elsewhere, either with other insurers or alternative means of insuring the risk. The average level of risk rises, and the insurer must raise the average price of premiums to reflect the total risk exposure. This cycle repeats itself until eventually the pool will only contain the highest-risk individuals. Free and accurate risk classification can theoretically alleviate the effects of adverse selection.[19]

Moral hazard refers to the potential for insurance against loss to reduce incentives to minimise loss.[20] By redistributing the financial responsibility for loss, the risk is "being taken out of control of the individual".[21] Fair discrimination argues that pricing based on the risk an individual presents provides incentives to individuals to

[12] Kenneth Arrow, 'Uncertainty and the Welfare Economics of Medical Care' (1963) 53(5) The American Economic Review 941, 960.
[13] Swiss Re, 'Fair risk assessment in life & health insurance' (2011) 3.
[14] Jonathon Simon, 'The Ideological Effects of Actuarial Practices' (1988) 22 Law & Soc'y Rev. 771, 780.
[15] Spencer Leigh, 'The freedom to underwrite' in Tom Sorrell, *Health Care, Ethics and Insurance* (Routledge 1998) 33.
[16] Roberta B Meyer, 'The Insurer Perspective' in Mark A Rothstein (ed), *Genetics and Life Insurance: Medical Underwriting and Social Policy* (MIT Press 2004) 31.
[17] Xavier Landes, 'How fair is actuarial fairness?' (2014) 128 J Bus Ethics 519, 520.
[18] Tom Baker, 'Containing the Promise of Insurance: Adverse Selection and Risk Classification' (2003) 9 Conn. Ins. L. J 371, 373.
[19] Tom Baker and Rick Swedloff, 'Regulation by Liability Insurance: From Auto to Lawyers Professional Liability' (2013) 60 UCLA L. Rev. 1412, 1419; Rick Swedloff, 'Risk Classification's Big Data (R)evolution' (2014) 21 Connecticut Insurance Law Journal 342, 365.
[20] Tom Baker, 'On the Genealogy of Moral Hazard' (1996-1997) 75 Tex. L. Rev. 237, 239.
[21] Arrow (n 12) 961.

reduce their exposure to the risk to avoid high premiums.[22] The raising of the average price of premiums in response to limitations on risk factors may also have the effect of bringing the price of premiums for higher risks down. Where then, these higher risks are charged a premium which does not reflect their high risk, they are not motivated to change the level of their risk. Similarly, low-level risks who are taking active measures to keep their risk level low may no longer take those measures when their premiums do not reflect their efforts.

Because risk classification allows insurers to take control of the adverse selection and moral hazard,[23] it is argued that insurers should have a right to underwrite and use actuarial assessment to ensure the functioning of private markets.[24]

At the opposite end of the spectrum lies anti-discrimination arguments. Anti-discrimination focuses on a completely different account of distributive justice, centred on the "notion of human dignity, on individual rights and liberties".[25] Advocates of anti-discrimination oppose the use of protected characteristics as grounds for differentiation, arguing that it constitutes an unacceptable interference with an individual's freedom to have "decisions about how to live insulated from the effects of normatively extraneous features" of them, such as sex or race.[26] According to anti-discrimination, the existence of a statistical correlation between a protected characteristic and risk does not make its use in risk classification fair. Rather, the use of sex and other protected characteristics in risk classification is objected to on three core grounds: the potential to perpetuate past injustices; the lack of causal connection between the protected characteristic and risk; and the immutability of the protected characteristic.[27]

The potential of risk classification to perpetuate past injustices is inherent in actuarial methods which are used for risk classification. To determine the risk classes, actuaries carry out statistical analysis on *past* data and insurers identify the risk classifications, and accordingly what premiums should be charged, on this basis. This means that where the past data reflects historical inequities, the use of the protected characteristic risk classification will reinforce these inequalities and perpetuate stereotypes.[28] This causes "expressive harm by acknowledging and legitimating that prior unfair treatment".[29] This is a particular issue where the actuarial data insurers use are outdated and do not reflect current societal norms. This can be shown in the English law case of *Pinder v Friends Provident Life Office,* which affirmed actuarial fairness. The claimant had challenged the use of sickness statistics that were 30 years old were used to justify charging women 50% more than men for health insurance.[30] The court held that there was no requirement that the data be current. The actuarial data did not reflect shifting norms in society and could therefore be used to reinforce a protected group's position in society. Considering the

[22] Baker and Swedloff (n 19) 1419; Swedloff (n 19) 365.
[23] Rick Swedloff, 'The New Regulatory Imperative for Insurance' (2020) 61 B. C. L. Rev. 2031, 2041.
[24] Lisa Rebert and Ine Van Hoyweghen, 'The right to underwrite gender' (2015) 18 Tidjschrift voor Genderstudies 413, 420.
[25] Lima Rego (n 8) 6.
[26] Sophie Moreau, 'What is Discrimination?' (2010) 38(2) Philosophy and Public Affairs 143, 147.
[27] Davey (n 9) 146.
[28] Swedloff (n 23) 2042.
[29] Avraham et al. (n 7) 217.
[30] *Pinder v Friends Provident Life Office* (Westminster County Court, 1985).

importance of insurance today,[31] differentiation based on protected characteristics in insurance can repeat the cycles of institutional discrimination faced by protected groups.[32]

The concern that risk classification can preserve inequality between protected classes is heightened by the fact that risk factors are often merely *correlated* with risk, as opposed to being a *cause* of the risk.[33] Despite their directly predictive value, insurers do not necessarily seek to identify factors which cause the insured event. Rather, they will seek to use whatever factors or characteristics can be cheaply obtained and verified in practice. Protected characteristics tend to correlate with risk, as opposed to being a cause of the risk. Where there is no causal connection between the characteristic and the risk, the insurer is not able to explain exactly why the characteristic is predictive.[34] Using a characteristic which is not the cause of a loss can be problematic considering how its use infringes on the individual's freedom to not be unfairly discriminated against. This relates to the use of sex and other protected characteristics as surrogates for more direct, but unidentifiable, information. For instance, men have historically been charged more in premiums than women for motor insurance. This was because statistically more men were involved in motor accidents of greater severity. Men, however, did not get into more accidents *because* they were men but because their sex could correlate with behaviour and care. Sex, therefore, was merely a convenient correlation. As opposed to causative factors, correlated factors can appear quite arbitrary and are not necessarily predictive of *individual* behaviour.[35]

The issue of causation is closely related to the lack of control an individual has over their protected characteristic. Many protected characteristics, including biological sex and race, are uncontrollable: they are not the result of a choice an individual made. The use of noncontrollable characteristics can be criticised on the basis that they remove the relevance of "the exercise of individual responsibility" and choice from pricing in insurance.[36] The use of uncontrollable characteristics seems to contravene the moral hazard arguments often used to support actuarial fairness. Following the reasoning advanced by actuarial fairness, one may question how an individual may be motivated to alter their risk profile when the risk profile has been determined on uncontrollable and immutable characteristics. The main objection lies once again with fairness. Allowing individuals to be disadvantaged by characteristics they cannot change contributes to the systematic mistreatment of protected groups. This is starkly illustrated by the denial of life, health, and disability coverage to victims of domestic abuse based on actuarial findings that battered women are more likely to be victims of domestic abuse again in the future.[37] As opposed to choices such as smoking and eating habits, the use of sex and other protected characteristic in risk classifications "saddles people with all the consequences of their higher premiums, whether deserved or not...".[38]

[31] Insurance is required for many aspects of life such as driving a car or getting a mortgage. See Brian Glenn, 'The Shifting Rhetoric of Insurance Denial' (2000) 34 Law & Soc'y Rev 779, 783.
[32] Martin J Katz, 'Insurance and the Limits of Rational Discrimination' (1990) 8 Yale L. & Pol'y Rev. 436, 458.
[33] Ronen Avraham, 'Discrimination and Insurance' in Kasper Lippert-Rasmussen, *The Routledge Handbook on Ethics and Discrimination* (Routledge 2017).
[34] Swedloff (n 23) 2034.
[35] Barbara Underwood, 'Law and the Crystal Ball: Predicting Behaviour with Statistical Inference and Individualised Judgment' (1979) 88(7) The Yale Law Journal 1408.
[36] Kenneth S Abraham, 'Efficiency and Fairness in Insurance Risk Classification' (1985) 71(3) Virginia Law Review 403, 437.
[37] Deborah S Hellman, 'Is Actuarially Fair Pricing Actually Fair?: A Case Study on Insuring Battered Women' (1997) 32 Harvard Civil Rights – Civil Liberties Law Review 355.
[38] Baker (n 18) 394.

The competing conceptions of fairness came to a head in 2011 when the Court of Justice of the European Union (CJEU) was asked to contemplate whether an exception which allowed differentiation based on actuarial evidence was compatible with the fundamental rights guaranteed to citizens of the EU.

2.2 The legal landscape on discrimination: Test Achats

In 2004, the Council of the European Union passed Directive 2004/113/EC ("Gender Directive") aimed at implementing the principle of equal treatment between men and women in the provision of goods and services. Article 5(1) of the Gender Directive prohibited the use of sex as the cause for differentiation between individuals' premiums and benefits for the purposes of insurance.[39] Article 5(2) added a derogation from the rule in Article 5(1), allowing 'proportionate differences in the assessment of risk based on *relevant and accurate actuarial and statistical data*'.[40] This legitimised the use of sex in risk classification if it was backed by actuarial findings. The Belgian Government had transposed the Gender Directive and the derogation into national law. "Test Achats", a consumer rights organisation, challenged the freedom the national law gave to insurers to classify individuals by sex. The Belgian Constitutional Court subsequently referred the case to the CJEU, asking whether Article 5(2) was compatible with the general prohibition on the grounds of sex which is enshrined as a fundamental right within the Charter of Fundamental Rights of the European Union.[41]

Despite a nuanced analysis of discrimination law given by Kokott AG in her opinion,[42] the CJEU did not comment on whether actuarial fairness was compatible with the right to not be discriminated against. Instead, the court concluded that the derogation should be declared invalid because it could potentially apply in national laws without temporal limitation. It explained where the EU legislature has decided to take action on a particular issue, 'it must contribute, in a coherent matter, to the intended objective, without prejudice to the possibility of providing for... derogations of a limited scope'.[43] Because, however, the Directive had been silent as to the length of time actuarial discrimination could continue to be applied, the derogation from the principle of equal treatment represented in Article 5(1) could persist indefinitely. Their decision, in short, rested on poor legislative drafting. The CJEU did not conduct any meaningful analysis of whether the use of sex as an actuarial factor was consistent with principles of equal treatment, as insurers claimed it to be.

Nevertheless, the effect of the decision was that the derogation was invalid leaving only Article 5(1) to remain. The position on sex discrimination in insurance within the EU is therefore clear: insurers cannot use sex cannot be used as a risk factor where it results in differences in premiums and benefits unless the differentiation exists based on non-comparable situations.[44]

---

[39] 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 5(1) ('Gender Directive').
[40] i*bid*., art 5(2) (emphasis added).
[41] Charter of Fundamental Rights of the European Union [2012] OJ C326/391, art 21(1).
[42] Opinion of Advocate General Kokott delivered on 30 September 2010, *Association Belge des Consommateurs Test-Achats ASBL and Others v Conseil des ministers,* C-236/09, EU:C:2010:564.
[43] Judgment of the Court (Grand Chamber) of 1 March 2011, *Association Belge des Consommateurs Test-Achats ASBL and Others v Conseil des ministers,* C-236/09, EU:C:2011:100, paragraph 21.
[44] Gender Directive, recital 12.

This has subsequently been mirrored in English law in the Equality Act 2010. As service providers, insurers are prohibited from discriminating against any individual requiring their service.[45] After the decision in *Test Achats,* insurers cannot use any of the protected characteristics as risk factors, except for age and disability. Insurers may consider age within risk classification where they carry out an assessment of risk using information which is relevant to the risk assessment and from a source which is reasonable to rely upon.[46] Disability can be used as a risk factor where: (a) it is done by reference to information that is both relevant to the assessment of the risk and from a source on which it is reasonable to rely; and (b) it is reasonable to do that.[47]

2.3 A middle group? Professor Lima Rego's twofold test

Whilst the debate about discrimination and risk classification seems to be legally settled, it would be wrong to suggest that the issue is resolved. Because of the CJEU's lack of analysis into the compatibility of risk classification and the principle of equal treatment, the judgment has received criticism.[48] Lima Rego explains her dissent from the decision stating that while the decision is a "welcome landmark in the pursuit of equality between men and women", it nonetheless has "gone too far by saying too little on the question of what separates admissible criteria of differentiation from inadmissible forms of discrimination".[49] As some protected characteristics such as age and disability continue to be acceptable grounds for differentiation where based on relevant actuarial data,[50] it remains difficult to definitively answer whether the use of protected characteristics in risk classification is acceptable or not.

Due to the inconsistency in approaches, it has been suggested that a more balanced test for acceptable forms of discrimination should be adopted. In her paper, Lima Rego proposes a twofold test for the separation of admissible factors of differentiation from inadmissible forms of discrimination.[51] The test consists of two questions. The first asks whether the actuarial data on the protected characteristic as a risk factor has an explanation which is unrelated to some form of past discrimination. To pass this question, the use of the protected characteristic would need to obtain its predictive power from reasons unrelated to the historical treatment or social standing of the protected group. Instead, the protected characteristic would need have a causal connection to the risk. The second question asks whether any other factors would instead be suitable in predicting the risk. Where there is a potential risk factor which is more suitable or even equally or slightly less suitable and would be less burdensome on human rights, then use of the protected characteristic would not be justified.

---

[45] Equality Act 2010, s 29.
[46] Equality Act 2010, sch 3 para 20A(2).
[47] Equality Act 2010, sch 3 para 21(1).
[48] Phillipa Watson, 'Equality, Fundamental Rights and the Limits of Legislative Discretion: Comments on Test-Achats' (2011) 36(6) European Law Review 896, 904; Margarida Lima Rego, 'Statistics for the basis for discrimination in insurance business' (2015) 14 Law, Probability and Risk 119, 134.
[49] Lima Rego (n 8) 8.
[50] It has been suggested that age may an acceptable ground for differentiation as it is mutable: all individuals have the potential to be advantaged and disadvantaged by its use throughout their lives. *See* Opinion of Advocate General Kokott delivered on 30 September 2010, *Association Belge des Consommateurs Test-Achats ASBL and Others v Conseil des ministers,* C-236/09, EU:C:2010:564, para [37]. The same however cannot be said for disability, which often raises the same concerns about control as sex and race.
[51] Lima Rego (n 48) 127; Lima Rego (n 8) 13.

Lima Rego's approach, whilst falling short of an outright ban, addresses the main concerns about the use of sex and other protected advances by those opposed to discrimination. The first limb of the test prevents insurance from being used to perpetuate past injustices and stereotypes by requiring a causal connection with the risk. The second limb addresses the use of protected characteristics as surrogate factors merely for commercial convenience and economic efficiency. Lima Rego anticipates that in most instances, the use of sex for risk classification would fail her twofold test and would do little to change the position as it currently stands. As such, sex would remain to be an unacceptable risk factor. However, the twofold test would allow for limited instances where the use of sex advantage both the insurer and insured. Lima Rego theorises that the use of sex in life insurance may pass the test. On the first question, the use of sex would need to be unrelated to any past unfair discrimination. Lima Rego explains, "everywhere in the world and across time", women have had longer life expectancies than men.[52] The differences in life expectancies can be justified by both sociological and biological reasons. Of course, where the differentiation is based on social factors, the use of sex as a risk factor will be prohibited. If, however, data progresses to a point where the differentiation can be explained by biological factors alone, sex has a chance of proceeding to the second question.

Sex would only proceed to the second question where it is directly predictive of the loss, ie, where it is the cause of women living longer lives. The existence of the causal connection would indicate that there would unlikely be a risk factor which would replace the value that sex would bring to the risk classification completely unless it was simply a proxy for sex. Yet, it should be noted that to pass Lima Rego's second question, sex should not carry a disproportionate weight in the process, mirroring the jurisprudence of the CJEU in *X v Finland*.[53]

Lima Rego's test represents a more balanced approach to statistical discrimination, providing clear guidelines on when the use of protected characteristics would be fair. The practical impact of the test compared to the current legal approaches would seldom be different. In most cases the protected characteristics should not be used. Yet the approach allows for some limited applications of the protected characteristics in risk classification, allowing the advantages of "actuarial fairness" to be reaped whilst individual's freedoms remain protected. As shall be discussed in section four, an approach like Lima Rego's may be what is required as big data changes the landscape of differentiation.

3. The impact of big data and AI on discrimination in insurance: the emergence of unintentional proxy discrimination

Within her presentation, Lima Rego suggested that the rise of big data and AI will alter the way in which we think about discrimination in insurance.[54] Instead of direct statistical discrimination which has dominated the debate, new forms of indirect discrimination which are extremely hard to tackle will emerge. The rise of insurance-focused AI is promising to transform the risk classification process in insurance, guaranteeing increased accuracy and efficiency in underwriting and pricing.[55] This potential comes largely from big data and machine-learning algorithms, which will eventually lead to the automation of the risk classification process. Big data refers to "high-

---

[52] Lima Rego (n 8) 13.
[53] Judgment of the Court (Second Chamber) of 3 September 2014, *X,* C- 318/13, EU:C:2014:2133.
[54] Lima Rego (n 8) 20.
[55] Insurance Europe, 'Q&A on the use of big data in insurance' (2019) 1.

volume, high-velocity and/or high-variety information assets that enable enhanced insight, decision-making, and process automation".[56] It is data that was previously too large and too complex to be managed by humans or traditional data processing software.[57] Big data can be used to fuel machine-learning algorithms, which are algorithms which 'learn' from its experience and improve automatically. Based on the big data it has been fed, machine learning algorithms are able to identify patterns without any specific programming to do so.[58]

Through automating the risk classification process, big data and machine-learning algorithms allow insurers to tap into previously unobtainable data and uncover new risk related insights.[59] Insurers can use highly complex machine-learning algorithms to sift through big data from varying sources to find correlates to risk. Programmed to find the most efficient and effective patterns between data and risk and armed with big data, machine-learning algorithms have the capability to identify subtle patterns which are not identifiable by humans.[60] For instance, when seeking factors which correlate with property insurance risks, the algorithm may find a link with the risk and the frequency an individual tweets, what pages they like on Facebook, their Google searches, or online shopping habits.[61] One unnamed insurer has even admitted that their algorithm has identified a correlation between drinking tap or bottled water and motor risks.[62]

Machine learning algorithms can learn and improve their efficiency as it obtains new data and experience.[63] Algorithms will continuously seek the best predictors of risk, meaning that the factors it uses and the weight it attaches to them will constantly adapt.[64] Because of this, the predictive power of the risk factors the algorithm identifies can be quite substantial. With the automation of risk classification, insurers can use thousands of these correlates for risk classification and the underwriting of risk.[65] Combined, insurers are able to obtain a much more comprehensive view of the insured risk. The need to rely on large pools in risk classification is lessened and instead they can offer more individualisation in the underwriting and pricing process.[66] This is bolstered by the increasing use of Internet of Things (IoT) devices which are being used by insurers in personal insurance to obtain real-time information about the insured's behaviour. IoT devices are "nonstandard computing devices that connect

---

[56] Gartner, 'Big Data' (*Gartner*) <www.gartner.com/en/information-technology/glossary/big-data> accessed 30 December 2021.
[57] Oracle, 'What is Big Data?' (*Oracle*) <www.oracle.com/uk/big-data/what-is-big-data/> accessed 29 December 2021.
[58] Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104(3) CLR 671, 674.
[59] Duncan Minty, 'Ethics, Data and Insurance: 4 Developments Worth Watching' (*NFT,* April 2017) < nft.nu/en/node/2157> accessed 05 December 2021.
[60] Jane Bambauer and Tal Zarsky, 'The Algorithm Game' (2018) 94 Notre Dame L. Rev. 1, 1; Vincent Reillon, 'Understanding Artificial Intelligence, (European Parliamentary Research Service 2018) 1.
[61] Esko Kivisaari, 'Big Data is Coming Are you Ready? Insurance Principle and Actuaries in the Age of Fintech' The European Actuary (October 2017) 6; Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI' (*University of Oxford Faculty of Law*, 09 October 2018) <www.law.ox.ac.uk/business-law-blog/blog/2018/10/right-reasonable-inferences-re-thinking-data-protection-law-age-big> accessed 10 December 2021; Swedloff (n 23) 2057.
[62] Duncan Minty, 'Ethics, Data and Insurance: 4 Developments Worth Watching' (*NFT,* April 2017) <nft.nu/en/node/2157> accessed 05 December 2021.
[63] Harry Surden, 'Machine Learning and Law' (2014) 89 Wash. L. Rev. 87, 88.
[64] Swedloff (n 23), 2057.
[65] Lemonade, an insurance company powered by AI and behavioural economics, have claimed to consider 100 times the typical 20-40 data points in their process of 'precision underwriting.' Daniel Schrieber, 'Precision Underwriting' (*Lemonade*, 3 April 2018) <www.lemonade.com/blog/precision-underwriting/> accessed 31 December 2021.
[66] Laurence Barry and Arthur Charpentier, 'Personalisation as a promise: Can Big Data change the practice of insurance?' (2020) Big Data & Society 1, 2.

wirelessly to a network and have the ability to transmit data" to the insurance provider. The potential value of this technology to insurers is ample. Health insurers may be able to obtain personalised health information, such as heart rate, calories burnt and sleep patterns, from the insured's use of a smart watch. Motor insurers may be able to use IoT sensors or smart phones to obtain data on the insured's driving behaviour and their use of the vehicle.[67] The premiums charged to individuals may therefore reflect the risk they present with more precision than with traditional risk classification. The benefits of this are numerous. Insurers may achieve better economic efficiency by avoiding adverse selection and moral hazard whilst remaining competitive on the market. The individualisation of pricing results in individuals once again being charged a premium which more precisely reflects their risk which may have been surrendered in the pursuit of equality and anti-discrimination.

On the fact of it, the use of big data and AI in risk classification represents a step towards the more objective, "mechanical" assessment of an individual's risk.[68] The amount of big data now available to insurers to use in pricing possibly represents a fading desire to use protected characteristics for their availability. Where, instead, an insurer knows of the individual's "shopping habits for ten years, what this person buys at the pharmacy, how often this person goes to the gym etc.", as well as personal data from IoT devices, "one could... say it is unnecessary to know the age or gender of this person".[69] Even if sex and other characteristics remain to be relevant, Lima Rego points out that their use would be just one factor amongst thousands of others, reducing the reliance and weight insurers place on their use.[70] In short, the use of big data and AI could reduce the commercial desirability of protected characteristics as risk factors. This is seemingly a win-win situation. Insurers can obtain the kind of accuracy and precision in risk classification which they argued was only achievable by using protected characteristics. Meanwhile, algorithms can be denied data on protected characteristics and can be programmed to them from its analysis to remove the possibility of direct discrimination.

Yet, the use of big data and AI nevertheless creates worrying concerns about the risk factors which the algorithm identifies as being correlated to the risk and how their use may disparately impact protected groups. Moving away from the direct discrimination which is at the centre of traditional debates about risk classification, the focus switches to facially neutral factors which correlate with protected characteristics. This is a form of indirect discrimination known as proxy discrimination and is inevitable where big data and machine-learning algorithms are used.

3.1 Unintentional Proxy Discrimination

In contrast to direct discrimination which focuses on the practice itself, indirect discrimination focuses on the effects of practices. It occurs where a facially neutral practice negatively impacts a protected group more than others and is also prohibited in most circumstances within the UK and EU. Proxy discrimination is a particular subset of indirect discrimination, which has been well documented in the provision of financial services. It refers

[67] Simon Behm, Ulrike Deetjen, Sanjay Kaniyar, Nadine Methner, and Björn Münstermann, 'Digital ecosystems for insurers: Opportunities through the Internet of Things' (*McKinsey & Company,* 4 February 2019) <www.mckinsey.com/industries/financial-services/our-insights/digital-ecosystems-for-insurers-opportunities-through-the-internet-of-things> accessed 28 December 2021.
[68] Lima Rego (n 8) 20.
[69] Kivisaari (n 61) 5.
[70] Lima Rego (n 8) 20.

to the use of a factor or characteristic within decision-making *because* that factor correlates with membership of a protected class. Intentional proxy discrimination has been a well-documented concern within the provision of financial services.[71] Perhaps the most infamous example is that of "redlining", which describes the denial of services by US financial providers to individuals in particular geographic areas where those geographic areas correlated with race within the civil rights era. Big data and AI raise a new issue: *unintentional* proxy discrimination. Machine learning algorithms will simply have been programmed to minimise future claims.[72] To do so, it will seek out any factor which may correlate with risk to be used in the risk classification process, regardless of the nature of the correlation.[73] An algorithm which has been programmed to exclude sex and denied data on sex will therefore locate factors which correlated with sex where sex can help minimise future claims.[74] Such factors will obtain their predictive power *because* they correlate with sex. This is to be distinguished from factors which correlate with sex but are predictive in their own right, such as the type of car an individual drives.[75] Only the former will constitute proxy discrimination.

The impact that proxy discrimination can have on protected groups is starkly illustrated by Amazon's failed recruiting algorithm which was scrapped in 2015 after showing bias towards women.[76] The algorithm had been trained using resumes submitted to the company over a 10-year period with the hope that it could identify patterns for successful candidates. The algorithm, identifying that more men than women were hired in the male-dominated industry, taught itself that male candidates were preferable in the recruiting process.[77] On this basis, the algorithm rejected resumes from women, despite sex and obvious proxies (such as educational institution) being removed from the data. Instead, it used the language choices in the resume to differentiate between resumes from male and female candidates. A similar situation was reported to have occurred within the insurance industry itself in 2018, when a mystery shopping investigation by a British newspaper had found that insurers have given higher premium quotes to motorists named Mohammed compared with motorists named John, suggesting that their pricing algorithms contained racial biases.[78]

Proxy discrimination is most likely to occur in two scenarios. The first is to do with the 'cleanliness' of the data on which machine-learning algorithms are trained, relating back to the concerns that risk classification can perpetuate past injustices discussed in part 1.1. Whilst algorithms are not capable of being *intentionally* biased, bias can creep into algorithms from training data. Data can include biased human decisions (such as biases in the

[71] Larry Alexander and Kevin Cole, 'Discrimination by Proxy' (1997) 14 Const Comment 453.
[72] Anya E.R. Prince and Daniel Schwarz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2020) 105 Iowa Law Review 1257, 1262.
[73] Ibid. 1275.
[74] Ibid. 1273.
[75] Barocas and Selbst (n 58) 691.
[76] Jeffrey Dastin, 'Amazon scarps secret AI recruiting tool that showed bias against women' (*Reuters,* 11 October 2018) <www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G> accessed 21 December 2021.
[77] i*bid*.
[78] Ben Leo, 'Mo Compare: Motorists fork out £1000 more to insure their cars if their name is Mohammed' (*The Sun,* 22 January 2018) <www.thesun.co.uk/motors/5393978/insurance-race-row-john-mohammed/> accessed 21 November 2021; see Duncan Minty, 'The Ethical Problems at the root of the Mohammed and John Quotations' Ethics and Insurance (20 February 2018) <ethicsandinsurance.info/2018/02/20/mohammed-john-quotations/> accessed 02 January 2022; Centre for Data Ethics and Innovation, 'AI and Personal Insurance' (September 2019).

previous underwriting) or may reflect historic social inequity.[79] For example, the Amazon recruiting algorithm contained biases because of the previous recruitment data. A similar problem was reported in 1988, where an algorithm designed to sort medical school applicants based on previous admission decisions systematically disfavoured women and racial minorities.[80] An algorithm will take the training data as truth.[81] Where the training data contains biases, it will take those biases to be true and implemented them into the system. Where the algorithm believes that sex says something about the risk, it will seek out correlates for sex where it is prohibited from using sex directly. Left untouched, biases can be baked into the algorithm which perpetuate the past injustices and may even identify predictors which are not necessarily correct.[82] To ensure that an algorithm does not indirectly discriminate based on historic inequity, insurers should ensure that the algorithm is fed up-to-date and objective data which has been cleared of any biases.

The second scenario can occur even where the algorithm has been trained with data free from any latent biases. This is where the protected characteristic is *directly predictive* of the insured loss. Prince and Schwarz explain, "proxy discrimination by AIs is most likely to occur when prohibited traits are directly predictive of legitimate outcomes in ways that cannot be more directly captured by alternative data".[83] Where the protected characteristic's power to predict a 'target variable' cannot be replicated by facially neutral data, the algorithm will always seek to find close correlates to the protected characteristic to reproduce the predictive value. This is known as omitted variable bias.[84] A protected characteristic may be directly predictive of loss where: (1) it has a causal connection with the desired outcome; or (2) the protected characteristic is correlated with the desired outcome, but "its predictive character is not mediated through a presently quantifiable available variable".[85] A causal connection will occur where the protected characteristic is causally linked with the insured risk. Sex will be causally linked with a risk where, for example, the likelihood that it will occur is due to biological reasons. Where a causal link is present it will "always impact the probability of the targeted outcome in a statistical model, irrespective of any additional information that could be added to the model".[86] Where the protected characteristic is causally linked to the loss, there will be no other known factor which would have been as equally suitable as a predictor of the relevant outcome, nor will there ever be. As such, an algorithm will always seek out proxies for the protected characteristic where it is denied information on the characteristic. The instances where protected characteristics are causally linked to risks will be few and far between but, nevertheless, when the link exists, proxy discrimination is inevitable.

[79] A similar and related concern is the type of data that is used to train algorithms. Big data has opened the door to insurers to use a data from more diverse sources, such as social media or smart phones. These diverse data sources may be representative of all members of society, leading to some groups being underrepresented or overlooked. Where the data on these groups is less or more extensive, it may lead to them being disparately impacted. See Kate Crawford, 'Think Again: Big Data' (*Foreign Policy,* 10 May 2013) < foreignpolicy.com/2013/05/10/think-again-big-data/> accessed 19 December 2021.
[80] Stella Lowry & Gordon Macpherson, 'A blot on the profession' (1988) 296 Brit. Med. J 657.
[81] Barocas and Selbst (n 58) 682.
[82] Betsy Anne Williams, Catherine F Brooks and Yotam Shmargad, 'How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications' (2018) 8 Journal of Information Policy 78, 80.
[83] Prince and Schwarz (n 72) 1265.
[84] Devin G Pope and Justin R Sydnor, 'Implementing Antidiscrimination Policies in Statistical Profiling Models' (2011) 3(3) American Economic Journal: Economic Policy 206, 210.
[85] Prince and Schwarz (n 72) 1278.
[86] i*bid*, 1277.

The second instance in which a protected characteristic may be directly predictive refers to the use as a surrogate factor for unattainable and unquantifiable variables. This is known as opaque proxy discrimination.[87] Opaque proxy discrimination can occur in two scenarios. The first is where there may be a causative factor for which the protected characteristic is proxying which cannot be quantified because it is not fully understood. In other words, the protected characteristic may itself be causative or may be merely proxying a causative factor.[88] The second scenario is where the protected characteristic is a surrogate variable for the true causative variable which is fully understood but is nonetheless difficult to quantify. For instance, the care a motorist takes when driving.[89] Opaque proxy discrimination does not hold the same degree of perpetuity as causal proxy discrimination. As the development in technology advances our scientific understanding of certain risks and allows insurers access to previously unavailable information, it is possible that opaque proxy discrimination will eventually cease to exist. Instead, the use of protected characteristics as valuable surrogate factors will be replaced by neutral causative variables. Indeed, the insurance industry is already seeing technological advances which allow them to quantify factors which were previously difficult, if not impossible, to measure directly. For instance, the use of IoT devices now allows insurers to receive insights about a driver's care levels.

Where the protected characteristic is directly predictive, algorithms will seek out the closest possible proxies for the protected characteristic. Driven to make the most efficient decisions, it will likely identify a cluster of factors which taken together correlated very closely with the protected characteristic. [90] For instance, the algorithm may find that the likes of a certain combination of Facebook pages may be a correlate for sex or race, or the frequency of Google searches may correlate with religion, or even the pattern of their mouse movements when filling out an online policy form.[91] This raises many of the same concerns considered in the debate over direct statistical discrimination. For insurers, this once again represents a form of rational, objective, and scientific differentiation required to charged premiums which reflects an individual's risk. Indeed, big data and predictive algorithms allows insurers to rely less on large pools for risk classification and instead move towards more individualised pricing.[92] Yet, it is undeniable that proxy discrimination causes the same harmful impact on protected groups and remains to interfere with their freedoms, liberties, and dignity. Indeed, as discussed above, algorithms can internalise biases and eternalise past injustice. They are only capable of identifying correlates and do not find factors based on whether they are causally linked.[93]

The use of big data and algorithms do, however, add extra complexity to the debate and create concerns for advocates of anti-discrimination. These concerns largely emerge from the information asymmetry between humans and algorithms. The value of machine-learning algorithms comes from their ability to identify subtle

[87] i*bid*, 1278.
[88] Prince and Schwarz use the genetic context as an example: 'even for many pathogenic genetic variants, it is often unknown why a particular sequence in a gene leads to increased risk. It may well be that one gene has been identified as higher risk because it is correlated with some other more particular DNA segment that has yet it be identified and characterised. Alternatively, it may be that the true causative mechanism is epigenetic changes that turn on and off the gene in question.' Anya ER Prince and Daniel Schwarz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data' (2020) 105 Iowa Law Review 1257, 1278.
[89] i*bid,* 1279.
[90] Minty (n 78)
[91] Wachter and Mittelstadt (n 61); Swedloff (n 19) 362.
[92] Barry and Charpentier (n 66) 2.
[93] Ryan Calo, 'Artificial Intelligence Policy: A Primer and Roadmap' (2017) 51 U.C.D. L. Rev. 399, 415.

patterns at a speed and in an *unsupervised* manner.[94] Machine-learning algorithms will therefore identify patterns which are completely unforeseeable and unpredictable by humans.[95] What algorithms predict is tremendous: *Microsoft* can predict Parkinson's disease Alzheimer's disease from search engines interactions;[96] *Google* can predict potential flu outbreaks.[97] But these links cannot be predicted or anticipated by humans, whether the insurer or the insured. Thus, when the algorithm identifies a cluster of factors because they correlated with a protected characteristic, it is likely that a human will not appreciate that the factors correlate with the protected characteristic any more than they might anticipate that the type of water one drinks correlates with their motor risk.[98] Even where an insurer is completely transparent about the factors, the presence of proxy discrimination will be difficult to detect and explain.

Big data and AI have the effect of moving the discrimination debate away from direct and intentional statistical discrimination towards unintentional and unidentifiable forms of proxy discrimination. Whilst the effects of the two types of discrimination are broadly the same in that protected groups may be charged different premiums or receive fewer benefits; proxy discrimination will be significantly harder to police than direct discrimination. It will not simply be a matter of just prohibiting the use of protected characteristics and obvious proxies but will require close oversight of the algorithms to ensure that the underlying motivation for risk factor choice is not related to protected characteristics.

3.2 Legal Protection from Unintentional Proxy Discrimination

The nature of proxy discrimination may mean that the legal protection which is currently in place within the EU and UK is insufficient to protect protected classes from its effects. Proxy discrimination by algorithms is a kind of unintentional, indirect discrimination. Indirect discrimination refers to practices which are facially neutral, but which have a disparate impact on a protected group. It is currently prohibited in the EU and UK.[99] Generally speaking, where an individual wants to bring a legal claim against indirect discrimination, the law requires three elements to be shown. First, there must be facially equal treatment. Second, the practice must have a disproportionately adverse impact on a protected group. Third, and perhaps most significantly, there must be an absence of an *acceptable justification*.[100] The following will describe the obstacles that must be overcome to demonstrate proxy discrimination by machine-learning algorithms under current principles of indirect discrimination law.

The first obstacle that must be overcome is detecting that proxy discrimination has taken place. The rise of big data and AI has widened the information gap between insurers and insured more than ever before. Under

---

[94] Swedloff (n 23) 2056.
[95] Karni Chagal-Feferkorn, 'The Reasonable Algorithm' (2018) U Ill JL Tech & Pol'y 111, 133.
[96] Liron Allerhand, David Arkadir, Brit Youngmann and Elad Yom-Tov, 'Detecting Parkinson's Disease from interactions with a search engine: is expert knowledge sufficient?' (2018) CIKM '18: Proceedings of the 27th ACM International Conference on Information and Knowledge Management 1539.
[97] Donald R Olson, Kevin J Konty, Marc Paladini, Cecile Viboud and Lone Simonsen, 'Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales' (2013) 9(10) PLoS Comput Biol <journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003256> accessed 01 January 2022.
[98] Minty (n 59).
[99] Equality Act 2010, s 19.
[100] Sandra Fredman, *Discrimination Law* (OUP 2011) 179.

traditional underwriting methods, insurers obtained most of the information they needed to know about the risk from the insured themselves. Insureds could thus broadly tell what factors were relevant to risk classification and underwriting. They knew that the type of car they drove, their annual mileage, where they parked their car at night, their occupation were all factors which were risk related and influenced premiums they were charged.[101] The rise of big data means that insurers are increasingly obtaining information about insureds from third party sources. Aviva, for example, have in recent years launched their "Ask it Never" initiative, designed to substantially reduce the number of questions posed to insureds at the underwriting stage.[102] Instead, the insurance company "has a sophisticated system of third-party data collection running in the background", where it can get the majority of the information required to underwrite the risk.[103] Whilst asking the insured fewer questions, insurers will actually be taking into account *more* risk factors at the underwriting stage. Big data is therefore lessening the involvement that an insured has in the underwriting of their own insurance. Whilst under traditional underwriting, insureds understood what went into the pricing of their insurance, big data and algorithms have enlarged the gap between insureds and their insurer. Put in this removed position, it is difficult for the average insured to identify what is discrimination and what is not. This will likely be the case even if insurers are completely transparent about the risk factors which contribute to the pricing of premiums. Because insurers will be able to consider so many data points and algorithms will identify obscure correlations to risk not identifiable by human, it will be hard for insureds to identify any sort of link between the use of certain factors and their membership to a protected class. The lack of ability to recognise proxy discrimination will also be held by insurers themselves. As service providers, insurers have a duty to not discriminate against an individual requiring their service.[104] Minty has noted that there is an increasing impression within the insurance industry that underwriters no longer understand how pricing is calculated.[105] Armed with the same inability to recognise obscure correlates to protected characteristics as consumers (albeit with more technological knowledge), insurers will also be unable to recognise where the algorithm is using a risk factor because of its link to the protected characteristic. This gap of understanding between humans and algorithms makes proxy discrimination extremely difficult to detect and to police.

Where proxy discrimination has been detected, an insured may be able to bring a claim if they can show a facially neutral practice has a disproportionately adverse impact on their protected group. In other words, they would need to show that the use of a factor or cluster of factors places their group at a disadvantage, whether it be because the insurer charges different premiums, offers different benefits, or denies coverage altogether. At this stage, the insured would not need to demonstrate that the factor or group of factors are necessarily proxies for the protected characteristic but just that it causes a disparate impact. To be successful here, it would be best if the claimant knows which factors are the source of the differentiation. The disadvantage should not be too difficult to demonstrate using statistical evidence, given that statistical correlations found by the algorithm will be the reason for the differentiation.[106] Take, for example, the use of a factor such as "likes" of the specific Facebook page as a proxy for sex. To show that the use of that factor causes a particular disadvantage, all that would need to be shown

---

[101] Duncan Minty, 'Social Sorting: Could it be the stuff of nightmares for insurance firms' Ethics and Insurance (9 September 2014) <ethicsandinsurance.info/2014/09/09/social-sorting/> accessed 02 January 2022.
[102] Centre for Data Ethics and Innovation, 'AI and Personal Insurance' (September 2019) 12.
[103] i*bid,* 12.
[104] Equality Act 2010, s 29.
[105] Minty (n 78).
[106] Anne Beale, 'Core Rights and Duties' in Anthony Robinson, David Ruebain and Susie Uppal (eds), *Blackstone's Guide to the Equality Act 2010* (4th edn, OUP 2021) 46.

was that women are statistically more likely to "like" the Facebook page than men.[107] Even for more obscure factors, such the type of water one drinks, a particular disadvantage can be shown purely by using statistics on which the correlation is based.

Significantly, where the claimant has demonstrated that there is a particular disadvantage, this will not be enough to qualify indirect discrimination where the insurer has an acceptable justification. In contrast to direct discrimination, indirect discrimination can be justified where it is a 'proportionate means to achieving a legitimate aim'.[108] A proxy factor will therefore not be indirectly discriminatory where it achieves a legitimate aim. Where the factor in question is genuinely predictive but also happens to serve as a proxy for a protected class, it is likely that insurers could easily argue that it achieves a legitimate aim, that is, predicting the risk and pricing premiums accordingly. There may, however, be trouble in proving how the factor achieves that aim. To do so, insurers would have to demonstrate what the factor's predictive value is. Again, the human's incapability to understand algorithmic correlates could be an issue here. Most algorithms are not able to explain why they make the decisions they do, though there is a computer science movement that is working on developing algorithms which can explain themselves.[109] Even if there is a way to explain why an algorithm uses a factor to predict risk, the fast-moving nature of machine-learning algorithms can alter the weight and significance a factor is given from one moment to the next based on new data it is fed. Insurers would need to ensure that the evidence they provide to show the justification is relevant to the allegation of discrimination at hand and that it was proportionate at the time when the alleged discrimination occurred. Because machine-learning algorithms alter as they learn, insurers will have to capture the history of the algorithm. This could be quite a significant administrative burden.

Where it is known that the factor(s) in question gains its predictive power from its correlation with a protected characteristic, insurers will have some trouble in showing that it is a proportionate means of achieving a legitimate aim. This may bring up the same arguments used to support risk classification and actuarial fairness. The courts have been clear that anyone seeking to justify a discriminatory practice cannot rely purely on cost considerations.[110] However, insurers have maintained that one of the main benefits of the use of protected characteristics within risk classification is the advancement of *fairness*. It is possible, then, that insurers will advance the argument that the use of the proxy factor is necessary to charge premiums which reflect an individual's risk, and to promote equal treatment for equal risks. The courts have recently affirmed that cost considerations can be balanced with other justifications in a 'cost-plus' approach.[111] Whether the courts would be convinced that the use of proxy factors for protected characteristics is a proportionate means of achieving a legitimate aim, however, would remain to be seen.

---

[107] Statistics are a powerful tool to demonstrate that one group suffers a particular disadvantage. See Sandra Fredman, *Discrimination Law* (OUP 2011) 183-190.

[108] Equality Act 2010, s 19(2)(d).

[109] Cary Coglianese and David Lehr, 'Transparency and Algorithmic Governance' [2019] Administrative Law Review 1, 18; MIT Computer Science & Artificial Intelligence Lab, 'Making computers explain themselves' (*MIT Computer Science & Artificial Intelligence Lab,* 27 October 2016) <www.csail.mit.edu/news/making-computers-explain-themselves> accessed 20 December 2021.

[110] Judgment of the Court (sixth chamber) of 24 February 1994, *De Weerd, nee Roks v Bestuur van de Bedrijfsvereniging voor de Gezondheid,* C-343/92, EU: C:1994:71; *Ministry of Justice v O'Brien* [2013] UKSC 6.

[111] *Heskett v Secretary of State for Justice* [2020] EWCA Civ 1487.

The law on indirect discrimination therefore does prohibit proxy discrimination, though does little to prevent its unintentional and undetectable nature. In the rare event where proxy discrimination is likely to be recognised by humans, it is not certain whether the law will operate in the way it should do to award the claimant justice.

4. An alternative approach to discrimination? Professor Lima Rego's twofold approach

Discrimination law as it currently stands is likely to see some difficulties in managing proxy discrimination by machine-learning algorithms. Firstly, the ban on the use of protected characteristics even where they are directly predictive makes the use of difficult to detect proxies inevitable when machine-learning algorithms are used in the underwriting process. Secondly, because of growing information gaps between humans and algorithms, indirect discrimination is difficult to detect and to prove. In this context, a more balanced approach to discrimination proposed by Lima Rego may be more appropriate in regulating this kind of proxy discrimination.

It has been argued by Williams, Brooks and Shmargad that collecting and accounting for the use of data on protected characteristics "can lessen discrimination, though this idea may seem counterintuitive".[112] By allowing the direct use of protected characteristics, transparency will increase, and it will be easier to monitor how they are being used. Being able to use directly predictive protected characteristics directly, algorithms may no longer have as large a need to seek out proxies for protected characteristics. As proxy discrimination lessens, as do concerns about the undetectable nature in proxy discrimination. This provides an interesting contrast to the previous debate, where those opposed to discrimination advanced the view that the use of protected characteristics in risk classification would obstruct fairness. Instead, the (limited) use of protected characteristics could be a way of *promoting* fairness and protected groups from undetectable proxy discrimination.

Balance would be needed. Unlimited use of protected characteristics in risk classification can be an unproportionate interference with an individual's liberty and should always be approached with care. Lima Rego has proposed a twofold test distinguishing between admissible and inadmissible forms of differentiation, adopting a more considered approach to the question of differentiation.[113] Though proposed as a response to the debate on direct statistical discrimination, the test provides a good starting point to answering whether a more balanced approach to direct statistical discrimination could be useful in dealing with discrimination by algorithms.

Lima Rego's first question asks whether the link between the protected characteristic and the risk can be associated to past discrimination. This would require the presence of a causal connection. Machine-learning algorithms are exceptional at finding *correlations* to risk which in most cases will contain no causal connection. Straight away, the first question would prohibit most instances of the use of sex or another protected characteristic by algorithms. Significantly, however, it would allow algorithms to use sex where it is known that it is a cause of the loss. It has been referred to above that where the protected characteristic is a cause of the loss, it is *directly predictive* in a way that proxy discrimination will be inevitable.[114] Allowing the direct use of protected characteristics where they have a causal connection to the risk could therefore lessen the risk of proxy discrimination. In this sense, justified direct discrimination may be preferable to proxy discrimination because it is more transparent and controlled.

[112] Williams et al (n 82) 87.
[113] Lima Rego (n 48) 127; Lima Rego (n 8) 13.
[114] Prince and Schwarz (n 72) 1273.

Problems may emerge, however, with the practicality of this. Opaque proxy discrimination remains a risk where the protected characteristic has predictive value arising from its correlation with a causal variable that cannot easily be quantified. Whilst it is recognised that these situations will be lessened as our understanding of technology and science increases, proxy discrimination will continue to occur until we get to this point. In addition to this, it is possible that an algorithm will find a correlate to loss which will eventually be found to be causal when our knowledge advances. When the causal connection has been realised, the use of the protected characteristic will be allowed under Lima Rego's first question. However, up until that point, it would have been refused, leading to the undetectable proxy discrimination. By allowing the direct use of protected characteristics to differentiate, the risk of causal proxy discrimination will reduce. Only opaque proxy discrimination will remain an issue.

Lima Rego's second question focuses on whether another factor can be used to predict the risk which would be less burdensome on the rights of the insured individual. This is designed to prevent the use of protected characteristics as surrogate factors for less accessible information. Big data and IoT devices should mean that eventually the need to use protected characteristics as surrogate factors will be diminished (except where the causal variables cannot be quantified). Determining whether other suitable factors exist should be an easy task when using an algorithm. Simply training the algorithm to not use the protected characteristic will show the availability of other factors which can be used to predict the risk. Yet, this once again could cause issues. Specifically, it will be difficult to determine whether these characteristics are suitable alternatives or are merely proxies for the protected characteristic. Given the seemingly arbitrariness of the proxies, it will be hard for a human insurer to tell, once again leading to the risk that proxy discrimination will go under the radar.

Nevertheless, Lima Rego's test highlights that a more balanced approach towards statistical discrimination in insurance may alleviate some of the problems that may be faced as insurers use big data and AI in underwriting.

5. Conclusion

The emergence of unintentional and unobservable proxy discrimination has the power to change the perceptions of fairness within the insurance industry. Currently, the law on statistical discrimination signals that the use of protected characteristic for differentiation will amount to an unacceptable interference with individual's freedoms, except in the case of age or disability. Proxy discrimination raises questions as to whether this absolute approach guarantees that protected groups will be treated fairly. Specifically, the ban on use of causal protected characteristics contributes to the inevitably that proxy discrimination will occur. For this reason, a more balanced approach like Lima Rego's twofold test may be necessary to guarantee respect for human dignity and the functioning of insurance business.

Yet, causal proxy discrimination represents only one example of the kinds of proxy discrimination which can occur. Even with a more balanced approach, difficult to detect discrimination can nevertheless occur. To ensure the responsible use of big data and AI in risk classification, insurers, regulators, and the legislature alike are likely to encounter an uphill battle in the quest for equality.