UNDERSTANDING DISCRIMINATION IN THE SCORED SOCIETY

Tal Z. Zarsky*

INTRODUCTION: KEEPING SCORE OF THE SCORED SOCIETY

In *The Scored Society*,[1] Professors Danielle Citron and Frank Pasquale introduce a troubling reality: a society in which a small yet powerful group of individuals makes crucial decisions regarding a broad segment of the public. Such decisions are reached on the basis of a scoring scheme the group members secretly develop in advance. This seemingly arbitrary and possibly automated scoring process[2] allows powerful entities[3] to quickly and seamlessly promote their objectives while treating similar individuals differently. The scoring process also unfolds in a manner which is incomprehensible to those whom it affects. The description above is not of a distant dystopia. Rather, it is of a society we are quickly entering—or perhaps, as a technology commentator recently noted in a similar context—"sleepwalking into."[4]

Scoring has been carried out for years in the realm of consumer

* Professor of Law, University of Haifa, Faculty of Law. I thank Courtney Bowman, James Rule, Susan Freiwald, Mireille Hildebrandt, Joris van Hoboken, Serge Gutwirth, and Florencia Marotta-Wurgler for comments on previous versions of this work. Some of the ideas here discussed were presented at the Privacy Legal Scholars Conference, 2011, a Faculty Workshop at IVIR (University of Amsterdam) (September, 2011), and the LSTS Workshop at the Vrije University in Brussels (September, 2011). In all these instances, participants provided excellent comments. I also thank Eyal Mashbetz for his research assistance. Research for this paper was partially funded by an NWO (the Dutch Research Foundation) grant titled: "Data Mining without Discrimination" and I thank my co-researchers Bart Custers, Bart Schermer, and Toon Calders for their insights.

1. Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014).

2. For this Article, I rely on Citron and Pasquale's description of the process, as follows: "Predictive algorithms mine personal information to make guesses about individuals' likely actions and risks. A person's on- and offline activities are turned into scores that rate them above or below others. Private and public entities rely on predictive algorithmic assessments to make important decisions about individuals." *Id.* at 3.

3. *Id.* at 3–4 (noting the discussion in terms of "power"). The article also notes "mutual-scoring opportunities" and instances in which "individuals can score the scorers" but note that more often the situation is reversed. *Id.* at 3.

4. Evgeny Morozov, *The Real Privacy Problem*, 116 MIT TECH. REV. 32, 37 (2014), *available at* http://www.technologyreview.com/featuredstory/520426/the-real-privacy-problem/.

credit. Yet the age of big data is leading to the dissemination of these practices to many other contexts. The scoring practices are rendered a feasible option in business, governmental, and social settings due to a multitude of effects.[5] Scoring is made possible given the availability of vast quantities of personal information, collection of which is enabled by changes in business models, and the ever-lower price of digital storage facilities. It is further enhanced by advances in data analytics and the ability to effectively aggregate data.

The accelerating use of scoring brings about a variety of problems, which Professors Citron and Pasquale elegantly and eloquently detail. As they demonstrate, the problems related to the expansion of the scored society can be understood on a basic and deeper, analytical level. On the basic level, these processes are problematic given the vast amount of complaints they generate,[6] and the fact that they are used, at times, to the detriment of the individual.[7] Here, Professors Citron and Pasquale refer to use of credit scores to allocate loans as a case in point.[8] On an analytical level, the nature of these concerns could be linked to the way the process relies on biased and inaccurate datasets,[9] its inherent opacity,[10] or the lack of sufficient human review.[11]

Yet another important concern surfacing in discussions and analyses of the scored society and the troubles big data analytics bring about is that of *discrimination*. In this Article, I begin to examine the relation between discrimination-based arguments and the emergence of the scored society.[12] I also briefly examine what solutions could be applied to limit such concerns. This inquiry comes at a crucial time. The discrimination-based argument is already being examined and invoked when addressing the realm of scoring and big data by both academics

---

5. For a discussion of the rise of these dynamics almost a decade ago, see Tal Z. Zarsky, *"Mine Your Own Business!": Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion*, 5 YALE J.L. & TECH. 1 (2003).

6. Citron & Pasquale, *supra* note 1, at 12–13.

7. *Id.* at 4.

8. *Id.*

9. *Id.*

10. *Id.* at 11.

11. *Id.* at 20 (noting that at times the systems take on a life of their own).

12. As explained throughout this Article, discrimination doctrine has severe shortcomings. It is also so often applied that we might be witnessing "discrimination fatigue," and perhaps it is best we opt for broader theories and justifications such as human dignity. *See* Kenji Yoshino, *The New Equal Protection*, 124 HARV. L. REV. 747, 795 (2011). Yet given the salience of discrimination as a concept in the already existing discourse, it is important to engage in an analytical discussion as to its proper meaning.

and policymakers.[13] Specifically, a recent Request for Public Comment set out by the Department of Commerce (via the National Telecommunications and Information Administration) sought advice regarding the issues arising that involve discrimination and Big Data.[14] Yet there is much analytical work to be done before such arguments could be properly articulated in this ever-changing context. Perhaps most importantly, scholarship must connect the policy arguments and popular discontent noted in the context of scoring and big data analytics, with sound theoretical arguments voiced elsewhere while discussing discrimination.[15] Such work is critical, as it distinguishes between valid concerns and those that merely result from a Neo-Luddite sentiment or even manipulation by various interest holders.[16]

This Article draws out several antidiscrimination paradigms which on their face pertain to the dynamics discussed in *The Scored Society*, and big data in general. Such analysis allows for recognizing which discrimination-based concerns are especially acute in the scored society, as well as setting forth initial proposed responses for mitigating them, when possible. The Article proceeds as follows: after a brief Introduction mapping the confines of the debate and summarizing Professor Citron and Pasquale's contributions, the Article moves to Part I, where it generally addresses the notion of "discrimination" and its relevance to the issue at hand. Part II—the heart of this Article—identifies the discrimination-based concerns which relate to the mistreatment of "protected groups."[17] There, the Article demonstrates the possible concerns while relying on race as a key example of a

---

13. In the academic context (beyond *The Scored Society*), see generally Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact* (forthcoming), *available at* http://ssrn.com/abstract=2477899. In the policy realm, see EXECUTIVE OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 51 (2014), *available at* http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf. For a recent discussion in the New York Times which involves both academia and policy, see Seeta Peña Gangadharan et al., *Room for Debate: Is Big Data Spreading Inequality?*, N.Y. TIMES (Aug. 7, 2014, 12:22 PM), http://www.nytimes.com/roomfordebate/2014/08/06/is-big-data-spreading-inequality/the-dangers-of-high-tech-profiling-using-big-data.

14. Big Data and Consumer Privacy in the Internet Economy, 79 Fed. Reg. 32714 (June 6, 2014).

15. Discrimination indeed cannot be reduced to one theory. *See* Susan Sturm, *Second Generation Employment Discrimination: A Structural Approach*, 101 COLUM. L. REV. 458, 473 (2001).

16. For discussions of instances where these were indeed the reasons for concerns regarding technology in general, see K.A. Taipale, *Technology, Security and Privacy: The Fear of Frankenstein, The Mythology of Privacy and the Lessons of King Ludd*, 7 YALE J.L. & TECH. 138 (2005).

17. "Protected groups" is a term of art broadly applied in the discrimination discourse. *See infra* text accompanying note 49; *infra* text accompanying notes 103–105.

"protected group" and distinguishing between explicit discrimination, implicit discrimination, and instances of disparate impact. In Part III, the Article takes a brief look at selected discrimination concerns which go beyond protected groups. It generally finds these latter problems relatively easy to resolve. Finally, in the Conclusion, the Article argues that even though the scoring process is seemingly ridden with discrimination-based concerns, it certainly should not be categorically abandoned, as it might even *promote* antidiscrimination objectives when carried out properly.

The Scored Society provides an excellent point to begin an important discussion regarding the novel forms of discrimination the analysis of big data brings about. Indeed, Professors Citron and Pasquale make the crucial link between the emergence of scoring practices and a variety of troubling concerns—including discrimination.[18] Yet their article does not settle for merely pointing out problematic aspects. Professors Citron and Pasquale go even farther and offer innovative solutions.[19] They promote regulatory oversight through the introduction of licensing schemes,[20] audit logs[21] and mandatory test-runs of the scoring methods, using various data sets.[22] They also recommend the use of interactive modeling interfaces which allow those subjected to the scoring process to gain a greater understanding of its inner workings.[23] This objective could be achieved by allowing users to challenge the system with various hypothetical scenarios and examine the responses they receive.[24]

The various harms, concerns, and even solutions discussed throughout *The Scored Society* can be easily grounded in several existing moral and legal concepts. The most obvious choice is integrating the analysis of scoring into a broader discussion regarding *privacy* and *data protection* (or the lack thereof) in the digital age. Indeed, many of *The Scored Society's* concerns can be articulated through the terminology of the "Fair Information Practices"—or FIPs. FIPs are broad regulatory concepts set forth by several influential entities, most notably, the OECD.[25] Both the European Union and the United States incorporated

18. Citron & Pasquale, *supra* note 1, at 13–15.

19. Similar solutions were also set forth by others in the broader "Big Data" context. *See* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA 171–84 (2014).

20. Citron & Pasquale, *supra* note 1, at 21–22.

21. *Id.* at 28.

22. *Id.* at 25.

23. *Id.* at 28–29.

24. *Id.*

25. *OECD Privacy Principles*, OECDPRIVACY.ORG (Aug. 9, 2010), http://oecdprivacy.org/. For a

FIPs, to some extent, into their privacy laws.[26] FIPs spell out what obligations those collecting and using personal information have towards their users and "data subjects." For instance, an individual's inability to view personal data collected about her while creating the score is related to the principle of "Individual Participation."[27] The fear that said scoring-related information is inaccurate or incomplete is linked to the principle of "Data Quality."[28] Finally, the overall theme which calls for greater disclosure in the scoring process is closely related to the principles of "Transparency" and "Openness."[29]

Furthermore, the innovative solutions Professors Citron and Pasquale provide fit nicely with other regulatory proposals set forth in the privacy and data protection realm to achieve similar objectives. Most notable in this context is the European Union's Data Protection Regulation Proposal. This proposed legislation introduces novel and even controversial measures to enhance FIPs, such as "the right to be forgotten," mandatory interoperability, a guarantee for human evaluation and hefty fines, among others.[30] To some extent, integrating these proposed solutions, especially those calling for enhanced disclosure and enforcement, into existing law will allow for overcoming the various concerns regarding inaccuracy and opacity which the scored society brings about.

Yet beyond the notion of privacy and data protection, the discussion at hand invokes an additional set of concerns, which call for a serious discussion regarding *discrimination*.[31] Indeed, Professors Citron and Pasquale note that the scoring process, especially when used to formulate credit scores, can prove discriminatory, have a disparate

---

recent discussion of the formulation of Fair Information Practices (FIPs), see Robert Gellman, *Fair Information Practices: A Basic History*, ROBERT GELLMAN: PRIVACY & INFO. POL'Y CONSULTANT (Aug. 3, 2014), http://bobgellman.com/rg-docs/rg-FIPShistory.pdf.

26. *See* Marc Rotenberg, *Fair Information Practices and the Architecture of Privacy (What Larry Doesn't Get)*, 2001 STAN. TECH. L. REV. 1, 15 (2001).

27. *OECD Privacy Principles*, *supra* note 25.

28. *Id.*

29. *Id.*

30. *Commission Proposal for a Regulation of the European Parliament and of the Council: On the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation*, COM (2012) 11 final (Jan. 25, 2012), *available at* http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf.

31. For a discussion as to the relation between antidiscrimination and data protection law, see Raphaël Gellert et al., *A Comparative Analysis of Anti-Discrimination and Data Protection Legislations*, in DISCRIMINATION AND PRIVACY IN THE INFORMATION SOCIETY: DATA MINING AND PROFILING IN LARGE DATABASES 61, 61–89 (Bart Custers et al. eds., 2013).

impact,[32] and even allow for "systemizing" discrimination.[33] They also note that scoring systems are stigmatizing[34]—a term often used in conjunction with discrimination.[35] These negative outcomes can result from deliberate or accidental actions of the scoring entities.[36] Thus, a solid link between the concerns stemming from the scored society and the broader concept of discrimination is apparent. However, policy-makers and scholars must engage in additional work on the theoretical and policy level to further incorporate insights from the antidiscrimination discourse into the emerging realm of "scoring." It is also important to recognize the limits of antidiscrimination theory, and the point at which referring to this concept will not prove helpful.

Before proceeding, note an important methodological caveat. The analysis below focuses on the concept of discrimination in *theory*, as opposed to discrimination *law* and *doctrine*. Indeed, in many instances, existing laws provide important points of guidance and insight. Yet given the very dynamic nature of the context here discussed, the current analysis is almost exclusively theoretical, leaving a doctrinal analysis for another day.[37] This analytical strategy could be partially justified by noting that the laws and regulations which will address these settings in the near future have yet to be written. Thus, merely focusing on concepts is acceptable. Relying on concepts rather than doctrine provides for the questionable liberty to crop together arguments set forth in scholarship related to the constitutional aspects of discrimination (which mostly pertain to the actions of the state), and those invoking civil rights (which also pertain to private parties) into one analytical discussion.

The analysis, therefore, strives to establish an acceptable form of conduct in the scored society, while relying on a normative standard which must govern the actions of the firms engaged in scoring. Formulating such a standard in the context of discrimination is, however, extremely difficult. Explaining discrimination, as Professor Larry Alexander notes, "is much more difficult than most people assume."[38] There is no full blown normative theory of discrimination, as Alexander

---

32. Citron & Pasquale, *supra* note 1, at 5.

33. *Id.* at 13.

34. *Id.* at 7.

35. *See* FREDERICK SCHAUER, PROFILES, PROBABILITIES AND STEREOTYPES 3 (2006).

36. Citron & Pasquale, *supra* note 1, at 24.

37. For a recent discussion of some of the doctrinal issues these practices bring about, see Barocas & Selbst, *supra* note 13, at 23–43.

38. Larry Alexander, *What Makes Wrongful Discrimination Wrong*?, 141 U. PA. L. REV. 149, 151 (1992).

further explains, but rather a "messy blend of deontological and consequentialist considerations."[39] Given that this Article addresses the actions of private parties, the law must also provide such entities with the autonomy and ability to exercise their preferences. Finding fault in actions which rely upon these preferences should usually be coupled with not only a negative outcome, but with negative intentions. Yet given the severity of some actions, this general rule has exceptions.

Rather than beginning this discussion by setting out the overall theory of justice this Article will rely upon, the analysis notes several antidiscrimination justifications throughout the discussion. Some focus on the wrong of a discriminatory decision, while others focus on the wrong of a discriminatory outcome. These normative baselines are briefly detailed separately in the chapters to follow. This strategy is crucial to properly explain the very different discrimination-based challenges the scored society brings about, and the manner in which they might be resolved.

I. SCORING, DISCRIMINATING

The Scored Society identifies the fear of *discrimination* as part of a broader set of concerns such a society sets forth.[40] Discrimination is a fundamental, yet illusive concept. In a scored society, individuals who are seemingly similar are treated differently when they receive different scores. Furthermore, they are subsequently allocated different benefits.[41] However, applying the concept of discrimination to the "scoring" discussion brings about several wrinkles. In addition, the questions as to how we might resolve these concerns and take anti-discriminatory action calls for a different and specific set of tools.

Discrimination is a broad term, which has generated a breadth of legal thought and case law. It is also a charged term, which triggers visceral reactions and responses. Discussing discrimination quickly leads to considering racial discrimination and other repugnant practices of the past. Such discussions also commonly call for seeking out discriminatory intent.[42] Yet, discussing discrimination in the context of

---

39. *Id.* at 154. In his important analysis of discrimination-related issues, Schauer also explains that rather than selecting a "moral standpoint at the outset" he chooses to explore a diverse collection of topics, while examining the intuitions and arguments which apply to them. SCHAUER, *supra* note 35, at 24.

40. *See* Citron & Pasquale, *supra* note 1, at 7, 13–15, 31, 33.

41. *Id.* at 5.

42. Indeed, intent is required in terms of applying the Equal Protection Clause. *See* Alexander, *supra* note 38, at 179.

the scored society challenges existing thought paradigms. It calls for closely examining what form of racial or intended discrimination is indeed occurring. It also requires inquiring whether other forms of discrimination should be prohibited and perhaps more importantly, why.

A preliminary note must address a puzzled reader's initial concerns.[43] This hypothetical reader might argue that discrimination-related problems are of limited relevance in the scored society; discrimination pertains to instances in which those that should be treated as equal are not treated as such. In addition, discrimination transpires when those that are different are treated the same. However, the scoring dynamics allow for treating different individuals *differently*. Indeed the entire scoring system is premised upon locating differences and using them to generate unique responses. Therefore, the scored society does not enhance discrimination, but equality instead!

Yet this view mischaracterizes both the scoring process and the proper meaning of anti-discrimination policy. Antidiscrimination policy is not only about assuring equal treatment to equals, but also about assuring that specific differences among individuals should be ignored.[44] Furthermore, the principles of antidiscrimination policy go beyond making choices among individuals and pertain to removing social frameworks of inequality from society. In addition, the scoring mechanism should not be viewed as a neutral and precise measure. While it indeed strives to scientifically distinguish between the different and unequal, it at times systematically errs in doing so. Instead of identifying differences, the scoring system might even create and reinforce them. For these reasons and others, the discussion of discrimination set out below is surely warranted.

The following analysis distinguishes between various forms of discrimination-based arguments. It begins by distinguishing between discrimination against protected groups (the definition of which will be briefly discussed below) and other forms of discrimination. Many of these arguments are also stated in *The Scored Society*. Yet it is crucial to distinguish among them, and point out the relative strengths and limits of each one.

Before proceeding, a few brief notes and caveats are required regarding the specific role discrimination plays in the analysis to follow. This discussion focuses on the fairness-related aspects of discrimination,

---

43. This assumption is also noted in passing in *The Scored Society*. Citron & Pasquale, *supra* note 1, at 4.

44. SCHAUER, *supra* note 35, at 215.

without accounting for the relationship the concept of discrimination has with efficiency. Indeed, in many instances, discrimination might be inefficient and thus present an unsustainable business or social practice.[45] In other instances, discrimination might be cost-beneficial given the business settings in which the firms operate, and thus prove everlasting.[46] Yet discrimination-based concerns are relevant regardless of the specifics of the efficiency-based analysis because they are premised on the concepts of fairness and justice. Indeed, in many cases, the economic benefits discriminatory conduct provides cannot compensate for the moral wrongs it involves. Furthermore, even if the discriminatory effects of scoring are merely an inefficient phase the market is due to correct, one could deem these effects morally unacceptable, even in the short run. Notably, at times, current law sets a balance in which some forms of discriminatory conduct may be justified given their benefits, yet such actions might be still considered discriminatory nonetheless.[47]

In addition, the following discussion is premised upon several non-trivial assumptions regarding the scoring process. First, that the scores formulated, which are premised upon the individuals' previous behaviors, rely upon previously noted non-spurious correlations between individual attributes and problematic behaviors the process is trying to predict. (A default, a risk, or poor work performance are some key examples.) The correlations will not be universal; the scoring processes will never be able to indicate that when specific social variables meet specified criteria, a future event pertaining to human behavior is *sure* to occur. In the social sciences—the context which relates to the scoring of humans this Article addresses, and as opposed to rules of physics—universal correlations are quite rare.[48] A prediction made by a scoring scheme will always include an inherent chance of error. However, I here

---

45. *See* Zarsky, *supra* note 5, at 25 n.73 (referencing relevant work by Ian Ayres and Gary Becker).

46. *See, e.g.*, Ian Ayres, *Fair Driving: Gender and Race Discrimination in Retail Car Negotiations*, 104 HARV. L. REV. 817, 843–45 (1991). Ayers introduces several forms of statistical theories of discrimination, such as cost-based and revenue-based discrimination. The former refers to seller's inferences that specific types of consumers impose additional costs on the firm and therefore should be avoided, while the latter refers to cases in which the sellers draw inferences regarding the revenue which could be derived from different customers given their specific traits and therefore charge higher prices. *Id.*

47. For instance, in the context of employment and Title VII antidiscrimination provisions, disparate impact is not prohibited if it can be explained by a business necessity. 42 U.S.C., § 2000e-2(k)(1)(A) (2012).

48. *See* Alexander, *supra* note 38, at 168 (discussing this distinction).

assume that when a scoring scheme is relied upon, the error level pertaining to the schemes involved will be below a specific, acceptable, threshold. This threshold level might be dictated by either economic forces or direct governmental intervention. Assuring that the scoring firms meet this requirement is easier said than done, yet I set aside the discussions as to its enforcement for another day.

A second assumption regarding the scoring process assumes that the scoring protocols dictated by the analysis of historical statistical data are indeed followed through diligently. At times, those vested with the authority to operate the scoring schemes might feel the urge to tinker with the mechanism at their own discretion. Such tinkering might lead to various abuses, discrimination, unfair outcomes for the affected individuals, and even financial loss for the relevant firms. To sidestep these problematic aspects of a scoring dynamic, the Article assumes that the scoring schemes structured by statisticians in the back office are indeed followed to a tee by those in the field. Again, an analysis as to how this might be achieved, and whether such an objective is indeed achievable is beyond this current discussion.

II. SCORING AND THE DISCRIMINATION AGAINST PROTECTED GROUPS

A. Overview and Explicit Discrimination

Discrimination against "protected groups" is probably the most salient analytical framework within antidiscrimination law, and with good reason. Generally speaking, this form of discrimination pertains to instances in which the scoring mechanism implicates members of protected groups as higher risks (or inflicts other negative attributes) at a higher rate than others outside these groups. Naturally, this is established while accounting for the groups' proportional size in society.[49] Such discriminatory conduct subjects members of the protected groups to burdens and hardships other groups do not suffer. It also leads to additional, secondary repercussions.[50]

The correct definition and extent of protected groups is a difficult and thorny question. The United States Supreme Court has recognized a very limited number of "protected groups," such as those classified by race,

---

49. Achieving this analytical task is obviously easier said than done. *See* Barocas & Selbst, *supra* note 13, at 32 (noting the "four-fifths rule" applied in the employment context); Uniform Guidelines on Employment Selection Procedures, 29 C.F.R. § 1607.4(D) (2014).

50. *See infra* text accompanying notes 103–105.

gender, and national origin.[51] Specific federal and state laws have further expanded these categories.[52] While some laws prohibiting discrimination merely pertain to actions of the state (as opposed to the private parties engaged in the scoring dynamic here discussed), the protected group's definition gives insights as to society take regarding this matter. The notion of what constitutes or should constitute a "protected group" is constantly in flux, yet it generally pertains to insular minority groups that in most cases suffered from past subordination. The discussion here will focus on the most salient form of such discrimination—one that is premised on the individual's *race*.

Professors Citron and Pasquale argue that the scored society increases the risks of discrimination towards protected groups and minorities. Yet examining this argument calls for the further parceling of this subset of discrimination. Discrimination against protected groups could result from explicit or implicit actions. In addition, it might result from neither, but the outcome of applying the scoring mechanism might generate a disparate impact against protected groups. An important element for distinguishing between various forms of discrimination is, therefore, the notion of *intent*. Forbidding intentional discrimination is easy to theoretically justify. But the unintentional discriminatory consequences of profit-seeking yet prejudice-free business-related conduct call for additional thought. In the next subsections, the Article will briefly examine the relevance of every one of the latter two concerns (implicit discrimination and disparate impact) to the issue at hand.[53]

Explicit Discrimination: Explicit discrimination here refers to the unacceptable instances in which the score formulated explicitly includes,

---

51. Yoshino, *supra* note 12, at 756 (drawing out the five classifications the Supreme Court has formally accorded heightened scrutiny: "race, national origin, alienage, sex, and nonmarital parentage"). Yoshino argues there will be no more classes accepted by law. *Id.* at 757 ("this canon has closed").

52. *Id.* at 757 n.73 (explaining that some states, such as California, have added "sexual orientation" as a protected class). In addition, age and disability have been recognized as classes worthy of protection under U.S. federal laws. For a discussion as to the groups recognized in the U.S. and other countries, see SANDRA FREDMAN, EUROPEAN COMM'N, EUROPEAN COMM'N DIRECTORATE-GENERAL FOR JUSTICE, COMPARATIVE STUDY OF ANTI-DISCRIMINATION AND EQUALITY LAWS OF THE US, CANADA, SOUTH AFRICA AND INDIA 25 (2012), *available at* http://ec.europa.eu/justice/discrimination/files/comparative_study_ad_equality_laws_of_us_canada_sa_india_en.pdf.

53. Distinguishing between these notions of discrimination when structuring an analytical framework is a common analytical move among scholars discussing discrimination. For instance, Richard Ford explains that "[w]e have three *distinct but related* theories of discrimination: formal discrimination, discriminatory intent, and discriminatory effects," thus mapping out a similar framework to the one noted in the text. RICHARD FORD, THE RACE CARD 183 (2008) (emphasis in original).

as a factor, an indication as to whether the individual belongs to a protected group (or, in the key example used thus far, the individual's race). For such an outcome to unfold, the creation of the score was preceded by a process which featured the collection and subsequent analysis of information regarding the specific individual's race.[54] Because current legal, social, and ethical rules as well as market sentiments will more than frown upon scoring which is explicitly based on race, such discrimination will most likely not transpire.[55] A complicated subset of this form of discrimination, which indeed might unfold, is the use of race and other protected factors in scoring to promote affirmative action. This issue will be set aside for now.[56]

Three analytical points strengthen the case for categorically banning such practices, yet nonetheless explain why this risk would probably fail to materialize in the scored society. I start with *justification*. The justifications for prohibiting these forms of discrimination can be articulated on several levels, which are both deontological and consequential. In other words, they both focus on the processor's intent or on the process' outcome.

The *intent* to discriminate is unacceptable, in that when engaging in these forms of scoring, firms act differently and adversely towards minorities *because* they are minorities. Such conduct undermines basic notions of fairness.[57] In addition, explicit and intentional discrimination is usually intrinsically immoral, as it features one individual incorrectly judging another to be of lesser moral worth.[58] Yet explicit discrimination generates an additional specific harm given the actual and intentional use of membership of a protected class in the analysis process.[59] Referring to

---

54. For a discussion advocating some exceptions for using these factors when the consequences of neglecting them would prove "catastrophic" and their use will substantially increase the process's effectiveness, see SCHAUER, *supra* note 35, at 186. This argument mostly pertains to the realm of national security and thus will be set aside in the context of this Article.

55. *See* FORD, *supra* note 53, at 180 ("Today, facial discrimination is rare. Because the law flatly forbids it with very few exceptions and because it is conspicuous when it happens, few people pass facially discriminatory laws or adopt facially discriminatory policies.").

56. For more on this issue, see generally Richard Primus, *The Future of Disparate Impact*, 108 MICH. L. REV. 1341 (2010); FORD, *supra* note 53, at 245–63.

57. *See* SCHAUER, *supra* note 35, at 203–04 (discussing the basic Aristotelian concept of equality which is a measure of fairness).

58. Alexander, *supra* note 38, at 159.

59. Barocas & Selbst, *supra* note 13, at 25. For a discussion on the degrading nature of racially discriminating laws and the importance to reduce this effect, see FORD, *supra* note 53, at 180–83 (explaining the insult caused to minorities when learning that others perceive them as inferior); Alexander, *supra* note 38, at 192 (discussing the insult biases entail); Jed Rubenfeld, *Affirmative Action*, 107 YALE L.J. 427, 467 (1997).

this specific form of attribute in a scoring process has a degrading effect,[60] regardless of the negative impact of the detrimental treatment and to the extent its usage is publicly known. Interestingly, this final rationale might be somewhat weaker in the scoring contexts here discussed, given its automated nature.

The o*utcomes* of explicit discrimination often lead to unfair distributions of wealth and resources. In many cases, structuring scoring systems to discriminate on the basis of immutable traits such as race amounts to subordinating these minority groups. In some instances such discrimination contributes to their seclusion. One can argue that minorities must not receive negative treatment in view of wrongs inflicted upon them in the past and the persistence of negative stereotypes about them in the present.[61] Yet given the importance of maintaining the firms' ability to freely exercise preferences, the justifications which rely on mere outcome are somewhat weaker, they will be the focus of the "disparate impact" discussion below.

A second point concerns the actual form of intention such discrimination involves. Explicit discrimination need not require the presence of bigotry.[62] Indeed, a firm might apply such a racial factor in its analysis while merely striving to achieve various economic objectives, such as exploiting a minority group's disadvantages, or catering to the (at times, implicit) prejudice of its customers.[63] Even in the absence of explicit bigotry, such other forms of intent are more than enough to justify a categorical ban on these practices because most of the deontological and consequential justifications noted above apply.

To further strengthen the argument for categorically banning this form of scoring, it is worth emphasizing that scoring practices based on racial characteristics must be considered forms of intentional discrimination even when the scoring method is formulated automatically. Consider a scenario in which a multitude of data input points—including those pertaining to race or other protected parameters—are collected in the field and aggregated into one dataset. Thereafter, data mining is used automatically to identify factors that

60. One might note a possible caveat that this harm is reduced in the situation at hand, given its secretive nature. Yet this response is unacceptable. Not only might information regarding such conduct leak, the fact that those carrying out the scheme know of it generates substantial harm.

61. For an additional discussion of these elements, see *infra* notes 108–121 and accompanying text.

62. This form of discrimination is referred to as "first generation discrimination." *See* Sturm, *supra* note 15, at 461.

63. *See* Ayres, *supra* note 46, at 843–45. Note that in some of these cases, minorities have chosen to discriminate against minorities as well for such reasons.

correlate with risk-predicting behaviors. Finally, such factors are grouped together to formulate a score.[64] Those applying the score might not have intended to use race in this process and thus argue that this was merely the result of the automated process which "chose" this factor out of hundreds of options. Yet these practices nonetheless feature intent, even though its manifestation is somewhat unconventional. The indications of intent in this setting could be derived from either the inclusion of race in the initial set of analyzed factors or the practice of applying a scoring mechanism which relies on race, even among many other elements. The firm should not be able to hide behind the automatic nature of the process, and must actively examine the factors it applies in the scoring process before it is launched.[65]

Even though the nature of intent in automated process is different, most of the justifications for banning intentional discrimination apply. The firms' actions intentionally single out minorities, treat them differently and thus act unfairly. In addition, the scoring process will also prove degrading towards minorities as it would be understood to generate an intentional discriminatory process. However, the scoring entities might not be making a moral judgment which is intrinsically unjust, as they might not be actively finding another group of humans to be of lesser worth when relying on the automated output.

Third, battling this form of explicit discrimination will not prove to be simple, but is ultimately achievable. While setting forth rules which ban such practices might be relatively easy, enforcing such a ban in a world in which the nature of the algorithm used is secret might prove to be a challenge. However, many of the solutions set forth by Professors Citron and Pasquale will be helpful in revealing such practices, thus leading to their termination.[66] For instance, various forms of auditing and licensing will contribute to revealing explicitly discriminatory scoring.[67] Furthermore, I suspect that the usage of race in the scoring process will most likely leak to the authorities or to the press. Therefore, firms will be reluctant to make use of such factors explicitly to begin with. They might, intentionally or not, opt for implicit forms of discrimination, which the Article now addresses.

---

64. For a full description of this process, see Zarsky, *supra* note 5, at 9–15.

65. This, in fact, will require the scoring process to be deemed interpretable—understandable at least to the firm's own analysts. For a broader explanation of the concept of interpretability and its implications, see Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1519 (2013).

66. *See supra* text accompanying notes 17–22.

67. *See supra* text accompanying notes 17–22.

B. Implicit Discrimination

Even when scoring firms do not explicitly rely upon race or other factors related to protected groups, scoring methods can prove discriminatory. For example, this will occur when the scoring mechanism applies proxies for race as part of its protocol. Such a dynamic will feature the inclusion of factors which, when applied in concert, are correlated with race beyond a specific threshold (i.e. the proxy need not prove perfect to be considered problematic). There might also be other subtle ways to engage in this form of discrimination. As Professors Citron and Pasquale explain, biases can be "embedded into the software's instructions."[68] The crucial point here is that when the final outcome of the scoring process would be adverse towards a minority or other protected group, hidden forms of intent might be in play. These are instances commonly referred to as implicit discrimination.

Implicit discrimination (at times described as "second generation discrimination"[69]) might indeed prove intentional. Yet identifying such intent and applying the antidiscrimination justifications explained above is quite a challenge. Three common forms of implicit discrimination within the context of a scored society include masking, subconscious discriminatory motivations, and relying upon tainted datasets or tools. In what follows, this Article details the specific theoretical aspects of each form; examines the role intent[70] might play; and provides a limited discussion as to possible actions which might be taken to mitigate this concern.

Masking: The decision to engage in implicit discrimination might be a clear and calculated one. As Barocas and Selbst recently explain in their working paper titled *Big Data's Disparate Impact*, discriminating firms might choose to "mask" their discriminatory practices in a way in which the discrimination might prove undetectable, or at least defensible.[71] These might be actions of firms which strive to introduce specific and organizational structures to assure that social status quo is maintained and minorities receive inferior treatment.[72] These need not be the work

---

68. Citron & Pasquale, *supra* note 1, at 4.

69. Yoshino, *supra* note 12, at 768; Sturm, *supra* note 15, at 468; *see also* Kenj Yoshino, *The New Equal Protection*, CONST. IN 2020 (Dec. 1, 2004), http://constitutionin2020.blogspot.co.il/2004/12/new-equal-protection-post-by-kenji.html.

70. Note again that the following is a theoretical, not a doctrinal analysis. Legal doctrine has limited the meaning of "intent" in this context. *See infra* text accompanying note 84.

71. Barocas & Selbst, *supra* note 13, at 21–23.

72. Sturm, *supra* note 15, at 469 ("Exclusion increasingly results . . . as a byproduct of ongoing

of bigotry necessarily, but merely of calculated businesspeople striving to achieve their objectives.

Nonetheless, the "masking" of such discrimination is unacceptable. Here, most of justifications against explicit discrimination are directly relevant. Such conduct undermines a basic notion of equality.[73] Individuals must not be treated differently based on their race or *because* of their race[74] (or any other factor distinguishing protected groups). Moreover, if masking resulted from intentional discrimination against a specific group, it is intrinsically morally unacceptable.[75] In addition, masking generates unjust outcomes (a limited argument when presented on its own). Masking, however, raises analytical questions as to the applicability of the antidiscrimination justification which strives to limit the insult to minorities. Here, one might argue that the "masking" counters the insult such practices convey. Yet such insult will indeed unfold should society suspect (and in the digital age, such suspicions will quickly circulate) that such masking has transpired.

To effectively identify and hopefully eliminate these forms of discrimination, the disclosure measures Professors Citron and Pasquale propose appear adequate. For instance, studying the firm's audit logs might allow for discovering such discrimination after the fact.[76] However, to initially start such an inquiry, the scoring systems must be tested and examined so to reveal whether minorities receive overall inferior treatment, and thereafter an investigation to try and uncover such masking might be launched. Indeed *The Scored Society* makes reference to such initial steps, while explaining that scoring systems "should be run through testing suites that run expected and unexpected hypothetical scenarios designed by policy experts. Testing . . . would help detect both programmers' potential bias."[77]

Subconscious Bias and Implicit Discrimination: A second form of implicit discrimination might result from long learned biases against minorities and other protected groups. These biases *subconsciously* (or unconsciously) impact the actions of the analysts who structure the scoring systems, and the algorithms that recommend what parameters

---

interactions shaped by the structures of day-to-day decision-making and workplace relationships.").

73. *See* Ronen Perry & Tal Z. Zarsky, *Queues in Law*, 99 IOWA L. REV. 1596, 1608–14 (2014) (explaining the analytical foundations of negative and positive equality and pointing to other relevant sources).

74. For a discussion in the context of employment, see Sturm, *supra* note 15, at 473.

75. *See supra* text accompanying note 58.

76. Citron & Pasquale, *supra* note 1, at 28.

77. *Id.* at 25.

should be used, and how they are used.[78] Analysts might be prone to make decisions that evidentially disenfranchise minorities given their own prejudice.[79] Subtle decisions may lead to an overall discriminatory outcome. A scoring process which is premised on the collection and analysis of vast databases of personal information offers a wealth of opportunities for subconsciously biased decisions to materialize.[80] For instance, at some points, analysts must decide which correlations and patterns should be incorporated into the scoring model and which must be set aside as "junk," random results, or statistical errors. Here, the analyst's biases might shape the final outcome and the discriminatory effect it will involve.

Subconscious discrimination is yet another form of implicit discrimination which society must find unacceptable.[81] The fact that actual "intention" in its customary form is absent indeed undermines some of the basic justifications for applying antidiscrimination measures. It is difficult to find discrimination undertaken subconsciously on par with conscious discrimination which expresses a lack of moral respect and clearly generates insult.[82] Nonetheless, implicit discrimination still generates an outcome which singles out minorities because they are minorities, and is therefore unjust. Prohibiting this form of discrimination could also be merely justified based on the imbalanced

---

78. *See* Jerry Kang & Kristin Lane, *Seeing Through Colorblindness: Implicit Bias and the Law*, 58 UCLA L. REV. 465, 484–85 (2010) (noting studies that indicate implicit racial bias and discrimination in employment); Amy L. Wax, *Discrimination as Accident*, 74 IND. L.J. 1129, 1138 (1999) (discussing the role unconscious bias plays in various contexts and how it might amount to disparate treatment). *But see* Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023, 1108–10 (2006) (calling for caution when applying the findings indicating biases in the lab to actual policy decisions which pertain to and impact actual interactions in the field).

79. For a famous discussion of such bias against married couples in a somewhat different context, see Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, *in* HUMAN VALUES AND THE DESIGN OF COMPUTER TECHNOLOGY 21, 30 (Batya Friedman ed., 1997) (one can further speculate that given the preference the system gave to the choices of the couple's predefined "leading member" this bias actually adversely impacted women).

80. Zarsky, *supra* note 65, at 1518–19; *see also* Steve Lohr, *For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights*, N.Y. TIMES (Aug. 17, 2014, 4:55 PM), http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html (explaining that the analysis of big data first requires a great deal of manual work—fifty to eighty percent of the overall time of analysis—collecting and preparing data for the subsequent process. These often ignored preparatory steps allow for a variety of instances in which the biases discussed could impact the scoring process.).

81. *See* Alexander, *supra* note 38, at 179–83, for an interesting discussion as to when this form of discrimination should be considered a conscious one.

82. *See id.* at 181.

outcomes which will follow. Yet, as noted,[83] the merit of such justifications is limited.

However, battling these forms of discrimination might require somewhat different responses than those noted above. As with other forms of subconscious discrimination, education might prove helpful.[84] In addition, assuring a higher level of automation and adherence to the statistical protocols (and a lower level of human discretion which fuels such implicit discrimination) is, somewhat counter-intuitively,[85] called for. Yet above all, novel methods of auditing and supervision of the analysts' actions must be developed to limit such discrimination.

Reliance on Tainted Datasets and Tools: In a third set of cases, those structuring the scoring mechanism might generate a discriminatory outcome by recklessly[86]—or perhaps, merely negligently—relying upon tainted datasets.[87] The analysis might rest upon existing datasets or data collection methods which systematically discriminate (or discriminated in the past) against minorities. For instance, the dataset pertaining to employees' previous achievements (upon which the analysis was later premised) might be tilted against minorities given their lack of past success because of a harsh work environment resulting from discriminatory attitudes, or the lack of minority hiring altogether.[88] In other instances, negative information regarding minorities might be overrepresented in the relevant database. This might result from an oversampling bias; given existing prejudice, minorities are sampled more often for indiscretions, and thus their indiscretions are over-represented in the database which is subsequently used for formulating

---

83. *See supra* text accompanying note 60.

84. Kang & Lane, *supra* note 78, at 500 (demonstrating how de-biasing could be achieved via juror education to battle bias in the courtroom).

85. This measure is counter-intuitive, as one would assume that enhanced automation would generate a greater, rather than a lesser, amount of this form of discrimination. Automation is often considered to invoke inherent opacity, which in turn will allow for various forms of implicit discrimination such as "masking" to unfold.

86. For an analysis of the standard for "recklessness" in the context of discrimination in employment, see generally Joseph A. Seiner, *Punitive Damages, Due Process, and Employment Discrimination*, 97 IOWA L. REV. 473 (2012).

87. A possible difficulty might arise in establishing whether those relying on the datasets are indeed reckless, or merely negligent. Applying the "reckless" paradigm to this juncture allows for easily including this discussion within the broader topic of intentional discrimination. For a discussion of discrimination via negligence in a somewhat different context, see David Benjamin Oppenheimer, *Negligent Discrimination*, 141 U. PA. L. REV. 899 (1993) (focusing on workplace discrimination).

88. *See* Barocas & Selbst, *supra* note 13, at 13. For a different form of misrepresentations of minorities, see Kate Crawford, *The Hidden Biases in Big Data*, HARV. BUS. REV. BLOG NETWORK (April 1, 2013, 2:00 PM), http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/.

the scoring scheme.[89] Even though some of these errors will be resolved over time as samples are renewed, others will persist. The likely outcome of an analysis (be it manual or automated) that relies upon a database ridden with discriminatory data would itself be a discriminatory score and a process which systematically disfavors minorities.

This form of discrimination does not involve the explicit discriminatory intent of those who structure the scoring schemes.[90] Yet the recklessness (or negligence) of relying on a tainted dataset leads to similar discrimination via scoring. Therefore, such behavior should be actively countered. The justifications for applying anti-discriminatory measures at this juncture are similar (and limited) to those offered to justify actions to counter subconscious discrimination. Here again, minorities and other protected groups are subjected to negative outcomes, and are treated negatively because of bias and prejudice—actions which society deems unfair. Indeed, such biases and prejudice are reflected in the explicit, implicit, or even subconscious actions of *others*[91]—those either creating the dataset, or engaging in various actions which caused the dataset evidently used to be racially imbalanced and tainted. Yet the analysts constructing the scoring systems cannot escape responsibility. By relying on tainted datasets they directly contribute to the discriminatory conduct. Analysts and executives must proactively assure that the datasets used are not tainted prior to launching schemes based upon them, as these datasets will adversely impact minorities.

The firms' failure to act[92] and prevent discrimination is, therefore, the analytical hook which incorporates intent into this scenario and provides justification for intervening in the discriminating firms' actions. Additional policy discussions must establish the proper standard of care this normative justification calls for on behalf of the scorers. As well, it must be established whether merely recklessness or perhaps negligence might render scorers' actions immoral. However, once the scoring entity is put on notice as to possible problems in the dataset, much of this discussion should be rendered irrelevant. The firms relying on a dataset

---

89. *See, e.g.*, BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 29–30 (2007).

90. Current law only sees "intent" as conscious choices. *See* Linda H. Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161, 1231 (1995).

91. Analyzing the normative shortcomings of those whose actions contributed to the creation of the tainted dataset calls for a distinct discussion which is unnecessary at this juncture.

92. For a discussion of duties to act in a broad variety of settings related to discrimination law, see Oppenheimer, *supra* note 87, at 936.

known to be tainted are reckless and their actions are morally unacceptable. Thus, the importance of disclosing the inner workings of the scoring process—at least to a group of experts—is apparent. Such disclosures[93] assure that the scoring entity will no longer be able to claim it inadvertently relied upon tainted data (as third parties examining the process will call this fact to its attention). However, discriminatory outcomes which persist even after the scoring entity meets acceptable standards of caution must be addressed under the discussion of "disparate impact" in Part III below.

Turning to possible solutions for the "tainted dataset" problem leads to several challenging dynamics. Beyond transparency, discrimination which results from tainted datasets might be resolved by insisting that analysts take sufficient steps to scrub existing processes and datasets of previous biases prior to incorporating them into the scoring process (thus "de-biasing" them). Recent scholarship in computer science has set forth various suggestions as to how this might be both achieved and audited.[94] Auditing whether dataset de-biasing processes were properly carried out does not require the exposure of the proprietary scoring method, i.e. the factors used in setting the score and the balance among them. Scoring firms often resist such disclosure while claiming that it undermines trade secrets and enables gaming of the system.[95] The disclosure discussed here merely involves examining the databases used in the process. Thus, objections to allowing governmental or other external watchdogs to engage in such queries are less persuasive. However, this de-biasing process might be unachievable and calls for greater thought.

Blatant Proxies: Before concluding, let us consider a specific situation which does not neatly fit within the discussion of explicit or implicit discrimination. In this case, merely *one* of the factors used in the scoring process (and of course the process of analysis could apply hundreds of factors) is a suspicious factor which could only be understood as a proxy for a protected group. Antidiscrimination law has an extensive history of dealing with factors suspected to be mere "stand-ins" for protected groups, such as zip codes as proxies for race (i.e.

---

93. It is possible that the disclosure need not include the actual datasets but merely indicate the data sources used. Thereafter, experts can inform the scoring entity that such sources are historically tainted.

94. *See* Andrea Romei & Salvatore Ruggieri, *A Multidisciplinary Survey on Discrimination Analysis*, KNOWLEDGE ENG'G REV., 2013, at 39–40; *infra* text accompanying notes 126–127.

95. *See* Citron & Pasquale, *supra* note 1, at 5, 17 (noting certain credit companies have refused to divulge their algorithms on the basis of trade secret protections); *id.* at 26, 30 (discounting concerns that divulging the logic of predictive algorithms would result in individuals gaming those systems).

redlining),[96] and height and weight as proxies for gender.[97] The age of Big Data might bring with it a new wave of blatant proxies. These proxies will now be available to the data analysts, which could easily produce a variety of accurate predictors of membership in a protected group. For example, consider an insurer or retailer using the consistent consumption of Hallal meat (an indication of adherence to the rules of Islam) or the systematic avoidance of shopping during Jewish holidays (an indication of the consumer being an orthodox Jew) as one of the many factors applied in its pricing or marketing schemes. These are merely simple examples of precise proxies for religion which could be inserted into a scoring mechanism.

I argue that the use of such factors is socially unacceptable, and that these factors must be removed from the scoring algorithm when highly correlated with a protected group.[98] As opposed to the last two forms of discrimination (the existence of subconscious bias and the use of tainted datasets), the normative emphasis here should be on the insult such discrimination generates. Pointing to minorities, even via proxies, must not be allowed because the precise nature of the proxy is largely capable of generating an insulting impact on par with that of explicit discrimination.[99] With such a blatant proxy, it need not matter what the internal motivations for applying these factors prove to be (thus distinguishing this category from that of "masking"). It also need not matter whether applying such proxies has any form of effect (as opposed to "disparate impact" discussed below). The analytical hook for finding such actions morally unacceptable is again that of the firms' reckless or negligent conduct—this time while applying proxies—are highly correlated with race.

Those regulating scoring processes must take steps—within reason—to actively block the usage of blatant proxies. Scoring entities should apply auditing methods to reveal these practices, even at the cost of efficiency. Such steps should examine correlations between the factors selected in the scoring practice and those factors which implicated minorities (which must be made available through other data sources) in an effort to detect such blatant proxies. Given the fear that exposing the

96. Barocas & Selbst, *supra* note 13, at 19.

97. Dothard v. Rawlinson, 433 U.S. 321, 323–24 (1977).

98. Additional policy work is required to set parameters as to how to distinguish between an unacceptably structured blatant proxy, and a simple and less accurate one that would be allowed.

99. Adverse outcomes towards minorities might follow as well, yet this section focuses on harms related to intent. The extent to which outcome-based concerns are sufficient to generate a regulatory response is discussed *infra*, Part II.C.

nature of the scoring process might reveal the firms' trade secrets, such auditing might be carried out either internally, or through an automated process carried out by a certifying authority.

To summarize, reliance upon scoring might lead to implicit discrimination against protected groups. Recognizing the methods of discrimination allows for sharpening the normative justifications and setting forth policy recommendations which call for various forms of disclosure, as well as other specific regulatory responses. I concede that this discussion expands the notion of intent and includes within it a vast set of behaviors. Such an expansion was crucial to analytically consider the discriminatory actions in view of a broader set of justifications, which do not merely involve the unacceptable effects of the discriminatory outcome.

Yet even such an expansive definition of intent cannot include all instances in which scoring methods lead to discrimination. For example, discrimination might occur when non-blatant proxies for race are applied. Or the datasets used might not be tainted to the extent that relying on them is reckless (or negligent). To capture the cases which cannot be categorized as implicit (or explicit) discrimination, I now introduce and address the concept of disparate impact.

C. Disparate Impact

Beyond the forms of discrimination discussed thus far, there might be yet another set of scenarios which *The Scored Society* brings about; one that merely features disparate impact. To explain, consider the following example. Here, an audit of the scoring mechanism indicates that a specific racial minority or other protected groups received adverse treatment. Those administrating the scoring system are baffled and upset by the audit's conclusion. They launch an extensive inquiry to understand why this outcome unfolded. The investigation revealed no smoking guns. An investigator reasonably examining the auditing logs found no intention to engage in masking. The initial actions of the scoring analysts were all within reason and with no hint of subconscious bias. The datasets used were all reasonably updated and even properly corrected to assure they do not account for past mistreatments. Yet nonetheless the scoring system is biased towards minorities;[100] the scoring process that emerged—as a whole—proved to be a proxy for

100. This description is not entirely hypothetical, but reflects recent lawsuits against various entities such as lenders and insurance companies. For one example of such a lawsuit, see Citron & Pasquale, *supra* note 1, at 15.

race.

This example leads to a crucial question—must scoring mechanisms which produce a disparate impact, yet contain no trace of prejudice or intent, be banned?[101] If so, what normative theory upholds this conclusion? Before proceeding, a quick doctrinal note is called for. In the United States, current law only forbids disparate impact in specific instances, such as employment.[102] However, even in many of these cases, practices which generate a disparate impact on minorities can still be justified when a "business necessity" for such conduct exists.[103] The definition of the term "business necessity" is murky and somewhat unclear.[104] In addition, the Supreme Court does not recognize discrimination via disparate impact alone as countering the constitutional protection provided by the Equal Protection Clause.[105] Nonetheless, examining disparate impact in this novel context is an important task given scoring's unique attributes.

Disparate Impact as Disparate Treatment in Disguise: One reason to prohibit disparate impact is that it provides a clear indication of one of the various implicit or explicit forms of discrimination—discrimination practices to which responses could be justified, as noted above. In other words, the rationale for prohibiting disparate impact lies in the implausibility of the fact pattern noted above: discriminatory intent (broadly defined) was not found because those looking for it have not looked hard enough. Or, those engaging in discrimination were very successful in masking their actions. Alternatively, the datasets used were not sufficiently examined to remove tainted data (although the firm might argue that they were). Disparate impact should therefore be considered as smoke indicating the fire of intentional discrimination.[106]

---

101. In many instances, these issues are not resolved by courts, but settled outside of them. *See, e.g.*, Charlie Savage, *Wells Fargo Will Settle Mortgage Bias Charges*, N.Y. TIMES, July 13, 2012, at B3.

102. *See* Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 253 (codified as amended at 42 U.S.C. §§ 2000e to 2000e-17 (2012); *see generally* Primus, *supra* note 56, for a discussion of the proper way to understand disparate impact provisions in view of a recent Supreme Court precedent; Richard A. Primus, *Equal Protection and Disparate Impact: Round Three*, 117 HARV. L. REV. 493, 515 (2003) (drawing out the legislative motives behind disparate impact doctrine).

103. 42 U.S.C. § 2000e-2(k)(1)(A)(i).

104. *See* Barocas & Selbst, *supra* note 13, at 32 ("[T]he definitions of . . . business necessity have never been clear . . . .").

105. *See* Washington v. Davis, 426 U.S. 229, 242 (1976); Stacy E. Seicshnaydre, *Is the Road to Disparate Impact Paved with Good Intentions?: Stuck on State of Mind in Antidiscrimination Law*, 42 WAKE FOREST L. REV. 1141, 1154 n.65 (2007). Note, however, that other countries have broader recognition and protection for disparate impact. *See* Fredman, *supra* note 52, at 50–55.

106. *See* Primus, *supra* note 56, at 1376 (referring to this rationale as an "evidentiary dragnet").

It must be treated just as explicit or implicit discrimination, with all the theoretical justifications that follow for prohibiting those types of discrimination. Even if a few specific cases of disparate impact do not actually involve intentional discrimination, yet are nonetheless considered as such, an over-inclusive legal rule regarding disparate impact is still an acceptable social compromise. It is merely a small and fair price to be paid (most likely by the powerful majority) in the overall battle against racial discrimination (or the discrimination of other "protected groups").[107]

Those who suspect the existence of implicit discrimination, even when it cannot be found, often have good reasons to do so. The argument that intentional discrimination is afoot when disparate impact exists is particularly strong in contexts where specific groups have been harshly discriminated against in the past. However, additional studies of this matter might indicate that some forms of intentional discrimination have been set aside, especially if carried out in the automated environment of the scored society. Therefore, it is important to further examine theories which find disparate impact to be normatively unacceptable, irrespective of intent.

A Normative Theory of Disparate Impact: Broadly speaking, the justifications for barring disparate impact could be summarized under two (somewhat overlapping) themes.[108] Naturally, these are consequential arguments which focus on the severe outcomes of disparate impact towards a minority group. They explain why society must not permit such outcomes to unfold, given the substantial harm they will unequally inflict on one social segment. Such severe harm justifies placing legal requirements on private entities which are seemingly exercising their legitimate preferences.

One such justification is derived from the harm of *social segregation*. Providing inferior treatment to minorities removes them from society as

---

He also sees this rationale as weak, noting that the rule does not include a "good faith" exception. *Id.*

107. Alexander and Cole make a similar point regarding the logic of applying an overall ban on discrimination—be it rational or irrational, see Larry Alexander & Kevin Cole, *Discrimination by Proxy*, 14 CONST. COMMENT. 453, 456 (1997) ("[W]e would be . . . wise to pay the cost of forbidding some rational racial discrimination in order to ensnare nonrational and irrational racial discrimination that might otherwise sneak through masquerading as rational discrimination.").

108. Prominent antidiscrimination scholars have noted the difficulty with unwrapping the theoretical components of disparate impact. Richard Ford, for example, has noted that the theory of disparate impact remains "the most controversial doctrine in contemporary civil rights." FORD, *supra* note 53, at 217.

a whole and segregates them.[109] In some instances, disparate impact might even lead to social subordination[110]—a situation in which a specific group is treated poorly as a whole. These dynamics have extensive and long lasting repercussions. Systematically discriminating against one group removes its members from positions of power and wealth in society,[111] causing them to remain internally segregated.[112] With time, this social conduct might generate additional negative dynamics impacting the segregated group, such as social disengagement, distrust, higher crime rates, and, at the end of the day, greater inequality between the minority and majority groups. These are all outcomes which society must try and avoid.

Another related theory finding disparate impact problematic is that of *stigmatization*.[113] A social system which generates a disparate impact by providing minorities with inferior treatment either creates or most likely contributes to a negative stigma already attached to these minority groups. For instance, when a scoring system indicates that, on a whole, members of a protected group are less creditworthy in a specific setting, this indication sends a message, which contributes to a historical prejudice that all members of this group are not trustworthy in many other contexts. Such stigma impacts the way the group is viewed by others, as well as how the group members view themselves.[114] Stigmatization leads to the lowering of aspirations, energy, and productivity within the group.[115] It might also lead to adverse treatment of this group in other contexts by others. In many instances, "stigmatization" and "segregation" are closely linked. The isolation of a specific group feeds the negative stigma about it, and vice versa.

---

109. *Cf.* SCHAUER, *supra* note 35, at 189 (explaining how the inclusion of race or ethnicity in an algorithm may produce racial or ethnic separation, and that this might be undesirable morally and socially); *see also* Primus, *supra* note 56, at 1376 (explaining that one of the justifications for prohibiting disparate impact is to "redress self-perpetuating racial hierarchies inherited from the past").

110. *Cf.* Barocas & Selbst, *supra* note 13, at 54 (noting anti-subordination theory as a principle that "undergirded anti-discrimination law").

111. Citron & Pasquale, *supra* note 1, at 18.

112. These are dynamics that reinforce themselves, possibly replicating a "caste system." J.M. Balkin, *The Constitution of Status*, 106 YALE L.J. 2313, 2353 (1997).

113. *See* Citron & Pasqual, *supra* note 1, at 7; *cf.* SCHAUER, *supra* note 35, at 189.

114. *See generally* Alexander & Cole, *supra* note 107, at 457 ("Members of groups discriminated against might suffer psychically even if not personally affected by the law."). Note the difference between this argument and the one pertaining to the insult derived from intentional discrimination. Here, the stigma attaches as a result of the discriminatory outcome, regardless of the perceived or actual intent of those carrying out the scoring process.

115. Alexander, *supra* note 38, at 162.

Fleshing out the theories behind an overall prohibition of disparate impact allows for identifying the limits of such theories, when applied to scoring. The theory of segregation is indeed relevant in the contexts at the focus of Professors Citron and Pasquale's work—those of *credit* and *employment*. Here, lower rankings and scores applied to a specific group quickly translate into transfers of wealth and changes in social stature.[116] It is therefore no surprise that in these specific contexts the United States Congress stepped in to prohibit private parties from engaging in discrimination—even in the form of disparate impact alone—against protected groups.[117] However, one might ask whether this rationale equally holds for other instances in which scoring is or will be applied to such groups, such as in advertising, marketing, and perhaps insurance. The wealth of social groups that receive different or even inferior offers for the purchase of goods will likely only be marginally impacted and this process will probably not lead to the groups' segregation. Therefore, prohibiting disparate impact in advertising and marketing practices cannot be justified on the basis of the "segregation" theory. I further note, cautiously, that scoring practices applied in most forms of insurance will probably *not* generate the signs of segregation or its effects as well. If an individual is unable to obtain insurance due to high rates, this lack of insurance is surely detrimental for that individual, especially if the uninsured risk materializes. And the disparate impact of higher rates of insurance might encumber an entire group. Yet stating that higher rates will generate a segregating effect is somewhat of a stretch, when compared to the very real segregation that lack of access to certain types of employment and credit will create.

On the other hand, considering the effects of "stigmatization" that may occur through the disparate impact of a scoring process may lead to a different conclusion. Indeed, this second justification might provide greater credence to the prohibition of disparate impact in other scenarios. Here, the key to an extensive adverse impact on a specific segment of society is how visible and salient the process is in the eyes of those discriminated against, as well as to other segments of the public. *Advertising* and *marketing*, for instance, are highly salient by nature and the messages they produce reach a broad segment of the public. A stigmatizing message delivered in such contexts will have a wide-

116. *Cf.* Lea Shepard, *Toward a Stronger Financial History Antidiscrimination Norm*, 53 B.C. L. REV. 1695, 1765 (2012) (noting the effect of low credit scores on racial equality and social mobility).

117. For protection with respect to creditors, see *id*. at 1729. With respect to employment, see Title VII of The Civil Rights Act of 1964, 42 U.S.C. §§ 2000e to 2000e-17 (2012).

ranging impact. Therefore, the potential for stigmatization should raise concern, even if no discriminatory intent is found.[118] The same could probably be said of insurance. Here, a discriminating scoring scheme forcefully signals to the public that a specific group generates greater risks—a message which could be easily misinterpreted and applied wrongfully to a variety of other contexts. The scientific aura of the scoring process will most likely further exacerbate the stigma-based concerns.[119]

A possible caveat, limiting the strength of the "stigma" argument in the context of scoring, pertains to the process's limited visibility.[120] The internal workings of the scoring process are notoriously opaque. Given the process's complexity, it is very difficult for an external observer to establish trends derived from its implementation, beyond some anecdotal observations. Therefore, one must question whether the outcome of this process might generate a substantial change in the way a group experiences stigma.[121]

In view of this latter critique, prohibiting disparate impact merely on the basis of the stigma rationale (in situations in which the segregation-based rationale is unconvincing) calls for a nuanced approach. If the protected group already suffers from a negative social stigma and the stigmatizing message derived from the scoring scheme is highly negative also, action against disparate impact is justified. In other instances, more lenient steps such as greater emphasis on education regarding the destructive nature of stereotypes, the proper understanding of statistics, and the importance of tolerance might suffice.[122] In addition, stigma-based concerns call into question the wisdom of exposing the inner workings of the scoring process to the broad public.[123] Perhaps some of the more subtle forms of disclosure discussed in *The Scored Society* (such as interactive modeling) might prove a

---

118. For an example of such a setting, see Latanya Sweeney, *Discrimination in Online Ad Delivery*, 56(5) COMM. OF THE ACM 44 (2013), *available at* http://privacytools.seas.harvard.edu/files/privacytools/files/p44-sweeney.pdf.

119. *See* Zarsky, *supra* note 80, at 1562 (presenting a similar argument).

120. Alexander and Cole, *supra* note 107, at 457, make a similar argument regarding covert users of classifications.

121. James Grimmelmann, *Regulation by Software*, 114 YALE. L.J. 1719, 1736 (2005) (explaining that when dealing with software on a case-by-case basis, it might be very difficult for those interacting with it to recognize the existence of systematic discrimination).

122. On the importance of explaining the proper meaning of statistical inferences and how they could be misunderstood in the age of data mining and data analysis, see Zarsky, *supra* note 65, at 1565.

123. For a similar argument, see *id.* at 1567.

better fit.[124] In such cases, individuals can gain insights to the scoring process without aggravating existing stereotypes.

Beyond the disclosure measures noted above,[125] two additional steps might be considered to mitigate disparate impact in scoring. First, the scoring factors which contribute to the creation of a disparate impact (i.e. disproportionately indicating minorities) could be removed from the overall score applied or the databases used. Second, the factors within the datasets used or the scores applied could be recalibrated in a way which will limit the disparate impact resulting from analyzing and using them.[126] Properly understanding both of these steps requires a discussion which goes beyond this Article. Yet such a discussion is nonetheless crucial, as computer science literature has already begun addressing these options.[127] Legal scholarship must move to introduce its perspective regarding these matters.[128] Indeed, both technical steps raise difficult questions. The first step might render the entire scoring process unusable, as over time all the relevant factors are removed from it, thus diminishing its predictive power. The second step might be considered as an illegal form of racial or other discrimination,[129] this time discriminating against the majority groups who find the scoring mechanisms deliberately tilted in their disfavor.[130]

I conclude with two important insights. First, I provide a quick reference to the "wrongs of the past" justification to prohibit

---

124. *See supra* text accompanying notes 21–22.

125. *See supra* text accompanying notes 19–23.

126. *See* Romei & Ruggieri, *supra* note 94, at 39–40.

127. In their bibliographical review of these issues, Andrea Romei and Salvatore Ruggieri sum up the current strategies examined in the computer science realm as follows (while referring to a variety of sources noting these studies): "We categorize three non mutually-exclusive strategies toward discrimination prevention: (i) a controlled distortion of the training set (a pre-processing approach); (ii) a modification of the classification learning algorithm (an in-processing approach), by integrating anti-discrimination criteria within it; (iii) a post-processing of the classification model, once it has been extracted, to correct its decision criteria." Andrea Romei & Salvatore Ruggieri, *Discrimination Data Analysis: A Multi-disciplinary Bibliography*, *in* DISCRIMINATION AND PRIVACY IN THE INFORMATION SOCIETY: DATA MINING AND PROFILING IN LARGE DATABASES 109, 122 (Bart Custers et al. eds., 2013) (citations omitted).

128. The first steps of doing so are carried out by Barocas and Selbst, *supra* note 13, at 51.

129. For instance, it is possible that such actions must be carried out by courts, but not private entities on their own. *See* Primus, *supra* note 56, at 1369. Another factor in the legality of these processes is whether there is a visible victim of discrimination when recalibrating these parameters. *See id.* at 1369. It is indeed possible that when tinkering with the scoring mechanisms no such visible victim exists, and therefore such actions are permissible.

130. *See id.* at 1362 (explaining how a recent Supreme Court decision might be understood to state that "any operation of the disparate impact standard is an equal protection problem," thus meaning that it constitutes discrimination against the majority group).

discrimination. A possible reason to prohibit scoring that generates disparate impact is the fact that it is unacceptable to adversely treat a specific minority group in view of wrongs inflicted against that group in the past.[131] This justification applies even if no stigma will attach or segregation will occur. Yet this theory raises serious questions and might have only limited implications.[132] It should indeed prove relevant when the implications of disparate impact are dire, and the previous wrongs are the reasons for the predictive correlations which indicate minorities as higher risks (for instance, wrongs of the past led to the group's disenfranchisement and thus current poor track record in employment and credit). In such cases, adverse disparate impact should be prohibited and the allocation mechanisms must be recalibrated to avoid this outcome. The stronger groups within society must all contribute to the minority group's quick compensation and rebuilding.[133] Yet fully developing this argument calls for a separate analysis.

Second, I note the importance of gathering sensitive data in the scored society. A recurring element throughout this discussion of discrimination is the need to examine and test whether disparate impact is occurring. Understanding the scope of disparate impact is essential either to battle it as a phenomenon directly or to recognize it as an indication of a form of implicit, yet intentional discrimination. However, discovering such forms of discrimination calls for comparing the way scoring systems treat various forms of groups (for instance minorities as opposed to other social segments), while using sensitive factors (i.e. race) as part of the comparison. To conduct such analysis, the auditing entity—be it the government, the scoring firm, or a third party—must obtain a dataset which includes a vast amount of personal and identifiable information regarding individuals, including sensitive information about them (i.e. their race, religion, etc).[134]

---

131. In the context of gender and race, see SCHAUER, *supra* note 35, at 153 n.28. This argument is rejected by Alexander in a similar context. *See* Alexander, *supra* note 38, at 187–89 (noting that past wrongdoings do not "appear to bear on the morality of acting on present ordinary preferences," beyond the notion that those subjected to past wrongs should receive reparations). *But see* Primus, *supra* note 102, at 523–32 (noting "integrating the workplace" as a possible justification for the disparate impact rules under Title VII).

132. *See* Alexander, *supra* note 38, at 189.

133. For a discussion of this issue, see SCHAUER, *supra* note 35, at 153.

134. Cynthia Dwork & Deirdre K. Mulligan, *It's Not Privacy and It's Not Fair*, 66 STAN L. REV. ONLINE 35, 37 (2013), http://www.stanfordlawreview.org/sites/default/files/online/topics/DworkMullliganSLR.pdf ("Protecting against this sort of discriminatory impact is advanced by data about legally protected statuses, since the ability to both build systems to avoid it and detect systems that encode it turns on statistics.").

Allowing the structuring of such sensitive databases generates both privacy and security concerns. There are fears that information related to racial and other sensitive groupings will be abused, used for manipulation, or hacked and thereafter fall into the wrong hands. Therefore, obtaining such datasets as part of the auditing or even database structuring process is a liability of its own. However, the newness of scoring processes brings about unique discrimination-related issues. The likelihood that discrimination will unfold renders the collection and usage of such datasets of sensitive information a social necessity.[135] Rather than blocking the collection and usage of such data, law should assure that the availability of this information will not end up exacerbating liberty-related concerns rather than mitigating them.[136]

As this extended analysis of discrimination concerns related to protected groups comes to a close, note Table I below, which summarizes the forms of discrimination addressed throughout this section.

Table I: *Discrimination − "Protected Groups" (such as race)*

Explicit Discrimination	**Implicit Discrimination**	**Disparate Impact**
	"Masking"	"Hidden Intent"
	Subconscious Tendencies	Social Segregation/ Social Subordination Examples: *credit, employment*
	Reliance on Tainted Datasets	Stigmatizing Examples: *marketing, insurance.*
	"Blatant Proxies"	[Wrongs of the Past]

135. *See* Reva Siegel, *Equality Talk: Antisubordination and Anticlassification Values in Constitutional Struggles Over* Brown, 117 HARV. L. REV. 1470, 1471–72 (2004) (noting that propositions for limiting the collection of such personal racial information have been struck down and further noting that courts have not struck down rules requiring the collection of such data). For an additional discussion of this proposition, see FORD, *supra* note 53, at 334.

136. For a recent case, where the lack of a proper methodology for measuring disparate impact led to the failure of this claim, see *EEOC v. Kaplan Higher Educ. Corp*., 748 F.3d 749 (6th Cir. 2014).

III. DISCRIMINATION AND SCORING BEYOND PROTECTED GROUPS: A PRELIMINARY EXPLORATION

Even given the vast advances of the recent generation, it is far too early to declare the arrival of a post-racial society.[137] Many of the "classic" forms of discrimination which are premised upon or pertain to protected groups and minorities are still very much alive. However, engaging in a mental exercise to examine the meaning, troubles, and acceptable boundaries of discrimination beyond the notion of race—or other protected groups—is certainly important. Indeed, the scored society introduces novel and unique discrimination-related concerns. These concerns require the implementation of a separate set of policy steps, even if all the previously discussed issues are resolved.

The notion of applying the term "discrimination" to a policy discussion that goes beyond race and other protected groups is not common practice. Scholars and policymakers considering the notion of discrimination (especially in the legal realm) quickly gravitate to a discussion which pertains to race. They will therefore examine whether the *factor* used to distinguish among individuals was race-related or whether the overall impact of the sorting process disadvantaged a racial minority group. Yet there are also other contexts in which the notion of discrimination comes to mind and protected groups are not involved.

Discrimination-related arguments can take many forms in the scored society. Throughout their article, when contemplating the notion of discrimination, Professors Citron and Pasquale set forth several arguments which pertain to this general discussion as well: the two most dominant being the "negative spiral,"[138] and the "arbitrariness-by-algorithm"[139] arguments. While the former focuses on the outcome of the scoring dynamic, the latter relates to concerns with the process.

First is the "*negative spiral*" argument.[140] The premise of this argument is that the scoring process generates extremely negative outcomes to some people, which are disproportionate to the actual differences among the individuals. The individual experiences this effect as a consequence of being a subject of a negative prediction. These dynamics will not be self-corrected, as they are misunderstood by the analysts studying the feedback of the scoring practices as mere reassurance of the scoring system's precision. To explain, consider a

137. For a critical discussion of this term, see FORD, *supra* note 53, at 338.

138. Citron & Pasqual, *supra* note 1, at 32–33.

139. *Id.* at 24.

140. *Id.* at 32–33.

situation in which individuals are provided with a relatively low score and subsequently treated adversely. The low score was allocated in a response to a minor transgression, which results in a prediction regarding the individual's limited ability to perform adequately in some future context. This low score leads to limited chances of obtaining employment, credit, or insurance—which are, at first, proportionate outcomes.

However, the negative impact of the low score is not confined to these specific contexts or to a single event or time, which the individual can ultimately set aside and move forward. In many cases, Professors Citron and Pasquale argue, the effects of such lower scores have severe repercussions on the individual's overall economic status and future prospects.[141] This is due to the fact that one unfortunate event leads to another; for instance, the inability to obtain credit leads to other financial and social pitfalls, such as the inability to secure employment. The predictive analysis finding that a one-time encounter leads to classifying an individual as a high risk ends up proving accurate, as the latter negative outcomes indicate that the risk indeed materialized. Furthermore, the model's predictive success is noted and the process is reinforced and continued. Thus, others will be similarly impacted.

Yet a closer look at what has just been described should lead to the conclusion that the precise prediction did not result from the predictive accuracy of the scoring process. Rather than predicting, it is the score itself that generated the outcome and initiated a self-fulfilling prophecy; the granting of the low score itself led to the outcome the scoring system was supposed to predict! The correlation this process involved is non-spurious, but the causation behind it is problematic. The scoring process is not merely measuring and predicting, but causing the effects measured. The "negative spiral" concern could also be articulated in terms of discrimination; an individual ends up suffering a problematic outcome as a consequence of the scoring system, while being treated unequally to others within his peer group. To some extent, it resembles the "segregation" argument against disparate impact—yet here the individual is a group of one.

Note that the "negative spiral" need not be necessarily linked to the "classic" forms of discrimination which involve minorities and protected groups. Of course, when those adversely affected by this dynamic are part of a protected or historically disenfranchised group, these problems are compounded. However, even when removing this issue from racial

141. *Id.*

or other protected contexts, the basic problematic outcome still follows; a group of individuals are adversely impacted as a result of the scoring process, and treated differently from their peers.

Conceptualizing the potential "negative spiral" dynamic in discrimination terms is a helpful analytical exercise. Such conceptualization reveals that this concern might *not* be as severe as it initially sounds. As explained above, discrimination is considered a severe problem when it *systematically* impacts an entire *group*. Group discrimination removes the individual, which is part of the group, from the sources of power within society. His or her social contacts are powerless to assist, as they are inflicted with the same problems as well. In addition, harming the group impacts the individual's self-confidence and generates other psychological harms, given the individual's affiliation with said group.[142] In other words, the severity of group discrimination results from negative dynamics unfolding in several dimensions—economic, social and psychological.

With this understanding in mind, one must conclude that the damages inflicted on individuals by the "negative spiral" dynamics are substantially different than those caused by the discriminatory practices addressed above (which pertained to "protected groups"). The situation at hand inflicts harm on specific individuals, almost randomly. The individuals are indeed part of a group, but one that is synthetic by nature, of unclear boundaries and structured by an algorithm. It is not a group that the individual feels a strong affinity to, or that the public easily identifies as such.[143] Therefore, group stigma need not attach, and the psychological damage caused will not be as negative. In addition and perhaps more importantly, others within this individual's social group are not part of this negative dynamic. Thus, those harmed will have someone to turn to for help in stopping this downward spiral.[144] These arguments cannot justify the wrongs caused to these individuals, but can strengthen the case that scoring must continue, nonetheless, while other steps are taken to soften their blow. In other words, the "negative spiral" does not feature outcomes that are substantially detrimental as to prohibit the actions of scoring firms, when these firms are meeting reasonable standards and exercising their legitimate preferences. After all, the firm's actions in this context do not feature the morally unacceptable intentions discussed in the previous section.

---

142. Alexander, *supra* note 38, at 185.

143. *Id.*

144. For a similar argument in the context of "general" discrimination theory, see Balkin, *supra* note 112, at 2359–60.

Regulators and scoring entities can take steps to limit this concern. However, it is questionable whether disclosure of the inner workings of the scoring process—the central remedy noted above—will assist in mitigating this problem. The statistical measures which led to the spiral and were used in scoring might prove sound to an external reviewer. A possible regulatory response might call for assuring that causation studies establish the relevancy of all factors used prior to launching scoring schemes. Note, however, that such an aggressive step might encumber the entire process, and is therefore not advisable.[145] Yet the "negative spiral" dynamic might be mitigated by taking a different approach, for instance by assuring that independent scoring mechanisms are applied in different contexts. In such a case, errors in one context are not compounded. Indeed, the discrimination literature notes that severe problems mostly arise when the same form of discrimination is exercised in a variety of settings.[146] Furthermore, encouraging greater competition between scoring systems in every sector will also limit this spiraling dynamic, for similar reasons. Different firms might apply different scoring systems and thus lead to varied results. Unfortunately, we are witnessing the opposite dynamic of growing concentration in scoring. Scoring carried out in terms of credit is spreading to contexts of insurance and employment as well.[147] Thus, recent laws limiting such horizontal spreading of scoring seem to be a step in the right direction.[148] Finally, in important instances, such as those involving healthcare, government will need to assure that the scoring mechanism does not undermine the individuals' ability to gain access to essential social needs, thus halting the negative spiral. To conclude, conceptualizing the "negative spiral" concern in discrimination terms is helpful in ultimately rejecting it as one that generates a severe, unique and incurable problem with the scoring mechanism, in the event the solutions noted are applied.

The second form of discrimination-based argument addressed in *The Scored Society* is that of "arbitrariness-by-algorithm."[149] This argument, partially attributed to FTC Chairwoman Edith Ramirez (who also refers

---

145. For a discussion as to the problems and limited utility of requiring that causation theories are to be established prior to including them in a predictive model, see Zarsky, *Transparent Predicitions*, *supra* note 65, at 1562, 1567; SCHONBERGER & CUKIER, *supra* note 19, at 61–72 (2014) (discussing the limited role causation is destined to play in the age of "big data").

146. Alexander, *supra* note 38, at 198.

147. *See* Lea Shepard, *Seeking Solutions to Financial History Discrimination*, 46 CONN. L. REV. 993, 1003–04 (2014).

148. *Id.* at 1011.

149. Citron & Pasquale, *supra* note 1, at 24.

to it as "Data Determinism"),[150] states that the scoring dynamic is problematic because it judges individuals based on what inferences and correlations suggest they *might* do, rather than for things they have *actually* done. In the scored society, individuals who are similar in all relevant aspects are *treated* differently. Scoring relies upon considering anecdotal differences, which do not translate into actual differences in reality. Or, from another perspective, the concept of equality is constantly violated by arbitrarily grouping together *different* individuals and treating them *similarly*. The fact that evidence constantly circulates as to these outcomes might render the firm reckless in its disregard to treat equal individuals equally. (I set aside the thorny question of whether private firms operating in a public sphere are even subjected to a duty to do so.) Note that this argument does not necessarily call for examining the negative systematic outcomes the process might inflict.

The severity of this concern is substantially undermined upon returning to this Article's foundational assumptions regarding the scoring process and its accuracy. This Article assumes that the scoring process is premised upon non-spurious (yet non-universal) correlations. Such correlations feature imperfect generalization. Given their nature, in some instances, the scoring will predict that individuals will take actions they ultimately would not have carried out. Yet the fact that a ranking process includes errors does not necessarily render it discriminatory if the errors are random and reasonable.[151] Indeed, any system providing for the allocation of benefits and burdens will include errors. It is unclear that a scoring system is more error-prone than any other. Even a system which allows for considering every individual on the merits will lead to mistakes, and there is no assurance that the mistakes of this latter process will be less severe.

Therefore, operating under the assumptions mentioned and taking a comparative view, the scored society is not necessarily drastically random or arbitrary, and thus unfair. It merely appears to be. The lack of clear and apparent reasons for actions, when joined with the fact that the process is guaranteed to produce errors, generates popular discontent.

---

150. Edith Ramirez, FTC Chairwoman, Keynote Address at the Tech. Policy Inst. Aspen Forum, The Privacy Challenges of Big Data: A View from the Lifeguard's Chair 7–8 (Aug. 19, 2013), *available at* http://www.ftc.gov/sites/default/files/documents/public_statements/privacy-challenges-big-data-view-lifeguard%E2%80%99s-chair/130819bigdataaspen.pdf.

151. Indeed, the morality of the process is compromised when the errors are not distributed equally and a specific social segment, such as the poor, is subject to a higher rate of errors. *See, e.g.*, Kate Crawford, *Think Again: Big Data*, FOREIGN POLICY 4 (May 9, 2013, 3:49 PM), http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data (discussing Boston's Street Bump App).

This is as opposed to other, accepted, courses of action which are premised on human discretion. In these latter situations the process is socially acceptable, as the existence of errors is not necessarily apparent. Yet the problem that equals will be treated differently exists with both options.

There is no real reason for society to concede to the visceral discontent scoring generates, set policy accordingly, and reject the process altogether. The discontent possibly associated with the so-called arbitrary nature of scoring might result from the fact that the process inherently and ex ante accepts that it will include substantial errors; errors which will cause real people severe harm. Yet what the reader must bear in mind is that alternative strategies will nonetheless introduce mistakes (a point FTC Commissioner Ramirez recognizes as well)[152]—and even errors that are more frequent, or systematic by nature.[153] Therefore, this latter discrimination-based argument does not carry sufficient analytical weight to transform it from a popular complaint to a legitimate point in a policy discussion. It also does not entail any perceivable intrinsic moral wrong, or evidence of any insult.

Rather than insulting, the process of scoring might lead to some psychological benefits. For this point, I note an interesting argument set forth by Professor Fred Schauer in the similar context of profiling, which might have some relevance to the current discussion.[154] Schauer explains that when an individual is indicated as a higher risk in a process which treats her differently based on her actions exclusively and directly, she may experience severe psychological anguish. However, when an individual is indicated as a higher risk by a profile which considers her as part of a larger group, she need not suffer such psychological harm. Rather, the individual might convince herself that attributing the negative score to her is merely a technical (even arbitrary) error resulting from wrongfully configuring the group's traits. The same might also be said of decisions-via-algorithms. The seemingly arbitrary scoring process might generate anxiety for some, but comfort for many others, given the ability to reflect the blame of failure.

The effects Schauer notes are merely psychological. Yet they are destined to have a broad impact, as they pertain to *all* individuals who receive a lower score, not only those treated unfairly (who might be

152. Ramirez, *supra* note 150, at 8.

153. For a discussion of the importance of considering alternatives of automated selection processes so to properly balance the understanding of the issues at hand, see Tal Z. Zarsky, *Governmental Data Mining and Its Alternatives*, 116 PENN. ST. L. REV. 285, 310 (2011).

154. SCHAUER, *supra* note 35, at 290–91.

quite limited in number). The psychological effects (or rather, lack thereof) impact the individual's ability to set aside a specific financial setback that results from a negative score (be it merited or mistaken) and move onward.[155] Thus, this psychological response might have a substantial positive real world effect. It presents a strong argument as to the benefits of the so called-arbitrary process scoring brings about when compared to other individualized alternatives. Again, it is the introduction of the vast literature on discrimination into the big data and scoring discussion which brings this interesting argument to light.

In addition to the "spiral" and "arbitrariness" arguments, other forms of discrimination might manifest in a scored society. Scoring might generate novel negative stereotypes and stigma attaching to social groups and personality traits. It might also discriminate on the basis of immutable personal traits, thus creating an additional set of discrimination-based concerns. These are both intriguing issues, yet beyond the current discussion.

CONCLUSION: SCORING AND ANTIDISCRIMINATION

The rise of the scored society brings several discrimination-based concerns to mind. This Article takes initial steps in articulating these issues, while striving to connect novel technological and business practices to the already existing vast literature on the troubles, harms of and response to discrimination. The Article shows that beyond mere intuition, and as argued by Professors Citron and Pasquale, these new practices generate unique concerns which must be urgently considered.

Given the evidence as to the connection between scoring practices and discrimination, this Article could lead to two conclusions. First, one can argue that engaging in scoring methods should be categorically prohibited, or at least substantially limited. Second, one might note that, in some instances involving scoring, discrimination-based concerns may arise, but these can be properly countered and balanced through specifically-tailored responses. This Article's central premise is that the latter set of conclusions should be applied. Designed and understood correctly, the discrimination-based concerns are manageable even in view of scoring.

Beyond catering to our general social aspiration to promote science and efficiency, scoring should be maintained and possibly even promoted for yet another reason. Elsewhere I have forcefully argued that

155. *Id.*

to some extent, scoring systems premised upon automated predictive modeling can *limit* severe discrimination-based concerns, especially those involving race. This latter form of discrimination is often practiced subconsciously. Yet automated decision-making systems provide the ability to limit the effects of human discretion, and the prejudices it involves, on the decision-making processes.[156] In addition, a standardized process carried out uniformly through the use of computer software and hardware would be easier to supervise. Here, the scoring entity can assure that the officials applying the score are not abusing their discretion to harm minorities, but rather sticking to their scripted protocol. Thus, given these two important attributes, the scored society can potentially prove to be a fair and equal one.[157] This will only occur, however, if the new forms of discrimination here addressed are fully acknowledged, and thereafter properly and quickly resolved.

156. Zarsky, *supra* note 5, at 27–29. *But cf.* Citron & Pasquale, *supra* note 1, at 4 (disagreeing with the argument voiced in the text).

157. Others have provided additional reasons why racial discrimination concerns might diminish in the age of big data—most notably Lior Strahilevitz. *See generally* Lior J. Strahilevitz, *Privacy Versus Antidiscrimination*, 75 U. CHI. L. REV. 363, 364 (2008) (explaining that increasing the availability of negative information regarding *individuals* will reduce decision makers' reliance on information regarding groups, which in turn often constitutes racial statistical discrimination). For a review of this strand in the literature, see Scott R. Peppet, *Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security & Consent*, TEX. L. REV. 34 n.191 (forthcoming 2014), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2409074.