The Fallacy of Equating “Blindness” with Fairness:
Ensuring Trust in Machine Learning Applications to Consumer Credit

by

Grace M. Abuhamad

B.A. Middle Eastern Studies
Wellesley College, 2013

Submitted to the
INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE IN TECHNOLOGY AND POLICY

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

© 2019 Massachusetts Institute of Technology. All rights reserved.

Signature of Author:

Institute for Data, Systems, and Society
May 15, 2019

Certified by:

Daniel J. Weitzner
Principal Research Scientist, CSAIL
Thesis Supervisor

Accepted by:

Noelle Eckley Selin
Director, Technology and Policy Program
Associate Professor, Institute for Data, Systems, and Society and
Department of Earth, Atmospheric and Planetary Sciences
The Fallacy of Equating “Blindness” with Fairness: Ensuring Trust in Machine Learning Applications to Consumer Credit

by

Grace M. Abuhamad

Submitted to the Institute for Data, Systems, and Society on May 15, 2019 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Technology and Policy

Abstract

Fifty years ago, the United States Congress coalesced around a vision for fair consumer credit: equally accessible by all consumers, and developed on accurate and relevant information, with controls for consumer privacy. In two foundational pieces of legislation, the Fair Credit Reporting Act (FCRA) and the Equal Credit Opportunity Act (ECOA), legislators described mechanisms by which these goals would be met, including, most notably, prohibiting certain information, such as a consumer’s race, as the basis for credit decisions, under the assumption that being “blind” to this information would prevent wrongful discrimination. While the policy goals for fair credit are still valid today, the mechanisms designed to achieve them are no longer effective. The consumer credit industry is increasingly interested in using new data and machine learning modeling techniques to determine consumer creditworthiness, and with these technological advances come new risks not mitigated by existing mechanisms.

This thesis evaluates how these “alternative” credit processes pose challenges to the mechanisms established in the FCRA and the ECOA and their vision for fairness. “Alternative” data and models facilitate inference or prediction of consumer information, which make them non-compliant. In particular, this thesis investigates the idea that “blindness” to certain attributes hinders consumer fairness more than it helps since it limits the ability to determine whether wrongful discrimination has occurred and to build better performing models for populations that have been historically underscored. This thesis concludes with four recommendations to modernize fairness mechanisms and ensure trust in the consumer credit system by: 1) expanding the definition of consumer report under the FCRA; 2) encouraging model explanations and transparency; 3) requiring self-testing using prohibited information; and 4) permitting the use of prohibited information to allow for more comprehensive models.

Thesis Supervisor: Daniel J. Weitzner
Title: Principal Research Scientist, CSAIL

This work was partially supported by the MIT-IBM Watson AI Lab and the Hewlett Foundation through the MIT Internet Policy Research Initiative.
# Table of Contents

Abstract 3

Acknowledgements 5

Executive Summary 6

Preface 7

Introduction 10

I. Research on Fairness in Machine Learning 16
   Sources of Unfairness in Machine Learning 16
   Proposed Remedies and their Trade-Offs 20

II. Background | Consumer Credit Context 31
   Procedural Fairness in the Fair Credit Reporting Act 33
   Substantive Fairness in the Equal Credit Opportunity Act 36
   Proving Wrongful Discrimination 39
   Revisiting a Twenty-Year-Old Proposal 43

III. Assessing “Alternative” Credit Today 46
   Categorizing “Alternative” Data 49
      Payment Data 50
      Personal Data 51
   Proposed Framework 53
   Model Examination Procedures 55

IV. Beyond “Blindness” and Other Outdated Mechanisms 58
   Procedural Fairness 60
      Expand the Definition of Consumer Report 61
      Encourage Model Explanations 62
   Substantive Fairness 64
      Require Self-Testing Using Prohibited Information 65
      Permit the Use of Currently-Prohibited Information 68

Conclusion 72

References 75
Acknowledgements

This work was partially supported by the MIT-IBM Watson AI Lab and the Hewlett Foundation through the MIT Internet Policy Research Initiative (IPRI). I am grateful to have the IPRI’s Founding Director, Daniel J. Weitzner, as my thesis advisor. He and Hal Abelson introduced me to the idea of Internet policy in a course\(^1\) I took with them as an undergraduate student, and they are the people who drove my choice in MIT for graduate school.

The IPRI has been my research home at MIT for the past two years, and its faculty, staff, and students all deserve my thanks for their support and encouragement throughout what were two very challenging years for me. In particular, I would like to note Julius Adebayo’s daily advising and support, which he somehow managed --always with perceptive humor-- amidst his own commitments as a first-year PhD student. In many ways, Julius picked up the baton from Jonathan Frankle, who spent the year prior, and significant parts of this year, teaching me to code and tutoring me through my technical courses. Over the course of both years, Ilaria Liccardi has been my guardian angel and my greatest advocate. David D. Clark guided me as if I were a student advisee of his own, even though I deviated from his areas of expertise. Lastly, Taylor Reynolds and Melanie Robinson kept many pep talks on hand for the tougher days.

With regard to this project, R. David Edelman had the insight to focus on consumer credit, an area with which I had no prior familiarity. Mikella Hurley used her limited personal time to brainstorm with me on the legal questions that arise in the recommendations, though I am solely responsible for their insufficient presentation. Jolina Cuaresma and Lily Fitzgerald provided valuable feedback on drafts in the midst of their own academic and research deadlines. I am also grateful to the following researchers and practitioners who might not recognize themselves as direct contributors to this thesis, but whose curiosity, teaching, and guidance inspired me in the process: Solon Barocas, Susan P. Crawford, Joi Ito, Howell E. Jackson, Margaret E. Tahyar, Antonio Torralba, and Jonathan L. Zittrain.

There are many family and friends who care for me irrespective of my professional success or lack thereof. They will go unlisted here, to protect their privacy and mine.

\(^1\) I took the course in Fall 2012. Information about the course, now named “Foundations of Information Policy,” is available at: [http://groups.csail.mit.edu/mac/classes/6.805/index.html](http://groups.csail.mit.edu/mac/classes/6.805/index.html).
Executive Summary

Deployments of decision-assistive tools across sectors call into question whether existing fairness frameworks are adequate, need modification, or need a complete overhaul. Since these tools can reflect or surface discrimination, their widespread integration creates a policy imperative to ensure trust by protecting consumers and reducing the risk of harm. The consumer credit system is a salient case study because it is a requirement of participating in the American economy. With as little as paying a bill, consumers are swept into a system of data collection that forms the foundation on which they are measured and given access to opportunities. An estimated 15% of consumers, primarily from marginalized communities, are underscored by the credit system. New data and modeling may promote financial inclusion.

The Fair Credit Reporting Act (FCRA) and the Equal Credit Opportunity Act (ECOA) envision a fair credit system that is equally accessible by all consumers, and developed on accurate and relevant information, with controls for consumer privacy. In addition to these two goals that ensure trust in the consumer credit system, a third goal is to expand credit and be more financially inclusive. In the context of “alternative” data and modeling techniques, these fair credit goals are still valid, but the mechanisms by which to achieve them are no longer.

“Alternative” data and models allow for more comprehensive models that promote fair credit goals. At the same time, these data and models also facilitate inference or prediction of consumer information, which may not be in compliance with the current mechanisms implementing the FCRA and the ECOA. Instead of limiting the technology, this thesis proposes modernizing some of these mechanisms. For example, the ECOA prohibits discrimination in credit decisions on the basis of certain protected characteristics, such as race. In implementing the ECOA, regulators interpreted this rule as prohibiting the use of protected information at every stage of the credit process. This “blindness” mechanism is ineffective in the wake of “alternative” credit processes since information can be inferred or predicted, and since models could be more comprehensive with this information.

This thesis concludes that updated mechanisms are necessary to keep pace with technological change, ensure trust in the consumer credit context, and usher in a new era of financial inclusion and credit expansion, including:

- Requiring model self-testing using prohibited information as correlatory benchmarks would provide creditors with more information about possible discriminatory inferences in their models;
- Permitting the use of prohibited information in models would allow for more comprehensive model design and may encourage creditors to curb systemic discrimination through product or service offerings; and
- Expanding the definition of consumer report under the FCRA and encouraging model explanations and transparency could increase trust in the consumer credit system.
Preface

“The impact of technology has been across the board, and we haven’t yet really absorbed how it’s going to change the way we do business.”

In a 1985 essay on computer ethics, James H. Moor described how advancements in technology require us to return to fundamentals of societal structure: “The question is no longer ‘How efficiently do computers count votes in a fair election?’ but ‘What is a fair election?’” In 2014, with the rise of machine learning, researchers founded what is now an annual event on Fairness, Accountability, and Transparency of Machine Learning (FAT/ML), echoing claims by Moor and others that, “finding a solution to big data’s disparate impact will require more than best efforts to stamp out prejudice and bias; it will require a wholesale examination of the meanings of ‘discrimination’ and ‘fairness.’” This research contributes to this broader examination of the impact of technology on societal values and structures, and vice versa.

Recognition of transformational effects of computing and data extends outside the realms of research, and into the public and private sectors, both in the United States and internationally. Since the mid-1990s, and perhaps before, each United States Administration has produced reports discussing broad effects of computing and data across the American economy, including “the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace.” Herein lies a dilemma of technology control, also known as the Collingridge dilemma: reviewing fundamental concepts, such as fairness, on which our society is built, is a task that requires

---

deliberation and an understanding of the effects which we seek to manage. And yet, effects of technological change can occur simultaneously, and can be harmful if not managed.

The aforementioned “potential to eclipse longstanding civil rights protections” has been realized on more than one occasion, decreasing trust in technology, or encouraging efforts to ban it altogether. A 2016 ProPublica report on bias in criminal justice risk assessment tools leading to wrongful discrimination sparked an ongoing debate about whether and how such tools should be used, especially in contexts that can impact the trajectory of an individual’s life.10

Increasingly, companies have deployed decision-assistive tools that use machine learning and other statistical methods to process massive amounts of data. In examining these decision-assistive tools, recent reports have demonstrated discrimination in housing, criminal justice, and hiring.11 These tools are designed to increase efficiency, and in some cases, reduce human bias. However, these tools are built using data that reflect or surface discrimination, and inadvertently reproduce it instead of eliminating it. Once the harm of these tools is exposed, stakeholders can take action, leading to, for example, the U.S. Department of Housing and Urban Development (HUD) charging Facebook with discrimination in housing advertisements,12 or Amazon recalling its hiring tool.13 Consumers can manage, to some extent, their exposure to such tools, although that control may diminish as these tools become more popular. The widespread integration of decision-assistive tools creates a policy imperative to protect consumers and reduce risk of harm. These examples leave us with a variety of tough policy choices, including whether to pre-empt technological advances knowing that our forecasting abilities are imperfect, whether or not to wait, and if so, for how long, and whether or not to accept some harm, and if so, how much.

There is interest in deploying decision-assistive tools in the consumer credit context, primarily as a tool for credit expansion and financial inclusion, though there is uncertainty as to how to manage the risks given failures in other contexts. Consumer credit differs from other contexts such as criminal justice in two ways. First, it is broader in scope, affecting almost every adult in the United States, and under-serving an estimated 45 million.14 Second, consumer credit is not voluntary in practice: it is a requirement to, and result of, participating in the American economy. Almost as soon as a consumer receives a bill for payment, this consumer is entered in the credit system, making the impact of the tools much more extensive than those used in other, perhaps less-popular, contexts. Additionally, information about the consumer is collected and shared mostly in a business-to-business environment, making the system opaque and outside of

---

the consciousness of many. The past two decades of legislation in consumer credit have given consumers more rights and access to their credit information, but that sense of participatory control that consumers may have for hiring tools, for example, does not exist for consumers in the credit context. The most recent example of this lack of awareness and control was the Equifax data breach in 2017, where the information of 143 million Americans was compromised, making this the largest data breach to date. Equifax is one of the three nationwide consumer reporting agencies that collect data and use them in their production of consumer reports. Unlike other data breaches to date, those harmed did not consent to their information being collected, they did not “sign up for Equifax,” and yet this information was, and is still, collected, packaged, and sold, and forms a foundation on which individuals are measured and given access to opportunities.

As the adage goes, “Trust takes years to build, seconds to break, and forever to repair.” This research examines assumptions that the consumer credit system in the United States is robust and trustworthy. This is not an effort to question the very premise of consumer credit or the use machine learning. Assuming that machine learning modeling techniques are being deployed in consumer credit and will become more widespread in the near future, this research separates policy goals from the mechanisms designed to implement them, arguing that the fair credit goals are still valid, but that mechanisms may need to change to account for new and unmitigated risks of technological advances.

---

15 Wu, Chi Chi. 2018. Fair Credit Reporting. 9th ed. Boston, MA: National Consumer Law Center. Page 5: According to the Consumer Financial Protection Bureau, “There are about three billion credit reports issued annually in the United States. Only about 44 million of these credit reports go to consumers.”
Introduction

“A poor credit history is the ‘Scarlet Letter’ of 20th century America.”

In the 1950s, the Boston Symphony Orchestra instituted “blind” auditions, primarily as a way to combat implicit bias against women candidates. Musicians would audition behind a screen, so that their gender was not visible to the judges. To the surprise of many, this initial “blindness” policy turned out not to be as “blind” as expected: it did not result in any changes to the gender balance. There were other factors that allowed for candidates’ gender to be inferred by the judges. Candidates’ gender manifested in another data point, the sound of their shoes as they walked behind the screen to audition. Shoe sound, such as the click of high-heels, was a proxy for gender. The consumer credit context also instituted a “blindness” policy to ensure fair and equal access to credit, by prohibiting the use of certain characteristics, such as race and gender, as the basis for credit decisions. However, as is the case with “blind” auditions, there are proxies that allow for these prohibited bases to be inferred or predicted. For example, given the history of redlining in the United States, zip codes can be used as a proxy for race, indirectly exposing information that would otherwise be prohibited from use. While orchestras were able to take measures to eliminate their proxy for gender in auditions by installing carpet to mute the sound of clicking heels on stage, doing so in the consumer credit context is far less attainable. New data and modeling techniques expose so many proxies that no amount of metaphorical carpet can mask the sound of clicking heels. “Blindness” as a mechanism for fair credit is no longer valid, due to technological advances in consumer credit reporting and scoring.

The United States has a credit-dependent economy: access to credit is also access to opportunity. Consumer credit reports and their numerical shorthand, known as a credit score, are primarily designed as indicators of creditworthiness, which is an individual’s expected likeliness to repay a loan. A good report or score indicates that a consumer can access more credit at lower interest rates. For the consumer, this translates as more opportunity for

---

20 See infra note 313.
21 Candidates removed their shoes or walked on carpet so as to mute the sound. See, Rice (2013).
22 See infra notes 24 and 25.
23 Minor, Kendrick A. 2013. The U.S. Credit Reporting System. New York, NY: Nova Science Publishers. Page 53. References to credit scores in this thesis are references to generic scores. There are other types of risk scores, such as the risk score used in criminal justice, or ones developed for specific creditors. Aside from consumer credit reports, there are “investigative consumer reports,” of which a portion summarizes information about the character, reputation, personal characteristics, or mode of living of a consumer. The information for these more detailed reports is obtained through personal interviews with neighbors, friends, and other third parties connected to the consumer. It is important to note that the information cannot simply be observed: there must be an interview of some kind, whether in-person or over the phone, online, etc. See, Wu, 695.
24 Conversely, a low credit score does not necessarily mean that a consumer will be denied a loan, but it generally indicates that a higher interest rate would apply to that loan. See, Solove, Daniel J., and Paul M. Schwartz. 2015. Consumer Privacy and Data Protection. 1st ed. New York, NY: Wolters Kluwer. Page 54.
consumption and wealth accumulation. In some cases, and perhaps increasingly so, consumer credit reports are used to predict more than credit risk: they are used to determine access to rental housing, utilities, premiums for home and auto insurance, and even employment. In the words of the first director of the Consumer Financial Protection Bureau (CFPB), “there is much at stake in ensuring that the credit reporting market is working properly.” Considering the impact of credit scores on an adult’s opportunities in life, lawmakers have developed an extensive set of regulations that promote fairness and equal opportunity in the credit system; protect consumers against fraud, identity theft, and violations of their privacy; and educate consumers about their rights.

Like other data-intensive industries, consumer credit is a commercial industry in which the consumer is the product. Private companies collect information about loan repayment history, build profiles of consumers, and sell these profiles. As consumer reporting and scoring become more expansive, and incorporate new data and model types, consumer credit provides a rich context in which to study systemic trust. In this industry, there is a long history of trust-building through fairness frameworks motivated by financial inclusion, especially in 20th century America. Following abuses of middle- and low-income consumers in the early 1900s, the United States embarked on a quest for consumer protection by establishing credit unions and creating the Federal Trade Commission (FTC) in the 1920s, passing level-setting legislation in the civil rights era, and most recently, by creating the CFPB, an agency dedicated to consumer financial protection. And even today, the quest is ongoing. While the aforementioned reforms have increased access to credit for many, there are still an estimated 45 million people in the United

---

25 According to estimates from the Credit Builders Alliance, having a low credit score can cost an individual up to $250,000 extra over his or her life. See, Credit Builders Alliance. “CBA reporter.” https://www.creditbuildersalliance.org/reporter.
26 For example, employers can check consumer credit reports, with the consent of the individual in question, and even though there is no statistical correlation between job performance and credit risk, credit bureaus do not dissuade employers from, and in some cases even encourage, doing so. See, Eric Rosenberg testimony in Oliver, John. 2016. “Credit Reports: Last Week Tonight with John Oliver” HBO via YouTube. https://www.youtube.com/watch?v=aRrDsbUdY_k. See also, White House Office of Science and Technology Policy (2014).
27 Minor, 53.
28 Planet Money.
29 Lauer, Josh. 2017. Creditworthy: A History of Consumer Surveillance and Financial Identity in America. New York, NY: Columbia University Press. Page 211, on the predictions made in the 1970s about the expansion of credit risk scoring: “We will see point tables to make decisions such as: lease / don’t lease, promote / don’t promote, insure / don’t insure, hire / don’t hire, open / don’t open a checking account, and so forth.”
30 Consumer Financial Protection Bureau. 2017. “Request for Information Regarding the Use of Alternative Data and Modeling Techniques in the Credit Process.” On Pages 9 - 10 new modeling techniques include, but are not limited to “decision trees, random forests, artificial neural networks, k-nearest neighbor, genetic programming, 'boosting' algorithms, etc.” New data include, but are not limited to: “Data showing trends or patterns in traditional loan repayment data; Payment data relating to non-loan products requiring regular (typically monthly) payments, such as telecommunications, rent, insurance, or utilities; Checking account transaction and cashflow data and information about a consumer’s assets, which could include the regularity of a consumer’s cash inflows and outflows, or information about prior income or expense shocks; Data that some consider to be related to a consumer’s stability, which might include information about the frequency of changes in residences, employment, phone numbers or e-mail addresses; Data about a consumer’s educational or occupational attainment, including information about schools attended, degrees obtained, and job positions held; Behavioral data about consumers, such as how consumers interact with a web interface or answer specific questions, or data about how they shop, browse, use devices, or move about their daily lives; Data about consumers’ friends and associates, including data about connections on social media.”
States who are underscored by the existing system, and these individuals are primarily from marginalized communities. This means that there is not enough information about them to build a credit profile.

Since most of these adults are from marginalized communities, inequalities in access to credit compound other inequalities they may already be facing. When these consumers are unable to access opportunities linked to their credit, such as their ability to find a place to live, commute, and secure employment, they are forced, in many cases, to resort to obtaining credit through lenders who may use unregulated or predatory practices which can further hinder their ability to access opportunities or the credit market at a later date, leaving them in a perpetual cycle of poverty. To be considered worthy of credit, a consumer needs to demonstrate good credit practices such as paying back loans, but in order to even get these loans at the outset, that consumer often needs to be considered worthy. For consumers with little or no credit history, primarily young and/or marginalized adults, the credit building process can prove to be difficult, creating more barriers for consumers who are already marginalized. Wrongfully limiting access to opportunities is unfair and can erode trust in the consumer credit system.

Given the importance of credit and the potential to better serve consumers, stakeholders throughout the consumer credit industry are considering new ways to increase access to credit. To distinguish the new from the existing, these are often referred to as traditional and “alternative” credit processes. In a 2017 Request for Information, the CFPB noted, “We use ‘alternative’ in a descriptive rather than normative sense and recognize that there may not be an easily definable line between traditional and alternative modeling techniques.” Simply put, “alternative” credit processes are any data and modeling techniques that are not used traditionally, encouraging the use of different data sources and modeling techniques so as to provide new insights into creditworthiness. These “alternative” processes have the potential to redefine the existing system and streamline access to credit, especially for those who have long been marginalized. There are risks associated with these advances, however, that would decrease trust in consumer credit if not addressed. The existing credit system, including its regulatory directives, is designed around specific types of data and models, and there are questions as to whether the existing system can and should withstand the possible changes from the “alternative” credit processes.

31 Brevoort et al.
33 Parker, Geoffrey G., Marshall W. Van Alstyne, and Sangeet Paul Choudary. 2016. Platform Revolution. 1st ed. New York, NY: W. W. Norton. Page 277: “Still another potential source of future growth in the hundreds of millions of ‘unbanked’ people, both in the developing world and in less affluent neighborhoods in the U.S. and other developed countries, who currently have no access to tools that can help them pay their bills, borrow money, save, and make investments. Because they live in areas without bank branches and lack the capital needed to qualify for a traditional bank account or a line of credit, the unbanked are forced to rely on costly, inconvenient, and sometimes fraudulent alternatives like check-cashing services, money order businesses, payday loan companies, and illegal loan sharks. These substandard financial operators represent another barrier to the self-sufficiency that makes it harder for the poor to escape poverty.”
34 Consumer Financial Protection Bureau (2017).
In order to evaluate these questions, Chapter II presents background on the legal and regulatory history that established the existing consumer credit system. In this system there are goals and mechanisms that are designed to help achieve those goals. The overarching goal in the consumer credit system is to ensure trust,\textsuperscript{35} broken down into three sub-goals:

- **Procedural**: the consumer credit system should be built on accurate and relevant information, with controls for consumer privacy. This procedural fairness goal is expressed in the Fair Credit Reporting Act (FCRA).

- **Substantive**: the consumer credit system should be equally accessible by all consumers. The Equal Credit Opportunity Act (ECOA) outlines this vision.

- A third goal is credit expansion and financial inclusion. This goal is a long-standing government policy that took root in credit reforms at the turn of the 20th century, and is the primary driver behind interest in “alternative” processes today.

These goals are developed further in Chapter II, and distinguished from their implementing mechanisms. In the context of “alternative” data and modeling techniques, these fair credit goals are still valid, but the mechanisms by which to achieve them are no longer.

Since there are a myriad of possibilities for “alternative” data and models, Chapter III proposes a framework in which to consider them and the risks they may pose across the credit system. First, the framework focuses the primary consideration for credit processes to be **predictive** of credit risk. While companies may want to consider credit profiles for other applications, their primary purpose is to be predictive of risk. Second, the framework looks at how **compliant** a certain data type may be, so that stakeholders can make decisions about which types of data may be more or less acceptable to use. Compliance suggests a focus on mechanisms, though these mechanisms are outdated. When the framework is populated with “alternative” credit processes and existing mechanisms, these “alternative” processes are not compliant, or at least indicate greater risk and uncertainty in their application to the credit system.

Based on the framework in Chapter III, it may be that “alternative” processes present greater risk and uncertainty. A resulting cautionary approach could challenge the idea in machine learning that more data are better than less, since, as developed in Chapter I, more data generally allow for more comprehensive model design. Including more data, while perhaps allowing for more comprehensive models, may not be in compliance with the FCRA or the ECOA. Compliance would require limiting and controlling much “alternative” data since these data can be proxies for prohibited bases,\textsuperscript{36} meaning that they can be, alone or in combination, highly correlated to information protected from use in credit decisions.\textsuperscript{37} Including these data would therefore violate the ECOA as it currently stands. However, it may also be the case that

\textsuperscript{35} In the findings of the FCRA, Congress described the following: “[... unfair credit reporting methods undermine the public confidence which is essential to the continued functioning of the banking system.” (FCRA § 602 or Pub. L. No. 90-321, tit. VI, § 602.) In this sense, public trust and confidence depend on the public’s perception of fairness in the credit system.

\textsuperscript{36} 15 U.S.C. § 1691 et seq.

\textsuperscript{37} The idea that not all data are created equal, and that some data deserve more protection than others, is not new: this idea is featured as a mechanism of the ECOA. Within consumer credit legislation, there is precedent for protecting against the use of certain data types: these are the “prohibited bases” in the ECOA.
“alternative” processes are being evaluated with outdated mechanisms. Chapter IV suggests another option, to modernize the implementing mechanisms, while maintaining the fair credit goals described above. These outdated mechanisms may not provide enough risk mitigation, or could be preventing the expansion of credit in their limitations. A primary example of this are the ECOA’s prohibited bases, which are so protected that their existence promotes a form of “blind,” yet ignorant and ineffective, justice, especially in the wake of “alternative” credit processes.

This thesis concludes that self-testing of credit scoring systems using protected class information as correlatory benchmarks should be a requirement.\footnote{Using protected information in self-testing is allowed/excepted, but not required.} Perhaps more radically, permitting the use of such information could also allow for more comprehensive model design, allowing policymakers to more directly assess discrimination and unfair credit access issues should they choose to allow, or perhaps encourage, some form of outcome calibration or affirmative action. The latter suggestion to explicitly use protected information in decision-assistive models would have its own set of risks and may therefore be less feasible, but no less important, than the first recommendation to bolster self-testing. Self-testing of models using protected information is done to examine models but is not a credit decision, and thus an exception in the ECOA prohibitions, whereas the use of protected information in credit decisions is currently prohibited in the ECOA. Permitting the use of protected information in decision-assistive models does not mean that the information itself forms the basis of a credit decision, but this is a question dependent on legal interpretation.

Chapters III and IV present two different scenarios analyses, one in which technology develops within an unchanging context, and one in which the context adapts to technological advances. These approaches are reconcilable in that they both focus on ensuring trust in the consumer credit system, and both are aligned with established fair credit goals. The second approach, however, differentiates between goals and mechanisms, recognizes that the existing mechanisms are no longer the most effective way of mitigating risks for advances in “alternative” credit processes, and proposes changes to the mechanisms so that they may be more effective. The tension for policymakers appears to lie in the choice between cautious and effective policy. As technology advances in the face of inaction, the chasm between the choices increases, placing consumers at risk of harm and decreasing trust in the system.

This thesis seeks to provide policymakers and technologists with an understanding of the consumer credit process, its vision of fairness, and its regulatory history for managing technological risks so that they may better consider and inform their policy and design choices. Technologists need to consider how their products, whether through data collection and management tools or through modeling techniques, can be inadvertent sources of unfairness in systems. Such risks may require either a cautious approach to “alternative” data and modeling, or a wholesale review of regulatory imperatives, which may be outside the realm of control for many technologists to tackle alone. For the purposes of this case study, assumptions include that
the consumer credit system in the United States is mostly robust and trustworthy, that “alternative” data and modeling techniques are being deployed in consumer credit and will become more widespread in the near future, and that the United States will remain a credit-dependent economy in the near future. This is not an effort to question the very premise of consumer credit or machine learning technology. As such, this thesis examines fairness in consumer credit as a lever for ensuring public trust in technology and the banking system, and contributes recommendations that align with existing policy goals while being cognizant of technological advances, making it a more realistic formalization, albeit a perhaps imperfect one.
I. Research on Fairness in Machine Learning

“Machine learning has become alchemy... I would like to live in a society whose systems are built on top of verifiable, rigorous, thorough knowledge, and not on alchemy.”

There is a long history of discourse on fairness in political theory, moral philosophy, and other fields, though the primary focus of this Chapter is to review work that addresses fairness in technology. There is ongoing concern about whether individuals are being treated fairly as technology advances: in recent years, with the rise of machine learning, researchers recognized sources of unfairness in data and models and broader societal context, and have been working to remedy these issues in order to promote trust. Since consumer credit is an industry dependent on large amounts of data and modeling, and as the industry incorporates new, or “alternative,” data and modeling techniques, such as machine learning, into their products and services, research that addresses fairness in computing is most relevant to understanding challenges arising in “alternative” credit processes and informing recommendations to ensure systemic trust. This chapter reviews work addressing sources of unfairness in data and models, and proposed remedies, including suggestions for measurements and increased transparency, and their trade-offs.

Sources of Unfairness in Machine Learning

While sources of unfairness external to technologies warrant some attention, this section will focus on the many sources of unfairness in data and modeling. The purpose of this section is to draw attention to the many sources of unfairness that exist in data and models, not just in their design, but also in their contextual application. Most of these sources of unfairness are rooted in human design choices, whether unfairness is intentional or not. Researchers, including some outside the machine learning field, have focused on elucidating and categorizing these sources of unfairness so as to better understand the problem of unfairness. For example, in Winner’s 1986 article “Do Artifacts Have Politics?” he argued that what matters about technologies is the ecosystem in which they are embedded, not necessarily whether there are politics embedded in

30 See description of “alternative” data and models at Consumer Financial Protection Bureau (2017), supra note 30.
the technologies themselves. He drew attention to two ways in which technologies can have politics. First, he discussed technologies that are used to create social order or settle issues in communities (e.g., the bridge designs in New York City). Second, he discussed inherently political technologies (e.g., the atom bomb). He showcased these examples to show that they are reflections of society in decisions to adopt the technologies, regulate their uses, etc., strengthening his argument that what matters most about these technologies is the context in which they are embedded. The same arguments can be made today with computing and machine learning tools. The context in which these tools are applied should be considered in the design of the tools, recognizing that it may not be possible to detach the tools from human values.

**What matters most about technologies is the context in which they are embedded.**

More recently, O’Neil developed this idea that context matters in her book *Weapons of Math Destruction*. She described how FICO credit scoring models, for example, are well-designed to fit the purpose of determining creditworthiness in a consumer credit context, but are not suited for applications in other contexts, such as determining an individual's employability. When tools such as credit scoring models are taken out of the context for which they are designed, they become not only less effective, but also harmful. Eubanks, in *Automating Inequality*, addressed the harm caused by such tools and highlighted how they can perpetuate poverty among the poor and continue to disenfranchise those who, traditionally, have had little power. Her focus was on the allocation of resources, which was thought to be made more efficient with data-based approaches, but was over-promised since the data reflected, and, once modeled, automated existing disparities. Even the best-designed technologies can have the worst impacts when applied in ill-suited contexts. Barocas et al. provided a framework for considering these harms as either allocative or representational: the former is a harm that limits the opportunities of a group, and the latter is a harm that diminishes the identity of a group.

---

42 Winner, Langdon. 1980. “Do Artifacts Have Politics?” *Daedalus* 109 (1). The MIT Press on behalf of American Academy of Arts & Sciences:121–36. [https://doi.org/10.2307/20024652](https://doi.org/10.2307/20024652). Page 122: “Hence, the stern advice commonly given to those who flirt with the notion that technical artifacts have political quantities: what matters is not the technology itself, but the social or economic system in which it is embedded.” See also, White House Office of Science and Technology Policy (2014), 10: ‘The historian Melvin Kranzberg’s First Law of Technology is important to keep in mind: ‘Technology is neither good or bad; nor is it neutral.’ Technology can be used for the public good, but so too can it be used for individual harm. Regardless of technological advances, the American public retains the power to structure the policies and laws that govern the use of new technologies in a way that protects foundational values.”

43 For context on bridge design, see Winner, 123-124: “It turns out, however, that the two hundred or so low-hanging overpasses on Long Island were deliberately designed to achieve a particular social effect. Robert Moses, the master builder of roads, parks, bridges, and other public works from the 1920s to the 1970s in New York, had these overpasses built to specifications that would discourage the presence of buses on his parkways. According to evidence provided by Robert A. Caro in his biography of Moses, the reasons reflect Moses’s social-class bias and racial prejudice. Automobile owning whites of ‘upper’ and ‘comfortable middle’ classes, as he called them, would be free to use the parkways for recreation and commuting. Poor people and blacks, who normally used public transit, were kept off the roads because the twelve-foot tall buses could not get through the overpasses.”


45 O’Neil, 142.


not distinguish between the types of harm, but their work included examples of both types of harm, occurring simultaneously or separately.

**Human values and bias are sources of unfairness in technological design and use.** While context matters, it is not the only source of unfairness. Human values and bias can also be sources of unfairness, whether they are imbued in the design of technologies or, more simply, exist in the human operator or user of the technologies. Kleinberg et al. (2019) described the extent to which humans are biased, consciously and unconsciously, and can be themselves “black boxes” of unexplainable behavior and choices.48 Because of this, implicit bias is difficult to extract in humans, but can be observed in “audit studies,” such as Bertrand and Mullainathan’s employment discrimination study where the authors used identical resumes, with different gendered- and racially-sounding names, to measure bias based on which candidates were deemed more employable.49 Similarly, Agan and Starr concluded, in part, that removing the information about criminal records in employment applications made employment discrimination worse for black Americans, whether they had a criminal record or not.50 Employers had implicit assumptions about an applicant, based on their race, that turned out to be even more discriminatory than had they had more information about the applicant.

**Data collection and use can manifest bias and have unintended consequences.** Human sources of unfairness also exist in data collection and use. Data form an important part of the machine learning and automation pipeline, and can become sources of unfairness if not properly processed prior to, or as part of, inclusion in model design. In 2016, Barocas and Selbst detailed five mechanisms by which data may be the source of unfairness in the context of anti-discrimination law: defining the target variable and class label; training data; feature selection; proxies; and masking.51 The authors noted that there has been much attention paid to

---

48 Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, Cass R Sunstein, and Areeda Hall. 2019. “Discrimination In The Age Of Algorithms.” https://www.fatml.org/resources/relevant-scholarship: “A growing body of research in economics suggests that the picture is complicated, in that suppressing information of one type (criminal record) can incentivize decision-makers to turn to other forms of information (race) in ways that may be overtly discriminatory.” See also, Barocas and Selbst: “Professor Lior Strahilevitz has argued, for instance, that when employers lack access to criminal records, they may consider race in assessing an applicant's likelihood of having a criminal record because there are statistical differences in the rates at which members of different racial groups have been convicted of crimes.”


50 Agan, Amanda and Sonja Starr. 2018. “Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment.” The Quarterly Journal of Economics. Volume 133, Issue 1. Pages 191–235. https://doi.org/10.1093/qje/qjx028. See also, Kleinberg et. al (2019): “A growing body of research in economics suggests that the picture is complicated, in that suppressing information of one type (criminal record) can incentivize decision-makers to turn to other forms of information (race) in ways that may be overtly discriminatory.” See also, Barocas and Selbst: “Professor Lior Strahilevitz has argued, for instance, that when employers lack access to criminal records, they may consider race in assessing an applicant’s likelihood of having a criminal record because there are statistical differences in the rates at which members of different racial groups have been convicted of crimes.”

51 Barocas and Selbst.
intentional forms of discrimination, but much less attention given to the unintentional. Since unintentional forms of discrimination may be harder to surface, yet still as harmful, Suresh and Guttag built on the Barocas and Selbst framework by focusing on sources of unintended consequences of machine learning: historical bias, representation bias, measurement bias, aggregation bias, and evaluation bias. Many of these biases are rooted in data, since data are the inputs to the machine learning pipeline. This idea is similar to the common computer science concept, GIGO, which stands for “garbage in, garbage out.”

In a popular example, Crawford illustrated data bias issues through the Boston crowdsourcing application known as Street Bump, where the city would allocate resources to fix potholes based on data collected from residents with the application installed on their phone. Since the allocation of resources depended on users with smartphones, and smartphone were less common among low-income and elderly residents, the data, and the subsequent resource allocation, were biased in favor of those who already had more resources available. This data bias example is one of allocative harm, and echoes Eubanks’ point that tools designed to increase efficiency and reduce inequality can actually end up increasing poverty and perpetuating inequality. Each of these sources of unfairness in model design originate as human choices, whether intentional or not. Unfairness can arise in choice of outcome, choice of predictors to use, and choice of training procedure.

---


55 Crawford, Kate. 2013. “The Hidden Biases in Big Data.” Harvard Business Review. https://hbr.org/2013/04/the-hidden-biases-in-big-data: “[...] people in lower income groups in the US are less likely to have smartphones, and this is particularly true of older residents, where smartphone penetration can be as low as 16%. For cities like Boston, this means that smartphone data sets are missing inputs from significant parts of the population — often those who have the fewest resources.” See also, Sweeney, Latanya. 2013. “Discrimination in online ad delivery.” Queue 11(3): 10. See also, Hurley, Mikella, and Julius Adebayo. 2016. “Credit Scoring in the Era of Big Data.” Yale Journal of Law & Technology 18 (148):69. https://yolot.org/sites/default/files/hurley_18yjolt136_jz_proffedits_final_7aug16_clean_o.pdf. See also, Barocas and Selbst.

56 Kleinberg et al. (2019): “For example, an employer that is screening applicant resumes might believe that college quality could matter in predicting worker performance. It might then invest time to assemble US News rankings of four-year colleges (which whites attend at relatively higher rates) without investing the same effort in measuring rankings of two-year colleges or for-profit universities (which black students attend at relatively higher rates).”
Feature selection and proxies are choices with unintended discriminatory effects. Among choices and data bias, two particular areas are examples of unintentional choices with discriminatory effects: feature selection and proxies. In feature selection, designers make choices about what attributes to observe in the data and their eventual outcomes. Proxies are data that are selected for their strong statistical correlation to other data, generally sensitive or personally identifiable data, such as race. As Calders and Zliobaite noted, certain features can have more statistical power and predictive accuracy than others for certain groups, meaning that the selection of features can lead to discrimination simply based on feature selection.\(^{58}\) The authors suggested that inference of a sensitive attribute, such as race, can be observed, even when that attribute is removed from or not included in the model, based on the features selected among other data in the model. In selecting features, model designers generally rely on statistics, looking for generalizations that are rational.\(^{59}\) As Schauer explained in his book *Profiles, Probabilities, and Stereotypes*, there exist unfair rational generalizations, such as those that are non-universal.\(^{60}\) Not only are certain attributes non-universal, but these attributes are also featured in more data than the sensitive data alone.

## Proposed Remedies and their Trade-Offs

Since data, models, and at a more macro-level, humans and application context, can all be sources of unfairness, researchers have suggested remedies to these challenges by: focusing on measuring and understanding the extent of the unfairness problem; formalizing and providing more precise definitions to aid, in part, with measurement; using fairness-enhancing tools and techniques to de-bias data; and, where solutions are unclear, providing a minimum level of increased transparency and explanations to increase trust among human users of these tools. Efforts to build such rigor into machine learning have contributed to an extensive, and still unsolved, debate about fairness in society.\(^{61}\) In some cases, researchers have revealed contradictions or trade-offs that are not universal across contexts, such as in the formalization of different fairness definitions.

---


\(^{60}\) Schauer, Frederick. 2003. *Profiles, Probabilities, and Stereotypes*. Cambridge, MA: The Belknap Press of Harvard University Press. Page 217. See also, Kleinberg et al (2019): “The algorithm can only select from among the candidate predictors made available to it by its human designers, but given the choice of training data and outcome, it is the underlying relationships between the variables in the training data that determines which predictors wind up in the screener, and how much weight they get; it is basically just a statistical matter of which variables are most correlated with the outcome.”

\(^{61}\) See, Corbett-Davies et al.
Concern about whether people are treated fairly led to formalizing fairness. In formalizing fairness definitions, researchers present a perspective of justice by which to evaluate their models and measure harm.\(^{62}\) Zemel et al. identified two broad categorizations for fairness: group fairness, where subgroup proportions are identical to whole group proportions, and individual fairness, where similar individuals are treated similarly.\(^{63}\) Verma and Rubin suggested that most fairness definitions can be measured based on predictive or actual outcome, or a combination of the two.\(^{64}\) The simplest notion of fairness focuses solely on a model’s prediction outcomes, though the authors suggest that combined measurements would be more useful in determining whether a model is fair and comprehensive. More recently, focusing on machine learning for decision-making, Corbett-Davies et al. categorized fairness definitions as anti-classification, where sensitive attributes are not explicitly used; classification parity, where groups are measured and equalized by sensitive attributes; and, calibration, where outcomes are independent of sensitive attributes but conditional on risk estimates.\(^{65}\) Prior discussion, as well as the following examples, indicate that categorizations for fairness definitions are useful frameworks, but do not always neatly align with models in practice.

Most fairness models trade-off rational generalizations and individuality.

In group fairness models, one suggestion is to equalize the false negative rate across groups, assuming that false negatives represent unfairness. This single threshold model is known as group-independent.\(^{66}\) Herein lies a fairness trade-off: group fairness models make assumptions in their categorizations that eliminate individuality, and individual fairness models would be more realistic reflections, but do not generalize and are harder to satisfy computationally.\(^{67}\) This trade-off between group fairness models and individual fairness models exists not just as a technical problem. Schauer suggests that rational generalizations, or group fairness models, may be the best way to ensure fairness, since: “Part of being fair is treating people equally, and part of treating people equally, we can now appreciate, is treating people the same even in the face of relevant differences. The just society is not necessarily one in which each individual is treated as an isolated collection of uniquely arrayed attributes demanding individualized attention. Rather, in some even if not in all respects the just society is one in which differences among individuals

\(^{62}\) One problem with measurement is that it is itself a series of political choices, including that counting requires decisions about categorization; creates norms about what is right; creates illusions of oversimplification; and gives power to those who lead the count. See, Stone, Deborah. 2012. “Numbers.” In Policy Paradox: The Art of Political Decision Making, 3rd ed., 183–205. New York, NY: W. W. Norton. Page 196.


\(^{65}\) Corbett-Davies et al.

\(^{66}\) Zemel et. al.

are often and desirably suppressed in the service of both equality and community.\footnote{68} Technical literature may not consider “community” as a constraint in the same way that Schauer describes, but his reasoning is nevertheless consistent with the present trade-off.

**A model incorporating group and individual fairness trades-off predictive power.**

As a middle ground, Dwork et al. (2011) built a classification model that incorporated both group and individual fairness, equalizing false negative rates over infinitely small groups.\footnote{69} Their demographic parity\footnote{70} model used a similarity metric, matched the demographics of a protected group, defined by sensitive attributes, to the demographics of the underlying population, and equalized outcomes across groups, protected or not. The authors recognized that the limits to this method included defining the subgroups and population: determining how fine-grained or coarse-grained should the classifier be can be traded-off for its predictive power. An example of this in consumer credit scoring would be whether to define the underlying population as the whole adult population in the United States, as a state-level population, or as the credit applicant pool. Depending on the choice of the underlying population and its demographics, the process of matching the subgroup demographics can vary widely: some states have greater racial/ethnic diversity than others and that choice can create more or less predictive results depending on the context in which the classifier is applied.

**An equal opportunity model is incompatible with risk calibration.**

Since the choice of the underlying population can be a trade-off for predictive power in the demographic parity view of fairness, Hardt et al.\footnote{71} focused on equalizing outcomes across groups, where the focus is on the subgroups only. The authors used profit maximization functions first to compare group-independent predictions and discovered that the results vary widely between groups. Then, the authors chose not just a single threshold in the false negative rate, but equalized the false positive and false negative rates across groups and showed that the trade-offs were the same across all groups. In this equal opportunity\footnote{72} view of fairness, there is a constraint ensuring that error types are proportionate across groups, but Pleiss et al. investigated this view of fairness in risk analysis tools, which require calibration of probability

\footnotesize
\begin{itemize}
\item \footnote{68} Schauer, 300: “Yet not only is generality not, in general, unjust, but justice itself maybe involve considerable components of generality. To the extent that justice is centrally about fairness, and to the extent that fairness itself is closely related to equality, then fairness, and therefore justice, cannot be seen as themselves resting on the idea of generality. Part of being fair is treating people equally, and part of treating people equally, we can now appreciate, is treating people the same even in the face of relevant differences. The just society is not necessarily one in which each individual is treated as an isolated collection of uniquely arrayed attributes demanding individualized attention. Rather, in some even if not in all respects the just society is one in which differences among individuals are often and desirably suppressed in the service of both equality and community. As such, the good society is one in which generality is not only inescapable, but is also necessary for justice itself.”
\item \footnote{69} Dwork et al. (2011).
\item \footnote{70} Also known as statistical parity.
\end{itemize}
estimates, and determined that the two are incompatible. The authors posited that, ideally, models should reconcile both equal opportunity and calibration. However, following empirical studies across different datasets, the Pleiss et al. concluded that reconciling the two views leads to a “relaxed” notion of both that is ineffective, since it creates randomized outcomes and introduces reverse-discrimination of qualified candidates. They advised that, in practice, decision-makers may need to choose between equal opportunity or calibration. As an example of this, Kleinberg et al. (2017) proposed models that can reduce crime with changing the incarceration rate, or reduce the incarceration rate without changing crime.

It is not clear how best to manage sensitive information and their proxies.

In both the equal opportunity and demographic parity views of fairness, there is a question around which characteristics to protect or make irrelevant, especially when many data can be correlated. Hurley and Adebayo proposed addressing the problem of proxies by preventing the use of credit scoring models that weighed data, or combinations of data, that were highly correlated with prohibited information, such as race. Their view was that tools were not adequate at detecting and preventing discrimination by proxy. Proxies can be legitimate or not, depending on their use. There are some proxies that are legitimate predictors even if correlated with prohibited information, such as education in the case of hiring, where education may vary by racial or ethnic background (even though it shouldn’t in theory). There may also be cases in which prohibited information is a very strong, if not the best, predictor. In recidivism models, for example, gender is an important predictor --women recidivate less than men-- so choosing not to consider gender in such models would allocate, by design, more harm for women.

---

73 Pleiss et al.: “Seeking equality with respect to a single error rate (e.g. false-negatives, as in the income prediction experiment) will necessarily increase disparity with respect to the other error. This may be tolerable (in the income prediction case, some employees will end up over-paid) but could also be highly problematic (e.g. in criminal justice settings).” See also, Chouldechova; Joseph et al.
74 Pleiss et al.: “Although we provide an algorithm to effectively find the unique feasible solution to both constraints, it is inherently based on randomly exchanging the predictions of the better classifier with the trivial base rate. Even if fairness is reached in expectation, for an individual case, it may be hard to accept that occasionally consequential decisions are made by randomly withholding predictive information, irrespective of a particular person’s feature representation.”
75 Zafar et al.
77 Hurley and Adebayo, 214: “If a data point or combination of data points is strongly correlated to any immutable characteristics or protected affiliations, it cannot be used. A data point can be used, however, if it is only weakly correlated to a prohibited characteristic or affiliation.”
78 Hurley and Adebayo, 199.
79 Corbett-Davies et al. See also, Kleinberg et al. (2019): “When we use the race-blind algorithm, we force the statistical model to assume the relationship is the same for both whites and blacks, which distorts the predicted outcomes for black applicants. This leads us inadvertently to say some high-performing black applicants are actually low-performing, that is, to understate their academic potential and hence reject them, while simultaneously admitting lower-performing black applicants instead.”
80 Corbett-Davies et al., 9. See also, Kleinberg et al. (2019): “One implication is that if we see a predictor included in the model, we cannot tell whether it is that predictor or some other correlated predictor that is actually causally affecting the outcome. Another implication is that if we see an algorithm that does not include a protected personal attribute like race in the final model, that does not mean that a correlated proxy for race is not playing a role. It is worth underlining this point: An algorithm that is formally blind to race or sex might be using a correlated proxy. Whether that is a problem, for legal purposes, depends on the governing legal standard. It is more obviously relevant to a disparate impact claim than to a disparate treatment claim.”
Where data is an important predictor, whether it is a proxy for or the sensitive attribute itself, removing or prohibiting the use of such information may reduce a model’s accuracy. This case brings forth another trade-off: removing sensitive information may reduce discrimination, but it might also reduce a model’s accuracy.

**Sensitive information can be inferred or predicted by humans and machines.**

Another problem with removing information is that removed information can be inferred or predicted. Humans can infer race or national origin by hearing a voice over the phone, or by looking at someone’s name. Even when trying to be “blind,” as was the case with the “blind” auditions discussed in the Introduction, judges inferred gender by hearing the sound of a candidate’s shoes walking across the stage. These inferences may be less accurate than the information themselves. Agan and Starr’s work, as described earlier, suggested that even without models, humans may make assumptions about individuals where information is lacking, and that these assumptions could be even more discriminatory: removing the information about criminal records in employment applications made employment discrimination worse for African Americans, whether they had a criminal record or not. These human inferences are perhaps limited in scope and scale, but this is not the case when data and modeling are applied.

Kosinski et al. used linear and logistic regression models to analyze the social network behavior of over 58,000 individuals and were able to to predict sensitive attributes with relatively high accuracy: “the model correctly discriminates between homosexual and heterosexual men in 88% of cases, African Americans and Caucasian Americans in 95% of cases, and between Democrat and Republican in 85% of cases.”

81 Calders, Toon and Sicco Verwer. 2010. “Presentation at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases: Three Naïve Bayes Approaches for Discrimination-Free Classification.”

82 Dwork et al. (2011) show that to treat individuals fairly, they may need to be more aware about demographics, using more information about individuals, with possible consequences to their privacy. While they suggest methods to preserve privacy, this approach also begs the question as to how much information is needed about individuals in order to build this fair classifier. See also, Corbett-Davies et al. See also, Grgic-Hlaca, Nina, Muhammad Bilal Zafar, Krishna P Gummadi, and Adrian Weller. 2016. “The Case for Process Fairness in Learning: Feature Selection for Fair Decision Making.”

83 Rice, Patricia. 2006. “Linguistic profiling: The sound of your voice may determine if you get that apartment or not” The Source at the Washington University in St. Louis.

84 “What’s in a Name?” 2003. MIT News.

85 Rice (2013).

86 Agan and Starr. For additional commentary, see supra note 50.

87 White House Office of Science and Technology Policy (2014), 8: “Similarly, integrating diverse data can lead to what some analysts call the “mosaic effect,” whereby personally identifiable information can be derived or inferred from datasets that do not even include personal identifiers, bringing into focus a picture of who an individual is and what he or she likes.”

correlated data, Volkova and Bachrach categorized Twitter data from 123,513 users by emotional tone and were able to predict demographics based on emotional tone and social network.\(^{89}\) These findings render null the idea of being “blind” to certain characteristics in deciding outcomes for individuals.\(^{90}\) It may be more accurate, rather than predicting or inferring, to include sensitive information in models, so as to more directly remedy the effect of bias in models.\(^{91}\) As Corbett-Davies et al. demonstrated, “the exclusion of any information— including features that are explicitly protected— can lead to discriminatory decisions.”\(^{92}\)

**Perhaps, instead of removing sensitive information, it can be de-biased.**

Given that excluding sensitive attributes does not necessarily protect them from being inferred or predicted, researchers have considered techniques for de-biasing data and models, which does not remove information, but rather uses techniques such as weighing and feature selection to control the value of the information in the outcome. Suresh and Guttag broadly categorize these techniques as data-based; model-based; and post-hoc.\(^{93}\) In data-based techniques, certain data can be weighed to fulfill a fairness objective.\(^{94}\) For example, Bolukbasi et al. and Zhao et al. removed gender from word embeddings.\(^{89}\) These manipulations become increasingly difficult with more data and complex models. In model-based methods, researchers added regularization terms or constraints to enforce specific fairness definitions, leading to accuracy trade-offs, among others.\(^{96}\) IBM’s AI Fairness 360 is an example of a model-based de-biasing tool.\(^{97}\) Such tools are valuable as a check, perhaps, but do not necessarily provide incentive for due diligence

---

\(^{89}\) Volkova, Svitlana, and Yoram Bachrach. 2016. “Inferring Perceived Demographics from User Emotional Tone and User-Environment Emotional Contrast.” *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.* [http://www.cs.york.ac.uk/semeval-2013/task2/](http://www.cs.york.ac.uk/semeval-2013/task2/). See also, de Montjoye, Yves-Alexandre, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. 2013. “Unique in the Crowd: The Privacy Bounds of Human Mobility.” *Nature Scientific Reports* 3. [https://doi.org/10.1038/srep01376](https://doi.org/10.1038/srep01376). This study uses location data to re-identify individuals, indicating that, even in coarser datasets, individual privacy is limited.

\(^{90}\) Hardt et al.: “Despite this demand, a vetted methodology for avoiding discrimination against protected attributes in machine learning is lacking. A naive approach might require that the algorithm should ignore all protected attributes such as race, color, religion, gender, disability, or family status. However, this idea of ‘fairness through unawareness’ is ineffective due to the existence of redundant encodings, ways of predicting protected attributes from other features.”


\(^{92}\) Corbett-Davies et al.

\(^{93}\) Suresh and Guttag.


in data processing at the outset, and have not been necessarily aligned with policy contexts. To this end, researchers argued that post-hoc methods aligned more with policy practices, by providing the information and giving decisionmakers the responsibility to identify decision thresholds. Dwork et al. (2016) proposed a decoupling technique that can be considered both model-based and post-hoc since it can be added to a machine learning model to learn different classifiers for groups categorized by sensitive attribute. This decoupling technique was designed to help maximize accuracy for majority and minority groups, using transfer learning for groups with less data, which is often a practical limitation for minority groups. Post-hoc methods account for context, which is especially valuable where formalizations of fairness definitions lead to much abstraction.

De-biasing techniques might not be as effective as expected.

Despite efforts to de-bias data and models through the machine learning pipeline, more recently, researchers have shown that de-biasing techniques are not effective. Building off of work similar to Bolukbasi et al., Gonen and Goldberg used text data to show that two de-biasing techniques hide, but do not remove, bias, like “lipstick on a pig.” The authors used their experiments to also show that they were able to recover bias, which exhibits similar concerns to research in privacy re-identification field, such as the work of de Montjoye et al. Approaching the question from another angle, Kohler-Hausmann, a sociologist and law professor, suggested that certain sensitive data, such as race, should not solely be thought of as social constructions, but also as a biological or genetic concepts, meaning that removing discrimination would extend beyond simple data manipulations. These sensitive data may be the “stickiest” and therefore the hardest to de-bias. Opportunity Insights, a group that studies upward mobility in the United States, showed that sensitive information, such as race, is correlated across other data, including life expectancy and income level. As proxies increase with the use of “alternative” data and models, there will need to be some agreement as to how to use sensitive information, whether to use it directly, remove it, or de-bias it in some way, and what the trade-offs of those choices would imply for privacy and fairness. Since there is not yet a

98 There are arguments against becoming too-context specific in, for example, Suresh and Guttag: “For instance, while ensuring group-independent predictions might make sense in hiring (when it is illegal to factor gender or ethnicity into decisions), it would not be appropriate in a medical application where gender and race can play an important role in understanding a patient’s symptoms.”
99 See, Hardt et al.; Corbett-Davies et al.
100 Dwork et al. (2016).
102 Bolukbasi et al.
104 See, de Montjoye et al.
clear solution, researchers have also developed interim proposals to engender trust through transparency.

**Transparency is an interim solution.**

In such cases, where formalized definitions are too abstract and not contextualized, and biased data cannot be removed since it can be inferred, more transparency, while not a solution onto itself, may provide more information by which humans can decide how to handle discrimination. In 2018, Woodruff et al.\(^{108}\) published research about perceptions of algorithmic fairness, noting that, as companies start to address algorithmic fairness, there could be an impact on consumer trust. Their study concluded that most individuals prefer human decision-making because of their perceptions that algorithms are unfair. In another study of perceptions, Binns et al. noted that individuals “not only care about whether the outcome of a decision benefits them, but also whether it meets standards of justice.”\(^{109}\) The authors suggested that, to be perceived as fair and trustworthy, decision-making algorithms need to provide information relevant to justice, which is perceived as existing in human decision-making. Gebru et al. suggested a standardized format for transparency.\(^{110}\) This “datasheet” would disclose information about the origins of the data, including demographics, and recommended uses or limitations. Building on this work, Mitchell et al. provided a similar standardized form for models.\(^{111}\) There are other ways to build more transparent models, through explanation and interpretability.

**Explainable and interpretable models are a form of transparency.**

As models increase in complexity of design, amount of input data, and training methods, correlations are harder to trace, and there are efforts underway to change this, including research in making machine learning models more interpretable and/or explainable.\(^{112}\) The purpose of this research is to increase trust by making machine learning models less opaque and difficult to analyze.\(^{113}\) Since biases may persist, even unintentionally, transparency and model

---


\(^{111}\) Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Deborah Raji, and Timnit Gebru. 2019. “Model Cards for Model Reporting.” [https://doi.org/10.1145/3287569.3287596](https://doi.org/10.1145/3287569.3287596).


explanations could reveal these biases and allow for deliberation as to whether or not these are acceptable. Where there is no clear technique or tool to use, transparency includes visibility into the design of the model and data used to train it. Transparency also includes some form of interpretability or explainability to increase trust among human users of these tools. Weller examines the idea that more transparency is always better and more helpful, and suggests cases where increased transparency could actually be a source of harm, such as when transparency is used as distraction from the actual problem or to invade privacy. In some cases, in the public sector particularly, there is a greater requirement for transparency and explanation. Algorithmic impact assessments, such as those suggested by Reisman et al., may help determine which contexts require more transparency and mitigation than others.

**Transparency is not itself a justification, but may help in providing one.** There is not always a greater requirement for transparency, Weinberger cautioned. He argued that transparency may not be needed as long as the results are satisfactory or optimal. Further, as Wachter et al. reminded that explanations are not justifications. A justification implies a causal relationship, whereas machine learning models are not designed to produce cause, but rather correlation. Algorithms are better at prediction than they are at causal inference. Wachter et al. asked whether it is justified to use algorithms or machine learning models that are not explainable or where the causal relationship is not clear. They suggested counterfactual reasoning as a remedy, but noted that, since counterfactual reasoning is most useful when individualized, they may not have broad applications. Counterfactuals can expose whether a protected attribute was used, but this does not always mean that the attribute was used unfairly or unjustly. Counterfactuals are a “minimal form of explanation,” providing some transparency with an emphasis on causal relationships between variables, and leaving the full explanation and reasoning to humans.

---


117 See, Reisman et al.


121 Wachter et al.: “In general, the best tools for uncovering systematic biases are likely to be based upon large-scale statistical analysis and not upon explanations of individual decisions.”

122 Wachter et al.
Transparency may be a way to verify human behavior.
Humans themselves are not always fully explainable or interpretable, especially when it comes
to discrimination. Where sensitive attributes are to be used, Pleiss et al. prefer that machines
take these into account rather than humans, given the inscrutability of human biases.\textsuperscript{123} Grgic-Hlaca et al. provide an example where feature selection, even in a simple model, can be
perceived as fair, but actually slightly increased the racism exhibited by algorithms, while
decreasing accuracy.\textsuperscript{124} Kleinberg et al. suggested that there should be safeguards for the humans
who build algorithms.\textsuperscript{125} The authors recognized the complexity of their proposal, but take an
inherently techno-optimist view towards algorithms: they believe that algorithms can be a tool
for social justice in helping to reduce human bias and discrimination. Hammond, in 1996,
suggested that artificial intelligence and machine learning were not capable of judgement and
decision-making.\textsuperscript{126} While there may still be areas that these tools cannot, and should not, be
providing decisions, there is reason to believe that such tools may not only be able to safeguard
against human biases, but also perform in ways that are more examinable than human behavior.

Fairness is an evolving discourse that differs by context.
There are a variety of remedies proposed to manage fairness in machine learning, including
among the humans who design these tools. Fairness is an evolving discourse, making it difficult
to formalize and measure, though such attempts have revealed an extensive interdisciplinary
debate. Fairness is also not just a substantive construct, but also a process that differs by
context, making it difficult to be precise. In absence of definitions or tools to measure fairness or
de-bias data and models, increased transparency can help foster trust in these systems.
Transparency, whether through explanations or otherwise, however, does not provide a causal
relationship or a justification: it simply exposes unfairness. In his 2018 FAT/ML tutorial,\textsuperscript{127} Narayanan showed how attempts by researchers to formalize and encode fairness definitions fail
to encompass the complexity of underlying political and ethical values questions. Instead of
pushing for a sole definition of fairness, he suggested instead that researchers accept, and even
celebrate, the variety of definitions, each unique to their context. In the spirit of Narayanan’s
tutorial, the following section brings forth a definition of fairness in context, the consumer credit
context.

\textsuperscript{123} Pleiss et al.: “If a risk tool for evaluating defendants were not calibrated with respect to groups defined by race, for
example, then a probability estimate of \( p \) could carry different meaning for African-American and white defendants,
and hence the tool would have the unintended and highly undesirable consequence of incentivizing judges to take
race into account when interpreting its predictions.”

\textsuperscript{124} Grgic-Hlaca et al.

\textsuperscript{125} Kleinberg et al. (2019): “The frequent inscrutability of human decision-making, combined with the relative rarity of
explicit archival evidence (memos, verbal statements) of discriminatory motives, makes violations difficult to police.”

\textsuperscript{126} Hammond, Kenneth R. 1996. “Irreducible Uncertainty and the Need for Judgment.” In Human Judgment and
Page 10: “I have not included work from the fields of artificial intelligence, expert systems, linguistics, logic, or
philosophy. There is too large a gap between these fields and the field of judgment and decision making.”

\textsuperscript{127} See, Narayanan, Arvind. 2018. “Tutorial: 21 fairness definitions and their politics.” Conference on Fairness,
Accountability, and Transparency. https://www.youtube.com/watch?v=iIXuYdnyyK.
Among the research in machine learning fairness, there is focus on sources of harm and proposed remedies, but less attention devoted to a comprehensive view that includes existing mitigations for unfairness in law and regulatory policy. In the context of consumer credit, the FCRA and the ECOA envision a system that is both substantively and procedurally fair: equally accessible by all consumers, and developed on accurate and relevant information, with controls for consumer privacy. A third goal is credit expansion and financial inclusion. This vision and its implementing mechanisms have provided the groundwork for managing fairness.

The current mechanisms may not be the only possible implementing mechanisms. The proposed remedies discussed in this Chapter are possible new mechanisms. The use of legal and regulatory precedent, combined with an understanding of the machine learning techniques and the “alternative” processes sought by the consumer credit industry, provides insight into how policymakers may address fairness most effectively in light of technological change. The combination of both is necessary, for example, since in some cases, mechanical application of such rules to machine learning may actually increase unfairness, going against intended policy goals. This is the case with the existing mechanisms that equate “blindness” to protected attributes as fair. The research presented in this Chapter reaffirms that this “blindness” policy is no longer an effective mechanism by which to create a fair consumer credit system. Following an overview of the legal and regulatory context in the next Chapter, Chapters III and IV will develop the argument that “blindness” is a fallacy as machine learning is applied to the consumer credit context.
II. Background | Consumer Credit Context

“We are coming to Washington in a poor people’s campaign. [...] We are coming to demand that the government address itself to the problem of poverty. We read one day: We hold these truths to be self-evident, that all men are created equal, that they are endowed by their creator with certain inalienable rights. That among these are life, liberty, and the pursuit of happiness. But if a man doesn’t have a job or an income, he has neither life nor liberty nor the possibility for the pursuit of happiness. He merely exists [...].”

Consumer credit laws in the U.S. are designed to provide economic efficiencies through the availability of credit, to remedy the trust issues associated with credit and lending, and to prevent abuse and discrimination. Among these laws, the Fair Credit Reporting Act (FCRA) and the Equal Credit Opportunity Act (ECOA) govern the decision-making process for extending consumer credit. In this system there are goals, and mechanisms that are designed to help achieve those goals. The overarching goal in the consumer credit system is to ensure trust, broken down into three sub-goals:

- **Procedural** fairness in the FCRA: the consumer credit system should be built on accurate and relevant information, with controls for consumer privacy.
- **Substantive** fairness in the ECOA: the consumer credit system should be equally accessible by all consumers.
- A third goal is credit expansion and financial inclusion. This goal is a long-standing government policy that took root in credit reforms at the turn of the 20th century, and is the primary driver behind interest in “alternative” processes today.

Broadly speaking, the FCRA addresses fairness in the consumer credit process and the ECOA focuses substantively on limiting discrimination in credit applications and evaluations. These goals are developed further throughout this Chapter, and distinguished from their implementing mechanisms. In the context of “alternative” data and modeling techniques, these fair credit goals are still valid, but the mechanisms by which to achieve them are no longer. Technological advances require contextualization, and can possibly evolve fairness and encourage change.

Our understanding of fairness is institutionalized today, primarily in the FCRA and the ECOA. This institutionalization represents an evolution in views on fairness. In the early 20th century, a credit culture began emerging in the United States. With this emerging culture came

---

129 15 U.S.C § 1681 et seq.
130 See supra note 35.
inequities: given interest ceiling and high administrative costs in processing a loan, the credit system did not extend to the average consumer, and was reserved instead for businesses and those with individual wealth equivalent to that of a business. There were some small credit setups, such as having a line of credit at the local grocer, but the average consumer in need of a loan often had no choice but to resort to loan sharks who offered unregulated, often very high, interest rates. Reforms such as the creation of credit unions in the 1920s, which emphasized pooling of resources among members, and issuing a common bond, and the creation of the Federal Trade Commission (FTC) with its mission to protect consumers against “unfair and deceptive practices” democratized access to credit while providing protective measures that increased trust in the system and encouraged its use.

Yet, by the late 1960s, despite prior efforts to democratize access to credit, 70 percent of low-income consumers had little or no access to credit. This time, consumer credit industry reform was anchored in the Civil Rights’ Movement. Reverend Martin Luther King Jr.’s march on Washington featured economic inequality as a core civil rights issue. The Poor People’s Campaign evolved the view of fairness and added further motivation and weight to the overall issue of access to credit and consumer protection. Given that the American economy revolved around credit, access to credit was synonymous to access to opportunities, including a home and employment. And yet, despite prior reforms, discrimination and unfair processes in consumer credit still limited minorities and the poor from such opportunities. Recognizing the compounding inequalities and perpetuating cycles of poverty, Congress passed the FCRA in 1970 and the ECOA in 1974. Together, these two laws contribute to the foundation for consumer credit protection in the United States.

The remainder of this Chapter is dedicated to the examination of these laws, and their respective evolutions, so as to better understand fairness in the consumer credit context. First, the FCRA addresses fairness in the consumer credit process. This procedural fairness includes requirements for information to be accurate and relevant to the consumer; for consumers to have controls over their information and privacy; and for obsolete or adverse information about consumers to be restricted. In implementing a fair consumer credit process, policymakers created an accountable and trustworthy environment for consumers. But focusing on the credit process alone was not enough, since there were also issues of access to the credit process by certain groups. Enacting the ECOA addressed fairness from a substantive perspective by limiting discrimination in credit applications and evaluations. The ECOA, and its implementing

---

134 Planet Money.
135 Carter, 6.
136 Carter, 6.
137 16 U.S.C § 239.1
138 Olegario, 147: “In 1967, 70 percent of low-income consumers either had no credit accounts at all or had them with only a few retailers in low-income areas.”
139 There are also other laws that are play a role in consumer credit protections including the Truth in Lending Act (1968) and Electronic Funds Transfer Act (1979).
regulation known as Regulation B,\textsuperscript{140} prohibit discrimination in credit transactions on the basis of certain characteristics; require notice and record retention of credit decisions in case claims of discrimination need to be evaluated; and incentivize verification and self-testing of credit scoring models to ensure compliance. Together, these procedural and substantive parts of fairness are the foundation of the consumer credit context.

**Procedural Fairness in the Fair Credit Reporting Act**

The FCRA was enacted in 1970 to regulate procedural aspects of consumer credit reporting, “to ensure fair and accurate credit reporting, promote efficiency in the banking system, and protect consumer privacy.”\textsuperscript{141} The FCRA primarily regulates practices of consumer reporting agencies that collect, compile, and sell the data produced in a consumer credit report. In its first iteration, the FCRA focused exclusively on consumer reporting agencies and consumer report users, such as employers.\textsuperscript{142} There was, at least among lawmakers, fears of privacy invasions, or worse, surveillance.\textsuperscript{143} The FCRA was not concerned with data sharing in the 1970s,\textsuperscript{144} but following a data boom in the 1990s, definitions were expanded, adding a new class of entities, data furnishers,\textsuperscript{145} and imposing more requirements on all entities --consumer reporting agencies, consumer report users, and data furnishers-- with regard to maintaining accuracy of information and resolving consumer disputes. This was significant because maintaining accuracy and integrity of information had become a significant challenge, and because the expanded definitions gave regulators more oversight. The FCRA’s scope depends in part on the definition of a “consumer report.”\textsuperscript{146} When data is not used for purposes related to credit, then it

\textsuperscript{140} Reg. B, 12 C.F.R. pt. 1002.
\textsuperscript{141} See, Safeco Insurance Co. of America v. Burr. 551 U.S. 47 (2007). Given its focus on consumer data protection, its accuracy and integrity, and its limits against illegitimate (“permissible”) uses of personal data, the FCRA can be regarded as the first data protection law in the United States. Permissible uses of consumer credit reports are: with consumer’s permission; for employment; for credit or credit transactions; for business transactions initiated by the consumer; and to review an account. The European Union’s General Data Protection Regulation (GDPR) went into effect in May 2018. The FCRA could serve as a model where there is interest in the United States to enact some form of comprehensive, as opposed to sectoral, consumer data protection law.
\textsuperscript{142} Federal Trade Commission. 2011. “40 Years of Experience with the Fair Credit Reporting Act: An FTC Staff Report with Summary of Interpretations.” Section II. This report was issued days before the CFPB was launched, as a way to share experience in interpreting FCRA.
\textsuperscript{143} Lauer, Chapter 8 “Database Panic.”
\textsuperscript{144} Lauer.
\textsuperscript{145} 16 C.F.R. pt. 660 provides the definition: “Furnisher means an entity that furnishes information relating to consumers to one or more consumer reporting agencies for inclusion in a consumer report. An entity is not a furnisher when it: (1) Provides information to a consumer reporting agency solely to obtain a consumer report in accordance with sections 604(a) and (f) of the Fair Credit Reporting Act; (2) Is acting as a “consumer reporting agency” as defined in section 603(f) of the Fair Credit Reporting Act; (3) Is a consumer to whom the furnished information pertains; or (4) Is a neighbor, friend, or associate of the consumer, or another individual with whom the consumer is acquainted or who may have knowledge about the consumer, and who provides information about the consumer's character, general reputation, personal characteristics, or mode of living in response to a specific request from a consumer reporting agency.”
In expanding the FCRA's scope, industry players needed to think of themselves as sharing in the responsibility of a fair consumer credit process. This shared responsibility led eventually to the standardization of data furnishing, processing, and provision (e.g., eOSCAR, Metric 1, Metric 2), which also increased the efficiency and accuracy of the credit system.

In a 2003 modification of FCRA, the Fair and Accurate Credit Transaction Act (FACT Act) addressed fraud prevention measures in response to identity theft that had plagued consumers. Emphasis was placed on consumer access to their information, by granting consumers the right to one free annual credit report from a consumer reporting agency, therefore allowing consumers to verify, dispute, and correct inaccuracies in their reports. This was important because of how much of an effect consumer credit reports could have on an individual’s opportunities. Finally, until 2011, the Federal Trade Commission (FTC) was primary enforcer of FCRA, under Section 5 of the FTC Act, but had limited rulemaking authority. Following the 2008 financial crisis, the Dodd-Frank Wall Street Reform and Consumer Protection of 2010 (Dodd-Frank Act) created a new federal agency, the Consumer Financial Protection Bureau (CFPB) which consolidated enforcement and rulemaking authority.

Key themes in the FCRA are accuracy, controls for privacy, and restrictions on obsolete or adverse information. The FCRA’s policy goal is to provide consumers with a fair process that will instill trust and confidence in the credit system and banking system more broadly. To achieve this goal, the FCRA has evolved to include a set of consumer rights, and rules for industry participants who take part in the credit process. Each of these are policy choices, reflective of values and visions for a credit process that is procedurally fair. In some cases, the FCRA includes not only the goals, but also mechanisms by which lawmakers thought these goals should be achieved.

With regard to accuracy, information must relate to the consumer and be verifiable. While accuracy is not defined explicitly in the FCRA, the liability structure, the mechanism, is such that credit reporting agencies and data furnishers are held liable for inaccurate information if reasonable procedures were not taken to ensure “maximum possible accuracy.” Consumers have the right to both dispute inaccurate or incomplete information in their files, and the right to verify, dispute, and correct inaccuracies in their reports. This was important because of how much of an effect consumer credit reports could have on an individual’s opportunities. Finally, until 2011, the Federal Trade Commission (FTC) was primary enforcer of FCRA, under Section 5 of the FTC Act, but had limited rulemaking authority. Following the 2008 financial crisis, the Dodd-Frank Wall Street Reform and Consumer Protection of 2010 (Dodd-Frank Act) created a new federal agency, the Consumer Financial Protection Bureau (CFPB) which consolidated enforcement and rulemaking authority.

Key themes in the FCRA are accuracy, controls for privacy, and restrictions on obsolete or adverse information. The FCRA’s policy goal is to provide consumers with a fair process that will instill trust and confidence in the credit system and banking system more broadly. To achieve this goal, the FCRA has evolved to include a set of consumer rights, and rules for industry participants who take part in the credit process. Each of these are policy choices, reflective of values and visions for a credit process that is procedurally fair. In some cases, the FCRA includes not only the goals, but also mechanisms by which lawmakers thought these goals should be achieved.

With regard to accuracy, information must relate to the consumer and be verifiable. While accuracy is not defined explicitly in the FCRA, the liability structure, the mechanism, is such that credit reporting agencies and data furnishers are held liable for inaccurate information if reasonable procedures were not taken to ensure “maximum possible accuracy.” Consumers have the right to both dispute inaccurate or incomplete information in their files, and the right to verify, dispute, and correct inaccuracies in their reports.

---

147 Wu, 77: “If the information is collected and used exclusively for marketing purposes unrelated to credit, employment, insurance, legitimate business need, or other FCRA purposes, it is probably not a consumer report and is not covered by the FCRA.”


149 Section 5 of the FTC Act gives the FTC enforcement authority for FCRA violations: such violations would be considered unfair and deceptive practices. This authority applied primarily to consumer reporting agencies, since banking agencies have authority over financial institutions. See, Federal Trade Commission (2011), Section III.


151 Wu, 126.

152 This standard came out of cases in the 1980s, such as Pinner v. Schmidt 805 F.2d 1258 (5th Cir. 1986) and Koropoulos v. Credit Bureau Inc.734 F.2d 372 (D.C. Cir. 1984). Cited in, Wu, 128. Additionally, a report may be considered inaccurate, if it is misleading (despite being technically accurate), though this is subject to Court decision: “Numerous district and state courts have held that a consumer report can be inaccurate, even if the report is technically true in some narrow sense, but the report is overly general, incomplete, out of date, or misleading.”
to seek damages from violators of this process.\textsuperscript{153} Part of accuracy involves a limited scope, giving consumers some control over their information and privacy.\textsuperscript{154} The FCRA limits access to credit reports to those with valid needs such as creditors, insurers, landlords, and employers.\textsuperscript{155} These “permissible purposes” are: with consumer’s permission; for employment; for credit or credit transactions; for business transactions initiated by the consumer; and to review an account. So, while employers may have a permissible purpose, the consumer’s consent is still required before the employer can access the consumer’s credit report.\textsuperscript{156} Finally, another part of this limited scope is that consumers may opt-out of pre-screened offers of credit based on the information their report. This gives consumers control over who has access to their information and for what purposes.

Consumers have the right to know what is in their file. Since 2003, consumers are given one free annual credit report from each of the three national consumer reporting agencies. Consumers also have access to their files under specific conditions such as unemployment or identity theft. This way, consumers can review their files, and dispute information that is inaccurate. Consumer reporting agencies have the duty to investigate as a result, and if the inaccuracy remains and a reinvestigation is required, consumers can claim harm for loss of credit opportunity.\textsuperscript{157} Though consumers have the right to access their file, and perhaps the annual responsibility to do so, policymakers chose not to leave the burden solely on consumers.\textsuperscript{158} If their credit information is being used against them, consumers must be informed.

Importantly, the FCRA places restrictions on obsolete or adverse information, but no restrictions on favorable information. Adverse information is not defined in the FCRA, but generally understood as information that would have an “unfavorable bearing on a consumer’s eligibility or qualifications for credit.”\textsuperscript{159} Information is obsolete when it is more than seven years old, with some exceptions such as criminal convictions, bankruptcies, etc.\textsuperscript{160} If there is favorable information about a consumer that more than seven years old, it can remain on credit reports.\textsuperscript{161} This gives consumers an advantage: their most favorable profile is put forward to creditors. Having time-limits on negative information also indicates that policymakers wanted

\begin{itemize}
\item \textsuperscript{153} Wu, 934.
\item \textsuperscript{154} Information control and privacy were main themes of the 1968 hearing that lead to the eventual passage of the FCRA. See, Opening statement of Rep. Cornelius E. Gallagher in “Commercial Credit Bureaus: Hearings before a subcommittee of the Committee on Government Operations,” House of Representatives, Ninetieth Congress, second session. 1968. Washington, DC: United States Government Printing Office: “the Special Subcommittee on Invasion of Privacy initiates an inquiry into one aspect of the seemingly endless trail of records left by individual Americans. The subcommittee has great interest in the confidentiality and accuracy of such records. It is the number and extent of commercial credit bureau files and the crucial question of possible unauthorized access to highly personal information which will concern us.”
\item \textsuperscript{155} Wu, 934.
\item \textsuperscript{156} Wu, 934.
\item \textsuperscript{157} Wu, 535 and 613.
\item \textsuperscript{158} Errors disproportionately impact consumer with lower levels of education. See, Hurley and Adebayo, 156.
\item \textsuperscript{159} FTC definition cited in Wu, 242.
\item \textsuperscript{160} Exceptions to the seven year rule are as follows: bankruptcies stay on files for ten years; prescreening inquiries stay on files for one year; and criminal convictions are permanently on files. See, Wu, 241.
\item Even though there are no restrictions on favorable information, in practice, consumer reporting agencies tend to stop reporting information that is more than ten years old. See, Wu, 242.
\end{itemize}
to give consumers the ability to renew their credit profiles to what amounts to about a dozen
times throughout their expected adult life.

Procedural fairness is achieved by establishing a credit reporting ecosystem that distributes
responsibility for producing accurate reports with information that is related to the consumer
and verifiable, that the consumer has some control over in terms of its distribution and
disputability, and that weighs favorable over obsolete or inaccurate where possible. In these
elements of procedural fairness are embedded values and visions of policymakers. However,
policymakers did not stop at mandating fair credit processes. Four years after the initial passage
of the FCRA, policymakers passed the ECOA, which addresses substantive fairness in the credit
reporting system by preventing discrimination on the basis of certain characteristics.

**Substantive Fairness in the Equal Credit Opportunity Act**

Following the enactment of the FCRA, which guaranteed procedural fairness in the credit
process, legislators worked to enact the ECOA, which addresses substantive fairness in the credit
process: it is the anti-discrimination law for consumer credit. There are other
anti-discrimination laws such as the Civil Rights Act (CRA) and the Fair Housing Act (FHA) that
protect against other forms of discrimination, indicating perhaps that certain discrimination is
deemed more harmful in some contexts than others, and, in addition to these federal laws, there
are state laws that offer more prohibited bases, such as sexual orientation or location of
residence. As with the case of “blind” auditions in orchestras, there are also contexts beyond
consumer credit, such as life insurance, that instituted “blindness” as a mechanism to ensure
fairness. This thesis focuses on the consumer credit context, and therefore the ECOA, though
some comparison is drawn to other anti-discrimination legislation where illustrative.

The ECOA prevents creditors --anyone who participates in a credit decision, as part of regular
business operations-- from denying loans or charging higher rates on the basis of one or more
protected attributes. When originally passed in 1974, the ECOA addressed sex and marital
status discrimination in issuance of personal credit. In 1976, Congress, upon advice from the

---

165 Which has been interpreted to include gender, and in many states, sexual orientation. See, Brattle, 51.
166 Brattle, 7. See also, Olegario, 148: “Women who were married, divorced, or separated had no credit identities of
their own but instead were regarded as extensions of their husbands’ financial identities.” In 1988, the ECOA was
broadened again to include business and commercial credit since those were areas where consumers (women in
particular) faced discrimination. See, Hall, Jon C., and Barbara D. Hall. 2006. *Adam’s Eve: A Handbook for the
Social Revolution: ECOA and the Story of Adam and Eve*. Bloomington, IN: AuthorHouse. See also, Brattle, 8;
Parker et al., 243: “In 1974, Congress passed the Equal Credit Opportunity Act, which prohibited credit
discrimination on the basis of sex and marital status. It was amended in 1976 to include race, color, religion, national
origin, source of income, and age. In 1977, the FTC began to devote a significant fraction of its resources towards
enforcing the act and addressing the discriminatory practices that had led to its passage.”

36
Department of Justice (DOJ) to make some protections common across all civil rights legislation, broadened the ECOA to include race, color, religion, national origin. Age,\(^{167}\) receipt of any public assistance income,\(^{168}\) and exercises of rights under the Consumer Credit Protection Act\(^ {169}\) were also added based on conclusions from congressional hearings in 1975.\(^ {170}\) In sum, the ECOA prohibits discrimination on the basis of:

- Race,
- Color,
- Religion,
- National origin,
- Sex,
- Marital status,
- Receipt of public assistance income, and
- Exercises of rights under the Consumer Credit Protection Act.

Today, the scope of ECOA is very broad and there are few exceptions. However, in his treatise on the subject, *Credit Discrimination*, Jeremiah Brattle suggests, “Despite the far-reaching remedies available, credit discrimination laws were not utilized significantly until the 1990s and may not yet be fully utilized.”\(^ {171}\) With increasing public awareness in 1990s on the existence of widespread discrimination, discrimination laws were strengthened (such as the FHA with the passage of the 1991 CRA) and more discrimination cases were brought in state and federal courts. The ECOA, and cases brought forth under the ECOA, benefit from evolutions in discrimination legislation and case law in other contexts, particularly housing and employment.

For example, the McDonnell-Douglas test, which was developed by the Supreme Court in an employment discrimination case, is often used in analysis for ECOA cases.\(^ {172}\) Finally, while several federal agencies, including the FTC, used to be responsible for the ECOA enforcement, those responsibilities have since been consolidated and allocated to the CFPB. This consolidation brought enforcement of procedural and substantive fairness under a single authority, the CFPB.

The prohibited bases for decision-making are an ECOA mechanism for substantive fairness. The ECOA includes other mechanisms, such as providing notices of adverse actions and record retention, that help implement substantive fairness and prescribe equality in the credit process. For example, the creditor must retain, for a period of twenty-five months, the consumer’s

---

\(^ {167}\) As long as the applicant is old enough to enter into a contract. The ECOA focuses primarily on discrimination against elderly and senior citizen populations. See, Brattle, 53.

\(^ {168}\) Such as Social Security Disability Insurance (SSDI) or the Supplemental Nutrition Assistance Program (SNAP).

\(^ {169}\) Pub.L. 90–321 (May 29, 1968). Included in the Consumer Credit Protection Act are the following acts: Truth in Lending Act; Fair Credit Billing Act; Consumer Leasing Act; Federal Garnishment Act; Federal Credit Repair Organization Act; Fair Credit Reporting Act; Fair Debt Collection Practices Act; Electronic Fund Transfer Act; and Equal Credit Opportunity Act.

\(^ {170}\) Brattle, 8.

\(^ {171}\) Brattle, 2.

\(^ {172}\) McDonnell Douglas Corp. v. Green, 411 U.S. 792 (1973). Complainants need to prove discrimination using the McDonnell Douglas test, which involves satisfying all three of the following: establishing a case of discrimination; producing evidence of a legitimate non-discriminatory reason for its actions (there should be none); and showing an inference of discrimination. See also, Brattle, 15.
application, information and criteria used in evaluating an application and/or extending an offer, notifications to the applicant, and any statements by the applicant alleging non-compliance with the ECOA. These rules educate consumers about the credit process, and may help create a body of evidence that can be used to determine whether there was discrimination in the credit process. With these retention requirements, the ECOA makes explicit the importance of non-discrimination at each stage of the credit process, including the application, evaluation, and extension of credit stages, without requiring disclosure of information, making it an elegant mechanism in service of its fair credit goal.

Throughout the credit process, though especially in the application stage, the ECOA prohibits creditors from requesting prohibited information about applicants. It is possible that the creditor may receive information from consumer reporting agencies, for example, but the creditor is prohibited from using that information in as the basis for a credit decision. In terms of retention, creditors may retain prohibited information so long as it was received, not requested, and they may not use the information in credit decisions, except in certain circumstances. These circumstances are summarized as follows:

- Prohibited bases may be considered to determine whether an applicant qualifies for special purpose credit programs, such as programs that favor the elderly and may need to consider age for eligibility.
- Age may be taken into account in a credit scoring system as long as elderly applicants are treated as favorably as younger applicants. In a decision system, age may be taken into account on a case-by-case basis if it helps evaluate other aspects of creditworthiness, such as limited employment history for a younger applicant.
- Immigration status and citizenship may have implications for a creditor’s ability to obtain repayment, so denial of credit on these grounds is not necessarily discrimination based on national origin.
- Self-testing of credit models is also an exception, further discussed below.

Self-testing of credit scoring models is a mechanism to ensure compliance with the ECOA. During regulatory reforms in the mid-1990s, the Federal Reserve Board (FRB) added self-testing and detailed rules for the process in Regulation B. According to Regulation B, a credit scoring system must be developed with a significant portion (or the entire) consumer credit file; be used to predict creditworthiness; be designed and tested using statistical methods; and be reviewed regularly and updated to reflect predictive ability. A credit scoring system is

---

173 Brattle, 22 and 85.
174 Brattle, 193.
175 Brattle, 218: “Such information would include information on birth control and childbearing; information about marital status even when it is irrelevant to creditworthiness; or references to the race of the applicant or ethnic makeup of the applicant’s neighborhood.”
177 Using age ranges does not qualify as scoring ages in a credit scoring system.
178 Brattle, 118. The FRB was one of the ECOA regulators prior to the creation of the CFPB in 2011. The exception was added as part of Congressional direction in the Economic Growth and Regulatory Paperwork Reduction Act of 1996 (Pub. L. 104-208, 110 Stat. 3009). See, 64 Fed. Reg. 44,582 (Aug. 16, 1999).
179 Reg. B, 12 C.F.R. § 1002.2. See also, Wu, 715; Brattle, 308.
“empirically derived, demonstrably and statistically sound,” earning the industry acronym EDDSS.\textsuperscript{180} Self-testing is currently voluntary. To incentivize verification of models, Regulation B grants privilege to creditors as long as the testing process is documented and appropriate corrective action is taken.\textsuperscript{181} Chapter IV further discusses self-testing and possible improvements in the context of “alternative” credit processes.

Proving Wrongful Discrimination

To discriminate is to treat an individual or a group more favorably than another. As Deborah Hellman explains in her book \textit{When is Discrimination Wrong?}, discrimination itself is not illegal, but lawmakers can impose protections when discrimination is irrational and/or demeaning.\textsuperscript{182} For example, discrimination based on merit is generally accepted, perhaps since it is seen morally as “earned” and therefore it is neither irrational nor demeaning. In some cases, however, even rational discrimination can be wrong and these cases are most of interest for anti-discrimination in consumer credit. Rational discrimination, also known as actuarial fairness, occurs when, for example, women are charged higher life insurance rates because they tend to live longer than men.\textsuperscript{183} In this case, it is considered rational, relevant, and acceptable to consider an individual’s sex in making a decision about life insurance rates. It is inevitable, perhaps, that decisions have different, and in some cases, worse results for certain people or groups, and these decisions, when morally wrong, need to be rectified.\textsuperscript{184} Building on the prior example, if women with red hair were charged higher life insurance rates, and having red hair was not predictive of life expectancy, this would be considered wrong because the trait “red hair” would not be relevant, rational or accurate. It may also be considered demeaning towards women with red hair. While the word itself does not necessarily imply wrongfulness, when this thesis refers to “discrimination,” it refers to wrongful discrimination, demeaning and irrational or rational. When the ECOA prohibits discrimination \textit{on the basis of} certain characteristics, it prohibits discrimination \textit{because of} these characteristics, meaning that the characteristics are

\textsuperscript{180} Robinson + Yu, 30. The authors claim that the EDDSS is a best practice but is not a hard regulatory floor.
\textsuperscript{181} Reg. B, 12 C.F.R. § 1002.15: “To qualify for privilege, appropriate corrective action is required when the results of a self-test show that it is more likely than not that there has been a violation of the ECOA or this part. A self-test is also privileged when it identifies no violations.”
\textsuperscript{182} Hellman notes that there is disagreement and differences of interpretation on what is demeaning, and that society solves questions of interpretation and disagreement through democratic and judicial processes. Hellman, Deborah. 2008. \textit{When Is Discrimination Wrong}? Harvard University Press.
\textsuperscript{183} Schauer, 11. Schauer illustrates this as follows: “As long as the probability of a dog’s having hip problems given that the dog is a bulldog is greater than the probability of the dog’s having hip problems given no information about the breed of the dog, we can say that the trait of being a bulldog is relevant, and we can say that generalizing from that trait meets the minimum threshold of statistical (or actuarial) soundness.” See also, Lauer, 200 on the rise of statistical credit scoring in the 1960s. \textit{See also}, Barocas and Selbst: “By definition, data mining is always a form of statistical (and therefore seemingly rational) discrimination. Indeed, the very point of data mining is to provide a rational basis upon which to distinguish between individuals and to reliably confer to the individual the qualities possessed by those who seem statistically similar.”
\textsuperscript{184} Schauer, 47: “[...] although it is often necessary to make decisions according to generalizations, it is, says Aristotle, a mandate of complete justice that there be a way of rectifying these deficient results, so that in the final analysis the generalization is applied only in those cases in which the results it generates are accurate.”
the cause for the discriminatory action. The purpose of this section is to go through the process for proving discrimination under the ECOA.

The law is clear when there is direct evidence of discrimination: when discrimination is directly, intentionally, based on a prohibited basis, then it constitutes disparate treatment.\(^{185}\) In this case, the prohibited base is the cause of the discriminatory action. However, cause is not the only way to prove discrimination: discriminatory effect also can be illegal. Indirect discrimination is discrimination that appears neutral, but has disparate impact on a protected group. When there is indirect discrimination, the disparate impact test, also known as the “effects test,”\(^{186}\) is used to determine whether a difference between two protected groups violates the principle of equal protection. In sum, the disparate treatment test is about cause, and implies intent to discriminate, versus the disparate impact test, which is about effect, and does not imply intent.

Proving discrimination is difficult because the plaintiff must show a relationship between the discriminatory action and one (or more) prohibited bases. For example, a consumer cannot claim discrimination simply given that they were denied credit and a member of a protected group. The connection between the two facts is not always evident, and creditors may be able to use other information as a business justification for their actions. Brattle explains, “As most applicants have some kind of blemish on their credit record, employment history, or income potential, it is relatively easy for a creditor to point to a specific legitimate reason for denying credit to an applicant.”\(^{187}\) In the case of disparate treatment, once the cause is made evident, the business justification, if presented as a defense, is irrelevant. However, if the cause is not evident, and the plaintiff is claiming discriminatory effect (the case of disparate impact), business justification can be relevant. In the case of disparate impact, effect alone is not sufficient to establish intent.\(^{188}\) In addition to effect, a plaintiff must determine whether there was a legitimate business reason for the discrimination. If so, the plaintiff must also determine if there was a less discriminatory alternative.

---

\(^{185}\) Barocas and Selbst: “Disparate treatment comprises two different strains of discrimination: (1) formal disparate treatment of similarly situated people and (2) intent to discriminate.” See also, Brattle, 2 and 67: “Disparate treatment can be proven either through direct evidence or through indirect or circumstantial evidence.”

\(^{186}\) Official Interpretations of Reg. B, 12 C.F.R. pt. 1002, supp. I, § 1002.6(a)-2: “The Act and regulation may prohibit a creditor practice that is discriminatory in effect because it has a disproportionately negative impact on a prohibited basis, even though the creditor has no intent to discriminate and the practice appears neutral on its face, unless the creditor practice meets a legitimate business need that cannot reasonably be achieved as well by means that are less disparate in their impact.” Cited in, Brattle, 76.

\(^{187}\) Brattle, 69.

\(^{188}\) Hellman notes that there is disagreement and differences of interpretation on what is demeaning, and that society solves questions of interpretation and disagreement through democratic and judicial processes. Hellman, 196: “The Supreme Court’s disparate impact cases suggest that impact alone cannot be sufficient to establish intent — though it is unclear, at least theoretically, why not. In my view, disparate impact alone could establish the objective meaning of the law, but in the kinds of cases that have come before the Court — like Washington v. Davis — there is additional important objective evidence that cuts in the other direction. For example, the fact that there are good reasons to test police officers for the skills the test in question evaluate it is relevant to the objective meaning of law. Note that I’m not suggesting one ask whether these reasons actually motivated the police in adopting the test.” See also, Barocas and Selbst: “Disparate impact is not concerned with the intent or motive for a policy; where it applies, the doctrine first asks whether there is a disparate impact on members of a protected class, then whether there is some business justification for that impact, and finally, whether there were less discriminatory means of achieving the same result.”
In requiring creditors to retain files, one could argue that the ECOA creates a body of evidence to be used in discrimination cases, though direct evidence of disparate treatment is rare, so plaintiffs often rely on circumstantial evidence. As was referred to earlier, the McDonnell-Douglas test has become the standard test for proving discrimination. To satisfy the requirements of the McDonnell-Douglas test, a plaintiff must show:

1. Membership in a protected class;
2. Qualified application for credit;
3. Rejection despite qualification; and,
4. Other similarly-qualified applications that were approved.

Once the plaintiff has shown the elements above, the burden of proof shifts to the defendant: the creditor must show legitimate reasoning, non-discriminatory, for the rejection to have occurred.

When a disparate treatment claim cannot be made, the alternative may be a disparate impact claim. Since disparate impact focuses on discriminatory effect, an advantage of the test, from a consumer protection standpoint, is that it “broadens the reach of the prohibited bases listed in the ECOA […], and, in effect, creates many new prohibited bases.” When data are highly correlated with protected bases, then these data can become prohibited bases by proxy. The use of proxies is not illegal unless they are proxies for prohibited bases. A common and historical example of a proxy for race is zip code, given the history of redlining in the United States. In the absence of race information, consumer credit scoring models are tested for ECOA compliance using zip code data, since creditors collect zip code information, but do not generally collect race information. Creditors do not, however, use zip code information in decisions about credit because zip codes are proxies. Proxies are used in lieu of protected information and in lieu of other information that may not necessarily be protected, but is not easily accessible. Proxies are legal, but they are difficult to test, as Barocas and Selbst explain: “Part of the problem seems to be that there is no obvious way to determine how correlated a relevant attribute must be with class membership to be worrisome. Nor is there a self-evident way to determine when an attribute is sufficiently relevant to justify its consideration, despite its high correlation with class membership.”

---

189 Brattle, 85.
190 Note that there is no rule establishing the McDonnell-Douglas test as the standard test for proving discrimination. See, Brattle, 71.
191 Brattle, 70, with note on 72: “Almost all of these courts do not interpret the fourth prong of the McDonnell-Douglas standard literally.”
192 Brattle, 69: “When the factors that creditors utilize in differentiating credit applications have a disparate impact on a prohibited basis, these other seemingly non-prohibited factors can become prohibited bases. For example, a creditor’s practice may have a disparate effect on a protected category if the creditor denies credit to anyone who relies on child support payments as income. As this practice would affect divorced and separated women almost exclusively, it would have a disproportionately negative impact on the basis of sex and marital status and thus, in effect, becomes a prohibited basis.”
194 See infra note 313.
195 Barocas and Selbst.
legality of proxies poses a challenge for the robustness of the legal and regulatory context, to be discussed in Chapters III and IV.

Case law since the passage of the FCRA and the ECOA has not provided further clarity with regard to questions of discrimination by proxy. In 2000, the Third Circuit suggested a distinction between more or less correlated proxies, as a way to indicate that the more correlated proxies could serve as stand-ins for protected attributes in a disparate treatment case. In *Erie Cty. Retirees Ass’n v. Cty. of Erie, Pa.*, Medicare eligibility was found to be an actionable proxy for age, whereas other proxies were not. This case does not clearly define a standard by which to establish which proxies are more or less correlated. Without providing further evidence or explanation, Pleiss et al. suggest that disparate impact doctrine “states that the prediction rates for any two groups should not differ by more than 80%.” Where “alternative” credit models may use many types of data, neither case law nor existing regulation provide clarity as to how proxies will be evaluated.

More recently, in a 2015 Supreme Court case on disparate impact, *Texas Dep’t of Housing & Cmty. Affairs v. Inclusive Communities Project, Inc.*, the Court ruled that organizations are not “liable for racial disparities they did not create” and that there should be a causal relationship between the policy and resulting discrimination, even in an effects test. Given this ruling, organizations could make the argument that they or their models only reflect existing societal discrimination and are therefore not responsible for discriminatory practices. This ruling might also suggest that organizations are not responsible for curbing systemic discrimination. Such a ruling may further raise the bar in terms of proving wrongful discrimination since the use of proxies is not clearly defined and replicating existing discrimination may not be viewed as problematic.

The FCRA and the ECOA envision a fair credit system that ensures trust, expands credit, and is more financially inclusive. *Procedural* fairness focuses on the credit process, how data are collected, protected against invasions of privacy, corrected when false, and made available when requested. *Substantive* fairness, focuses on who gets access to credit, and how the kinds of data used can have wrongly discriminatory results for certain groups. The implementation of prohibited bases is a foundational mechanism of fair credit. In requiring creditors to be “blind” to the characteristics that were most vulnerable to discrimination, the ECOA represented a

---

196 *Erie Cty. Retirees Ass’n v. Cty. of Erie, Pa., 220 F.3d 193, 211 (3d Cir. 2000). See also, Taylor, Winnie. 2012. “Proving Racial Discrimination and Monitoring Fair Lending Compliance: The Missing Data Problem in Nonmortgage Credit.” Review of Banking & Financial Law 31:199–264. Page 245: “Theoretically, proxies could supply the absent data in some cases. However, in practice, courts have not accepted general population statistics or census tract data as sufficient proxies for the racial composition of a lender’s applicant pool. Plaintiffs have also been unsuccessful in their attempts to use zip codes as a proxy for racial information.”

197 Pleiss et al.


200 Inclusive Communities Project, Inc., 135 S. Ct. at 2523.

201 See, Hurley and Adebayo, 194.
simple way to address fairness, and might have provided an elegant solution to discrimination brought to light by the Civil Rights Movement in the late 1960s. The logic was as follows: if the information was not collected, then it could not be used, and then it was not possible to discriminate on the basis of that information.

Unfortunately, that is not reflective of actual practices. Regulatory interpretations of the ECOA and nearly four decades of case law have established various methods for proving discrimination, by either using prohibited bases as the cause, or by having a discriminatory effect on members of protected groups. But even with the case law as examples to build from, there is some uncertainty around the standing for the disparate treatment and disparate impact tests moving forward, and how they will evolve in light of technological advances. In the context of “alternative” data and modeling techniques, these fair credit goals are still valid, but the mechanisms by which to achieve them are no longer. Aside from the tests, themselves also perhaps mechanisms of the ECOA, the policy goals from which these tests originate remain grounded in the ECOA and other anti-discrimination law.

Revisiting a Twenty-Year-Old Proposal

Whether “blindness” is an effective substantive fairness mechanism is a question that has been asked before in the context of traditional credit processes. In producing the ECOA’s implementing Regulation B, regulators considered how to implement the ECOA’s prohibition. They determined that there was too much risk associated with allowing the data to be collected, even if not used in decisions, and therefore implemented the ECOA with an all-together ban on the collection of prohibited information.²⁰² In both 1995 and 1998, the FRB, lead regulatory agency at the time,²⁰³ considered changing Regulation B to allow for collection of this information. In 1995, the proposal was to allow, but not require, the collection of prohibited information.²⁰⁴ It is not clear what motivated the proposal in 1995, but it was retracted within a year, with the FRB suggesting instead that such a change should be the prerogative of Congress.²⁰⁵ In 1998, following a request by Congress to evaluate the effectiveness of prohibited bases, the

²⁰² Taylor, 217: “When the FRB drafted the ban in 1977 to implement the 1976 ECOA amendments, it considered whether permitting all lenders to collect applicant personal data would advance or impede the ECOA’s antidiscrimination goals. Concluding the latter, the FRB decided on a general prohibition against data collection “on the theory that if creditors did not have this information, they could not use it to discriminate against applicants.”

²⁰³ Until the creation of the CFPB, the FRB set the most standards and rules related to the ECOA. The FRB’s version of Regulation B is the version adopted by the CFPB and still in use today.

²⁰⁴ Taylor, 217: “Perhaps the FRB recognized this infirmity when it drafted proposals to lift the ban in 1995 and 1998. The 1995 proposal would have amended Regulation B to allow, but not require, creditors to collect...”

²⁰⁵ Taylor, 218: “In December 1996, after receiving and considering more than 250 comment letters, the FRB withdrew this proposal, concluding, without elaboration, that given the political sensitivity of the issues involved, it might be more appropriate for Congress to decide whether the data collection ban should be lifted. Three years after announcing that it would defer to Congress on whether the data collection ban should be lifted, the FRB again proposed removal of the restriction. Like the previous proposal, the 1998 proposal did not have a mandatory data collection requirement; it merely proposed lifting the ban to allow nonmortgage creditors to collect race data voluntarily.”
FRB recommended, as a result of their study, removing the prohibited bases and allowing consumers to voluntarily supply such information, with the eventual goal of making it mandatory.\textsuperscript{206} The proposal was controversial: banking institutions were against the proposed change since they believed it would increased their regulatory compliance burden and be an invasion of consumer privacy.\textsuperscript{207} They were also concerned about the increased risk of litigation should the rule go into effect.\textsuperscript{208} On the other hand, regulators, government agencies, and advocacy groups favored the change.

In 2003, after further consideration, the FRB decided not to move forward with the proposed rule. Instead, as part of revisions to Regulation B, the FRB added incentives to increase self-testing of credit scoring models, for which the use of prohibited bases is excepted.\textsuperscript{209} As described in Chapter II, the added incentives included granting privilege to companies, regardless of the results of their testing, so long as the testing and any corrective action is documented. The FRB’s rationale was that retaining the prohibition was most effective against discrimination. As Taylor explains: “The FRB justified its decision on two primary grounds. First, the FRB said that retaining the ban was justified because the proposed voluntary data collection approach would not produce useful market-wide data. The FRB worried that many creditors would elect not to collect the data while those that did collect it would use inconsistent standards, criteria and methods. Consequently, the data would be of questionable utility because there would be no assurance of its accuracy nor would there be any way to compare it from creditor to creditor. The FRB’s second justification for retaining the ban was that allowing data collection would create opportunities for nonmortgage lenders to use the data in a discriminatory manner.”\textsuperscript{210} A criticism of the FRB’s proposal, in both cases, is that it focused on voluntary data collection, as opposed to the mandatory collection.\textsuperscript{211}

We recognize today that, even where equality is prescribed, differences are manifested in other ways, not just in the particular class of information to which we prescribe an equal value. The difference between men and women, for example, is not simply manifested in the data point about their sex, but also in their behavior, and possibly could manifest in other types of protected information as well. Some of these differences manifest naturally, such as childbirth, and others are societal factors, where differences are echoed even when the policy in one context prescribes equality of the sexes. In the case of blind auditions, for example, candidates’ gender manifested in the sound of their shoes as they walked behind the screen designed to “blind” judges from knowing the candidate’s gender. Shoe sound, such as the click of high-heels, was a

\textsuperscript{206} Taylor, 217: “The creditor must disclose that the applicant is not required to provide this information, that the creditor is requesting the information to monitor compliance with the ECOA, and is prohibited from considering this information or the applicant’s decision not to furnish this information in making the credit decision, and, when applicable, that this information will be collected based on visual observation or surname if not provided.”

\textsuperscript{207} Brattle, 111.

\textsuperscript{208} Brattle, 118: “Assuming corrective action is taken when warranted, the results of the self-test cannot be obtained by a government agency in an examination or investigation or by an agency or any applicant in any proceeding or lawsuit alleging a violation of the ECOA or Regulation B. The latter point appears to address creditors concerns about the risk of litigation if the ban had been lifted as proposed.”

\textsuperscript{209} Brattle, 118.

\textsuperscript{210} Taylor, 222.

\textsuperscript{211} Taylor, 224.
proxy for gender, and allowed judges to infer gender. This was not an expected proxy: it emerged in the process. In a world with “alternative” data and modeling techniques, where more data are being collected on individuals and tools are designed to surface correlations among these data, it is no longer effective to be “blind” to certain data, but it is possible to mitigate risks of discrimination by testing and verifying models using sensitive information.

For the traditional credit processes, years of regulations have helped manage risks associated to procedural and substantive fairness. With “alternative” credit processes, it is not yet clear what the risks are, and whether the existing regulations will be sufficient in managing them. The late 1960s and early 1970s established a framework for procedural and substantive fairness that has evolved over nearly fifty years, yet problems of inequality and discrimination persist today. In the words of Eubanks, “the Poor People’s Campaign is one of our nation’s great unfinished journeys. Its aspirations are as pressing today as they were 50 years ago.”212 Traditional inputs and traditional models, while useful from the standpoint of standards and accountability, as discussed in the context of FCRA reforms earlier in this Chapter, can lead to disproportionate results for certain segments of the population that do not fall squarely in these categories. New data types and modeling techniques could change that and expand access to credit. The following Chapter proposes a framework in which to assess these “alternative” credit processes.

---

212 Eubanks, 204.
III. Assessing “Alternative” Credit Today

“After decades of hard-fought legal battles, consumers seemed to be back where they started. Laws passed in the 1970s to rein in the information maximalism of credit reporting and to end discriminatory lending were at risk of being circumvented altogether. Instead of local hearsay or prejudiced credit managers, which colored past conceptions of creditworthiness, Americans would be judged by even more inscrutable surveillance networks and algorithms.”  

While one of the fair credit goals is credit expansion and financial inclusion, there are still an estimated 45 million people in the United States who are underscored by the existing system. These consumers are primarily from marginalized communities. There is not enough information about these consumers to build a credit profile. Inequalities in access to credit compound other inequalities they may already be facing. When these consumers are unable to access opportunities linked to their credit, such as their ability to find a place to live, commute, and secure employment, they are forced, in many cases, to resort to obtaining credit through lenders who may use unregulated or predatory practices. Such action can further hinder their ability to access opportunities or the credit market at a later date. To be considered worthy of credit, a consumer needs to demonstrate good credit practices such as paying back loans. However, in order to even get these loans at the outset, that consumer often needs to be considered worthy. For consumers with little or no credit history, primarily young and/or marginalized adults, the credit building process can prove to be difficult.

There are patterns of credit invisibility among specific consumers. In the CFPB “Credit Invisibles” report, the authors identified marginalized groups by age, ethnicity, and income: consumers aged 18-24 (somewhat expected since credit history is established over a lifetime) and 70-older; low-income; black or hispanic. For any of these demographics in combination, the effects are even more stark: “these differences are observed across all age groups, suggesting that these differences materialize early in the adult lives of these consumers and persist thereafter.” These consumers end up stuck in a perpetual cycle of poverty. One idea is to break this cycle by considering “alternative” data types, by collecting and using many more data about

---

213 Lauer, 267.
215 Brevoort et al.
216 Robinson + Yu.
217 See, Credit Builders Alliance.
218 Parker et al., 277. See supra note 33 for quote.
219 Brevoort et al., 6.
an individual in the formation of their consumer credit report and score. Using “alternative” data and modeling techniques may allow more Americans to access credit, especially those who have been underscored by the traditional system.

“Alternative” credit scoring was originally promoted as having greater benefits to the consumer, since using more data could paint a different profile of an individual who may not have the “right” profile using traditional credit score information. And in some cases, they do. The purpose of this Chapter is to discuss the advantages and risks associated to “alternative” credit processes. There are three primary advantages to “alternative” credit scoring: increasing access to opportunity; identifying insights with new data and modeling techniques; and driving down costs, both for the creditors and consumers. These advantages apply primarily to consumers underscored by the traditional credit scoring system, and while these “alternative” credit scoring systems are promoted as more financially inclusive, they are also advantageous for companies in providing new products, collecting more consumer information, and growing their consumer base.

While there are many exciting aspects of “alternative” credit scoring, there are also risks to consider. Some of these risks create a tension between what is practically feasible and ethically acceptable, and can outweigh the benefits of increasing access to credit. While there is great potential to better serve marginalized communities, there is also concern that unregulated and untested products could disproportionately affect them, and cause harm, through discriminatory behavior, intentional or not, and violations of personal data privacy in the development and use of consumer credit profiles. While financial inclusion is an important policy from the regulatory perspective, it may also be premise for expanding business if not properly targeted. Consumer credit reports are business products built on information that is collected, packaged, and sold. Credit scoring companies are competing to sell the most comprehensive and predictive profiles. The inclusion of “alternative” data can make these profiles more valuable and competitive. Credit products developed from “alternative” data have

220 This idea is not new. In the 1850s, creditworthiness used to be evaluated with many more criteria than more recent credit scoring systems: the Bradstreet Company used over 36 criteria in the early 1850s, versus FICO, which listed 5 criteria on its website in 2014. See, Olegario, 212.
222 More comprehensive credit information would allow for creditors to make better decisions and, theoretically, decrease the default rate of their loans. For consumers, the cost of borrowing may decrease since they might have access to more credit products at lower interest rates, based on more personalized assessments. Consumers can apply for credit online and get an almost-immediate assessment and line of credit.
not necessarily been more financially inclusive, or “better alternatives”225 than traditional credit products.

Some problems are not unique to traditional or “alternative” credit scoring, though they can be exacerbated by it. As discussed in Chapter I, a primary source of unfairness are data. These data may be poorly collected or labeled. They may also be poorly correlated. They may be proxies. They may be based on other people, who are considered similar theoretically, but might not be in practice.226 In addition to data quality, data privacy is also a risk. Chapter II described how consumer privacy was the primary motivation for the passage of the FCRA. Privacy was also a chief concern during the 1996 FCRA reforms. And, privacy still is a concern today.227 Consumer advocates argue that these “alternative” credit scoring processes may be too privacy-invasive, and in some cases, just as bad, if not worse, in terms of transparency, than traditional credit processes.228 This Chapter proposes a framework through which to assess “alternative” credit, using the technical background from Chapter I and the legal background from Chapter II.

Since there are a myriad of possibilities for “alternative” data and models, this Chapter proposes a framework in which to consider them and the risks they may pose across the credit system.

First, the framework focuses the primary consideration for credit processes to be predictive of credit risk. While companies may want to consider credit profiles for other applications, their primary purpose is to predict creditworthiness.229 Second, the framework looks at how

---

225 Yu et al., 31: “Some of the features of these loans are arguably “less bad” than those offered by traditional payday lenders, but these products still fail to meet the requirements to be considered genuine, better alternatives. They still feature three-digit APRs. With the exception of LendUp and MySalaryLine, all products accept installment payments; however, many of them require weekly or biweekly payments rather than monthly ones.”

226 O’Neil, 145: “Fair and Isaac’s great advance was to ditch the proxies in favor of the relevant financial data, like past behavior with respect to paying bills. They focused their analysis on the individual in question — and not on other people with similar attributes. E-scores, by contrast, march us back in time. They analyze the individual through a veritable blizzard of proxies. In a few milliseconds, they carry out thousands of “people like you” calculations. And if enough of these “similar” people turn out to be deadbeats or, worse, criminals, that individual will be treated accordingly.” See also, Barocas and Selbst: “By definition, data mining is always a form of statistical (and therefore seemingly rational) discrimination. Indeed, the very point of data mining is to provide a rational basis upon which to distinguish between individuals and to reliably confer to the individual the qualities possessed by those who seem statistically similar.”

227 Opening statement of Rep. Cornelius E. Gallagher in “Commercial Credit Bureaus:” “For we must not become data rich by becoming privacy poor. There is a growing compulsion today to make each action by every single American a permanent record. Once it is known that this record exist, there is a tendency for some person, inside or outside the government, to discover another purpose for that record. Because his goal and using the data frequently may be noble and public spirited, he will not consider confidentiality to be an overriding concern. He will ignore the fact that date it was originally gathered and used solely for administrative functions.” See also, Harris, Louis and Associates and Westin, Alan F. 1990. “The Equifax Report on Consumers in the Information Age.” New York, NY: Louis Harris and Associates. Cited in ’Amendments to the Fair Credit Reporting Act: Hearing before the subcommittee of Consumer Affairs and Coinage of the committee on Banking, Finance, and Urban Affairs, House of Representatives,” One Hundred First Congress, second session. 1990. Washington, DC: United States Government Printing Office: “The generation of Americans who have grown up with computers at their fingertips -- 18-29 year olds -- are more enthusiastic about the advantages of computer-based consumer services then the pre-computer generations. They are also more willing to see consumer information used, with existing safeguards, for credit, credit card, insurance, and direct marketing uses. But when it comes to general concerns about privacy, the computer-age generations joins its elders in voicing strong pro-privacy positions.”

228 Robinson + Yu.

229 According to Yu et al., 20, the FCRA does not require that credit scores be predictive, but the ECOA’s Regulation B does provide some insight as what a predictive score would be. In Regulation B, a credit scoring system must meet the following requirements: it must be developed with a significant portion (or the entire) consumer credit file; be used to
compliant a certain data type may be, so that stakeholders can make decisions about which types of data may be more or less acceptable to use. To assess compliance, this framework uses the three goals for fair credit (procedural, substantive, and financial inclusion) and their currently-implemented mechanisms.

In some cases, where the data may not be compliant with mechanisms in the FCRA or the ECOA, such as the “blindness” mechanism, then is it clear that these data would not be acceptable to use. With regard to the use of proxies, there is perhaps less clarity since proxies are not precisely defined. In these cases, companies are left to determine their level of risk and defend their position should there be an examination of their models. Finally, one of the main reasons “alternative” credit processes are of interest is that they could help with credit expansion and financial inclusion goals. As companies and regulators evaluate “alternative” data, they may want to consider whether the inclusion certain data would decrease the number of underscored people in the United States, and whether that is worth a possible trade-off, if at all, in privacy, accuracy, transparency, explanation, etc. Together, these considerations form the rows of the proposed framework (see Table 1, below).

The columns of the framework consist of traditional and “alternative” data types, allowing for comparison of risk considerations. Since credit scores are modeled on data from credit reports, the FCRA and ECOA focus on data regulation, and less on models. At the time, the logic was perhaps that if the data in credit reports were accurate and non-discriminatory, then the data used in developing credit scores, and the credit scores themselves, would also be accurate and non-discriminatory. As discussed in Chapter I, data is indeed a source of unfairness in the machine learning pipeline, but modeling techniques can also be unfair. With the rise of “alternative” credit data and modeling methods, and perhaps even before, this logic does not apply as simply. This Chapter will focus on data for the proposed framework, but, to be more comprehensive, will also address model unfairness.

Categorizing “Alternative” Data

Increasingly, companies have touted “big data” approaches to consumer credit: the more data a company can amass on an individual, the more targeted the credit decision can be. In short, the idea behind “alternative” credit scoring is: “all data are credit data.” “Alternative” credit scoring could consider many data types, broadly classified as either payment data or personal

---

230 Most attention to model is in Regulation B, which provides detail on credit scoring systems and their self-testing requirements.

231 See, Hurley and Adebayo.
Traditionally, credit scores are derived from consumer credit reports: personalized files that contain information on loan repayment, sources of debt, public records for civil judgments and bankruptcies. There are other types of non-loan payment data that may be as predictive as loan repayment data, and more common among underscored consumers, such as mobile phone, utility bill payments, or other regular payments. Personal data is more broad and could consist of a consumer’s social network, educational background, location data, shopping habits, etc. Personal data does not necessarily need to identify a consumer by name, though this depends on how the FCRA is interpreted. It is not yet certain that “alternative” payment data would be compliant with the key mechanisms in the FCRA and the ECOA, though it is likely more compliant than “alternative” personal data.

Payment Data

“Alternative” payment data is already in use by both incumbent and newer companies in the consumer credit ecosystem, though newer companies also include “alternative” personal data in their models. Of the incumbents, VantageScore started offering its VantageScore 3.0 score that factored in rent and utility data over five years ago, and reported in December 2018 that “unconventionally scored consumers do not default quicker or exhibit an increased risk when opening and paying new loans.” The FICO XD score incorporates data from National Consumer Telecom and Utilities Exchange (NCTUE) and LexisNexis. Experian suggested that, given that 96% of adults in the United States have a mobile phone, incorporating this data into a creditworthiness assessment may be more inclusive. “Alternative” payment data is generally predictive and would likely help with financial inclusion goals.

---

232 It may not always be possible to distinguish between the two, a risk discussed later in this Chapter. See also Robinson + Yu, who classify the data as “mainstream alternative data” and “fringe alternative data,” respectively. Hurley and Adebayo also prefer that classification.

233 The exact characteristics differ by credit scoring company. For example, the Fair and Isaac Corporation’s (FICO) characteristics are payment history, outstanding debt, length of credit history, pursuit of new credit, and debt-to-credit ratio. While the data used to determine these scores are regulated, the exact formula is itself proprietary. For an outline of baseline credit data, see Robinson + Yu, 8.

234 Robinson + Yu, 10.

235 Robinson + Yu, 10.

236 In McCready v. eBay, Inc., the Court ruled that a consumer under the FCRA had to be an identifiable person. However, the Federal Trade Commission (2011) interpreted the FCRA more broadly. See also, Federal Trade Commission. 2014. “Data Brokers: A Call for Transparency and Accountability.”


238 Wu, 721: “FICO has developed a specialty score for consumers who have no files or “thin” files at the nationwide consumer reporting agencies, called FICO XD. This score primarily relies on information from specialty consumer reporting agencies, including National Consumer Telecon and Utilities Exchange (NCTUE) and LexisNexis.”


With regard to accuracy, payment data is related to the consumer and verifiable.\textsuperscript{240} From a privacy perspective, using payment would not necessarily be an invasion of privacy, if consumers consented to this information being used. The FICO XD score, for example, currently does request consumer consent, but this might not be the case for other scores. Payment data would also be likely be compliant with the FCRA restrictions on obsolete or adverse information, though it may be worth considering if credit models would penalize late payments or consumption types.

For example, how would a consumer’s energy usage or type (e.g., renewables versus fossil fuels) be considered, if at all? Today, some creditors give preferential treatment to online applicants:\textsuperscript{241} this is a concern, and possible example of disparate impact, since Internet access is not equally distributed in the United States, affecting black, hispanic, older, disabled, and public assistance populations.\textsuperscript{242} According to a comment submitted by LendUp in the CFPB’s 2017 Request For Information, lack of Internet access appears to have a stronger relationship to credit invisibility than does the presence of a bank branch.\textsuperscript{243} Where access to certain energy types is also not equally distributed, preferential treatment could worsen impact on marginalized consumers. If such treatment had disparate impact on credit access for certain protected classes, it would pose concerns for ECOA compliance as well. To be sure, there would need to be regular testing done to determine whether certain “alternative” payment data correlate with prohibited class information.

Personal Data

There are some aspects of “alternative” payment data that may not be compliant with the FCRA and the ECOA, though further examination may provide more clarity. Compliance is much more challenging with regard to “alternative” personal data. These data have not been proven to be predictive of credit risk, help with inclusion goals, or compliant with the FCRA and the ECOA. Part of the challenge is lack of clarity as to how personal information is interpreted. If personal information is understood in the broadest sense, then compliance is very difficult since there are many proxies and methods to re-identify characteristics of an individual. Data quality plays a key role and can create conflicts between the FCRA and the ECOA.\textsuperscript{244} Depending on where and

\textsuperscript{240} Note that a 2014 study found that information from data brokers was mostly inaccurate. The assumption in this case, however is that creditors would use information from utilities, so this information would be more accurate and verifiable than compiled information from brokers. See Yu et al..

\textsuperscript{241} Brattle, 61.

\textsuperscript{242} See, Yu et al., 27.


\textsuperscript{244} Wu, 189.
how the data are collected, the results could be more or less accurate,
though attempts to verify the data could reveal protected class information.
For example, attempts to verify employment, one of the factors in the Equifax Decision 360 model, could reveal a consumer’s dependence on public assistance programs, which is a prohibited basis for credit decisions. It is also possible that models make “spurious correlations between a particular attribute and creditworthiness.” Such correlations are concerning in general, but especially so if that attribute is inaccurate.

Personal data may impede on consumer privacy and limit a consumer’s control over their information, even if consent is given. In evaluating social network data, as LendUp, Earnest, and Kreditech’s models do, a consumer may have given consent to their information being used, but that consent does not necessarily include the use of their friends’ information. Personal data may also not be in compliance with the adverse and obsolete information mechanisms of the FCRA. Since credit scores are generated from consumer credit reports, they change and are updated as a consumer’s report changes. As described in Chapter II, favorable information can remain on a consumer credit report, but adverse or obsolete information must be removed after seven years. This gives consumers an advantage: their most favorable profile is put forward to creditors. Having time-limits on negative information also indicates that policymakers wanted to give consumers the ability to renew their credit profiles to what amounts to about a dozen times throughout their expected adult life. However, there are varying degrees of permanence when it comes to personal data. An individual’s social network may change throughout their adult life, but is less likely to do so for lower-income or marginalized individuals. Including social network data could therefore end up furthering inequality when a purpose of “alternative” processes to be more inclusive. The inclusion of personal information may not be in compliance with FCRA goals or mechanisms.

The idea of using personal information in creditworthiness evaluations is not new. In the 1850s, the Bradstreet Company’s criteria for creditworthiness focused more on elements of moral character, which were difficult to evaluate fairly. The passage of the FCRA and the ECOA was

---

245 See e.g., Experian report cited in Hurley and Adebayo, 153.
246 Yu et al., 19.
247 See Robinson + Yu, 13.
248 Hurley and Adebayo, 177.
249 See, Robinson + Yu, 15.
250 See supra notes 160 and 161.
252 Olegario, 216.
essential to protecting consumers for unfair creditworthiness evaluations and assumptions about consumers moral character that were discriminatory or untrue. In creating its credit score, FICO was considered an improvement on the consumer credit system since its model “measured ‘character’ only indirectly, through the individual’s recent payment history.”\textsuperscript{253} It is not clear that returning to creditworthiness models more akin to those from the 1850s is the best way forward.

**Proposed Framework**

Regulators have provided guidance on such matters for traditional credit scoring models, some of which could be applied to the “alternative” credit processes as well. Following the 2017 Request for Information on the topic of “alternative” credit processes,\textsuperscript{254} it is also likely that the CFPB will issue specific guidance. Per the discussion above, Table 1 is a tabular representation of a framework to assess “alternative” credit, for consideration.

\textsuperscript{253} Olegario, 216.
\textsuperscript{254} Consumer Financial Protection Bureau (2017).
Table 1. Framework for assessing “alternative” credit in the existing consumer credit context.

<table>
<thead>
<tr>
<th>Are the data predictive of credit risk?</th>
<th>Traditional</th>
<th>“Alternative”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payment data data</td>
<td>Yes.</td>
<td>Yes.</td>
</tr>
<tr>
<td>Personal data</td>
<td>Yes.</td>
<td>Yes.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Are the data compliant?</th>
<th>Traditional</th>
<th>“Alternative”</th>
</tr>
</thead>
<tbody>
<tr>
<td>to determine, see next</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Are the data compliant with the FCRA mechanisms?</th>
<th>Traditional</th>
<th>“Alternative”</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Accuracy</td>
<td>Yes.</td>
<td>Yes.</td>
</tr>
<tr>
<td>→ Privacy controls</td>
<td>Yes.</td>
<td>Yes, with consent of consumer.</td>
</tr>
<tr>
<td>→ Restrictions on obsolete or adverse information</td>
<td>Yes.</td>
<td>Yes for obsolete information, though would need to evaluate how data types could be considered adverse information.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Are the data compliant with the ECOA mechanisms?</th>
<th>Traditional</th>
<th>“Alternative”</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ No prohibited bases or proxies</td>
<td>Yes.</td>
<td>Would need to test to confirm.</td>
</tr>
<tr>
<td>→ Notices of adverse actions</td>
<td>Yes.</td>
<td>Dependent on other compliance mechanisms.</td>
</tr>
<tr>
<td>→ Record retention</td>
<td>Yes.</td>
<td>Dependent on other compliance mechanisms.</td>
</tr>
</tbody>
</table>

When the framework is populated with “alternative” credit processes and existing mechanisms, these “alternative” processes are not compliant, or at least indicate greater risk and uncertainty in their application to the credit system. Compliance suggests a focus on mechanisms, though these mechanisms are outdated. It may be that “alternative” processes present greater risk and uncertainty. A resulting cautionary approach could challenge the idea in machine learning that more data are better than less, since, as developed in Chapter I, more data generally allow for more comprehensive model design. Including more data, while perhaps allowing for more

---

See, Domingos.
comprehensive models, may not be in compliance with the FCRA or the ECOA. Compliance would require limiting and controlling much “alternative” data since these data can be proxies for prohibited bases, meaning that they can be, alone or in combination, highly correlated to information protected from use in credit decisions. Including these data would therefore violate the ECOA as it currently stands. However, it may also be the case that “alternative” processes are being evaluated with outdated mechanisms. These outdated mechanisms may not provide enough risk mitigation, or could be preventing the expansion of credit in their limitations.

Model Examination Procedures

Since most of the proposed framework suggested a need for further testing to determine compliance, this section addresses how such testing occurs today. There is little known about how often and how companies test their credit scoring models, though they must do so to ensure compliance with the ECOA. This section will review the CFPB Examination Procedures and Baseline Review Procedures, which are publicly accessible, in order to understand how evaluation occurs from the regulators’ perspective. It is likely that companies use these examination procedures as a baseline to develop their own testing procedures. Using this information about current practices, Chapter IV will develop how such testing could occur in the future.

There are two primary ways in which credit scoring models are tested. First, creditors are incentivized to self-test their models, allowing for some self-regulation. Second, regulators can examine models, usually following complaint or as part of a Baseline Review, discussed below. While creditors are exempted from prohibitions on use of protected class information in their model self-testing processes, it is unclear whether they choose to use this information, since there is little disclosure of such information by creditors. VantageScore, for example, suggests that they use zip code information approximated to census data. This also appears to be the

---

257 The idea that not all data are created equal, and that some data deserve more protection than others, is not new: this idea is featured as a mechanism of the ECOA. Within consumer credit legislation, there is precedent for protecting against the use of certain data types: these are the “prohibited bases” in the ECOA.
260 This is possibly due to concerns about loss of privilege. Privilege is granted as part of the self-testing process, but can be lost if creditors reveal too much information that is supposed to be privileged. Official Interpretations of Reg. B, 12 C.F.R. pt. 1002.15: “To qualify for the privilege, appropriate corrective action is required when the results of a self-test show that it is more likely than not that there has been a violation of the ECOA or this part. A self-test is also privileged when it identifies no violations.” Cited in, Brattle, 349.
261 See VantageScore’s “Testing Credit Scoring Models for Statistical Bias” Page 3.
method in use at the CFPB for examination or study.\textsuperscript{262} If this is similar to other creditors, the self-testing is currently is less efficient and trustworthy than it could be. Unless examined by the CFPB, there is no guarantee that these self-tests are occurring.\textsuperscript{263} how often, how, and by which metrics. Chapter IV proposes remedies to bolster self-testing. This section will lay the groundwork for these remedies by describing how regulators examine models.

The CFPB uses its Baseline Review Procedures to identify risks of ECOA violations and inform agency enforcement priorities. Baseline Reviews are scheduled and depend on information requests for which the CFPB has standardized “click & type” forms. There are five modules in the Baseline Review:\textsuperscript{264}

- Module 1 Fair Lending Supervisory History
- Module 2 Fair Lending Compliance Management System (CMS)
- Module 3 Fair Lending Risks Related to Origination
- Module 4 Fair Lending Risks Related to Servicing
- Module 5 Fair Lending Risks Related to Models

Module 5 is of particular interest as part of answering the questions around self-testing. In this Module, review questions focus on the use of models, such as:\textsuperscript{265} “How often are the entity’s models validated or re-validated? Please describe the nature of validation or re-validation.” and “Does the entity conduct any fair lending related review or testing of models? If so, please note the frequency and nature of such review or testing; the part(s) of the entity responsible for such review or testing; the results of the last review or testing performed; and any corrective action(s) taken.” These questions, and others, review creditors’ model validations and verify that they remain in compliance with the ECOA.

In the case of a formal fair lending examination, the CFPB has similar standardized “click & type” forms, though these go into greater detail and require more information. To compare the processes briefly, the Examination Procedures document is twice the length of the Baseline Review Procedures document. Prior to initiating an examination, the CFPB scopes a specific area for examination and requests data to conduct its assessment. This information can be treated confidentially. For example, when the CFPB examined the “alternative” credit practices of Upstart in 2017, the company made a request for confidential treatment, and submitted one letter with public information and another with further detail, presumably on their data and modeling techniques.\textsuperscript{266} In its first-ever no-action letter response, the CFPB agreed to allow Upstart to use “alternative” credit processes so long as Upstart continued to report on its compliance with fair lending practices.\textsuperscript{267}

\textsuperscript{262} Brevoort et al.
\textsuperscript{263} Self-testing of models is voluntary.
\textsuperscript{264} Consumer Financial Protection Bureau (2019), 1.
\textsuperscript{265} Consumer Financial Protection Bureau (2019), 24.
While there is extensive documentation on the questions asked during Baseline Review and Examination Procedures, there is little information documenting the threshold for answers. For example, Upstart noted a risk of its model not being equally predictive across all demographic groups, and noted a commitment to fair lending practices, but it not clear to what extent this unequal treatment was acceptable to Upstart or to the CFPB, and what type of corrective action was required, if any. The lack of clarity around the legality of proxies poses a challenge for the robustness of the legal and regulatory context. The number of proxies will increase with “alternative” data and models as will correlations to protected attributes. These data and models are generally not audited until a complaint has be brought forth, and such a process does not engender trust. Testing with approximations for the actual data is suboptimal when the actual data can be collected. The ECOA’s prohibited bases are so protected that their existence promotes a form of “blind,” yet ignorant and ineffective, justice, especially in the wake of “alternative” credit processes.

This Chapter presents an overview of how regulators could assess “alternative” credit processes in using existing mechanisms and examination procedures. The conclusion is that “alternative” data and models are not likely to be compliant given existing mechanisms. The tension for policymakers appears to lie in the choice between cautious and effective policy. As technology advances in the face of inaction, the chasm between the choices increases, placing consumers at risk of harm and decreasing trust in the system. Chapter IV presents another scenario, one in which the context adapts to technological advances. In doing so, Chapter IV differentiates between goals and mechanisms, recognizes that the existing mechanisms are no longer the most effective way of mitigating risks for advances in “alternative” credit processes, and proposes changes to the mechanisms so that they may be more effective.

---

268 Upstart, 8.
IV. Beyond “Blindness” and Other Outdated Mechanisms

“[...] the claim that all are created equal is a claim that, descriptively, is simply untrue. Yet although it is false that all men are created equal, it is not false that it is possible to treat all men as being created equal even if they are not.”

In examining decision-assistive tools that use machine learning to process massive amounts of data, there is concern about inadvertently reproducing discrimination instead of eliminating it. Similar concerns about discrimination and fairness have motivated regulation in the past, notably in the late 1960s with the passage of the FCRA and the ECOA. The FCRA set goals for procedural fairness in a credit system that is developed on accurate and relevant information, with controls for consumer privacy. The ECOA envisions substantive fairness in a credit system that is equally accessible by all consumers. A third goal, perhaps most notably expressed in the creation of the CFPB, is to expand credit and be more financially inclusive. These three goals are still valid today, yet the mechanisms and regulations by which they are meant to be achieved have not kept pace with technology and do not help mitigate concerns about unintended discriminatory consequences of machine learning applications to consumer credit.

From a technical perspective, there are a variety of proposals to address unfairness and restore trust in machine learning applications. These are described in Chapter I and include managing input data by ensuring its diversity and respect for context, formalizing fairness definitions in algorithms, auditing outputs, or restricting the use of such tools in contexts where discriminatory results would be too harmful. These techniques, while useful research advances, lack context-specific application. The review of historical, legal, and regulatory background in Chapter II provides a context in which to apply these technical advances.

Unfortunately, the consumer credit context is itself disconnected from today’s risks and concerns brought forth by “alternative” credit processes. Machine learning models are inferring and predicting sensitive information. This finding raises concerns for machine learning

---

269 Schauer, 217: “Equality, as has often been recognized, has both descriptive and prescriptive dimensions. [...] yet although descriptive and prescriptive equality are different, it is important to see that claims of prescriptive equality coexist with and at times or even build on the fact of descriptive equality.”


271 See, among others from Chapter I, Kleinberg et al. (2019); Barocas and Selbst; Reisman et al.; and Gebru et al.

272 See, for examples of formalizations and their trade-offs from Chapter I, Zemel et al.; Dwork et al. (2011); Hardt et al.; and Pleiss et al.

273 See, among others from Chapter I, Zafar et al.; Bellamy et al.; and Green and Hu.

274 See suggestions by O’Neil and Eubanks, among others, in Chapter I.

275 See, for examples described in Chapter I, the work of Kosinski et al.; Volkova and Bachrach; de Montjoye et al.
applications to credit scoring since the ECOA prohibits certain information as a basis for credit decisions.\textsuperscript{276} From a technical perspective, the prohibition of certain information also reduces a model’s performance, leading researchers to attempt de-biasing techniques at different stages of the machine learning pipeline.\textsuperscript{277} More recently, however, research indicates that de-biasing techniques may not be as effective as indicated: they may be hiding bias, rather than removing it.\textsuperscript{278} In part because there has yet to be a definitive machine learning technique, researchers have focused on tools to increase transparency and model explanations.\textsuperscript{279} These can be interim ways to increase trust in machine learning applications. They may also be useful for legal and policy debates about discrimination, though these techniques do not provide a cause or justification, which case law seems to indicate as most valuable in proving wrongful discrimination.\textsuperscript{280}

Chapter III serves in part as a demonstration of the disconnect between the existing consumer credit context and advances in machine learning. The framework presented in Chapter III (see Table 1) compares traditional and “alternative” credit data to the existing mechanisms implementing the FCRA and the ECOA. Traditional data are compliant with these mechanisms, since the mechanisms were designed around those data. However, traditional data are limited and leave a portion of the population underscored, creating the motivation for “alternative” credit data. “Alternative” credit data is expansive, and could perhaps be all data.\textsuperscript{281} To breakdown “alternative” data slightly, the framework suggests two categories: payment and personal data. “Alternative” payment data are perhaps more compliant with the FCRA and ECOA mechanisms since they are more similar to traditional data, yet there would still need to be some evaluation and testing to confirm that these “alternative” payment data present fewer risks. With regard to “alternative” personal data, however, it is not clear that their use would be compliant with the ECOA and FCRA mechanisms, and might, depending on enforcement for proxies, be prohibited.

“Alternative” credit data, combined with machine learning modeling techniques, could result in increased discrimination, which is what motivated regulatory action in the late 1960s. Regulators are well-aware of these concerns, though exploring options: as described in the 2017 Request for Information, “the use of alternative data and modeling techniques could potentially lead to disparate impact on the part of a well-intentioned lender as well as allow ill-meaning lenders to intentionally discriminate and hide it behind a curtain of programming code.”\textsuperscript{282} Despite their risks in the existing context, “alternative” credit processes have the potential to be more inclusive and expand access to credit. The FCRA and ECOA mechanisms, implemented to achieve fair credit goals, themselves become a source of unfairness when they are not able to

\begin{itemize}
\item \textsuperscript{276} 15 U.S.C § 1681 \textit{et seq}.
\item \textsuperscript{277} See, for examples from Chapter I, Bolukbasi et al.; Zhao et al.; Zafar et al; and Bellamy et al.
\item \textsuperscript{278} See, Gonen and Goldberg.
\item \textsuperscript{279} See, Wachter et al.; Woodruff et al.; Selbst et al.
\item \textsuperscript{280} See, in Chapter II, the discussion of Texas Dep’t of Housing & Cmty. Affairs v. Inclusive Communities Project, Inc., 135 S. Ct. (2015).
\item \textsuperscript{281} See, Hurley and Adebayo.
\item \textsuperscript{282} Consumer Financial Protection Bureau (2017), 19.
\end{itemize}
mitigate risks to consumers, and no longer help attain fair credit goals. This Chapter will explore the idea that technological advances may have evolved fairness mechanisms.

Procedural Fairness

Credit scores are mostly viewed as problematic because they are mysterious, “black box” derivatives of credit reports, so, even while they source information from credit reports, they can reflect inaccuracies, or weigh credit information differently, perhaps even to the extent of making non-adverse information have an adverse effect on a consumer’s score. For example, reporting an account issue several times is not adverse information in itself, but depending on how the credit scoring system is designed, the model may weigh this information negatively. Consumer advocates suggest that more transparency around credit scoring systems would help evaluate whether credit scores are having adverse effects on consumers, both from procedural and substantive perspectives. Companies claim their credit scoring systems as proprietary, however, there is some movement towards more transparency in the industry. For example, the Fair and Isaac Corporation (FICO) has disclosed categories and approximate weights for its models, as follows: payment history (35%); amounts owed (30%); length of credit history (15%); new credit (10%); and types of credit in use (10%).

Transparency in consumer credit scoring may also be relevant for discrimination claims under the ECOA. As with the example above, there may be instances where the data or modeling techniques may have discriminatory impact on certain groups. While companies might argue that certain discriminatory effects in credit scoring systems have a business purpose, insight into the data and modeling techniques would help evaluate whether there exists a less discriminatory alternative to achieving a similar business outcome. In the case of traditional consumer credit scoring and reporting, the data used to determine information are regulated, but the modeling methods are not, and formulas or algorithms can remain proprietary. Despite this focus on data, there are still uncertainties around how they can adversely affect consumers, and there is interest in more transparency around consumer credit scoring systems in order to address questions of compliance within the FCRA and the ECOA, and their procedural and substantive fairness goals. To help manage these issues in the context of “alternative” consumer credit, there are two ways to modernize FCRA mechanisms: first, policymakers could expand FCRA’s definition of a consumer report to hold more companies accountable in the “alternative” credit

\[\text{Wu, 275: “[...] there is certain information that by itself does not seem adverse, but can lower a credit score (e.g., reporting an account several times). Information about how the credit score model works may be necessary to prove that the information is actually adverse.”}\]

\[\text{Wu, 722. See also, Robinson + Yu, 9.}\]

\[\text{Wu, 725: “They would also have to rebut the inevitable argument that the factors in the scoring model are required because of business necessity. In order to rebut the argument, the plaintiffs might try to show that there is a less discriminatory alternative. A disparate impact claim might be based on the argument that scoring models contain forbidden factors such as race or that correlate strongly with race.”}\]
context; and, second, policymakers could encourage more transparency around and explanation of credit models, given that such “alternative” models will have many more factors incorporated.

Expand the Definition of Consumer Report

“Alternative” credit scores can resolve issues raised with traditional credit scores, and increase the pool of individuals with access to credit. One of the issues, however, as discussed in Chapter III, is that the data could be inaccurate, or processed in a way that could lead to unfair results. As background, the FCRA was not initially concerned with data sharing, but following a data boom in the 1990s, definitions were expanded, adding a new class of entities, data furnishers,286 and imposing more requirements with regard to maintaining accuracy of information and resolving consumer disputes. This was significant because maintaining accuracy and integrity of information had become a significant challenge, and because the expanded definitions gave regulators more oversight. The liability structure in the FCRA is such that credit reporting agencies and data furnishers are held liable for inaccurate information if reasonable procedures were not taken to ensure “maximum possible accuracy.”287

Ushering in the era of “alternative” data gives reason to further expand the definition of a consumer report.288 Consumer reporting information needs to relate to credit or other FCRA purposes.289 This definition may too restrictive for the expanse of “alternative” data. As with the addition of data furnishers in the 1990s, broadening the definition of a consumer report would impose accountability on companies that collect and process consumer data. This would give the CFPB clear authority to enforce data protection violations that may currently be out of scope.290 Since the FCRA does not regulate the transparency of credit scoring models, and it currently does not expand transparency requirements beyond traditional credit processes, companies using “alternative” processes may not be regulated under the FCRA. This does not mean, however, that companies will not engage in similar activities.291 For example, in 2015, Facebook submitted patents for creditworthiness algorithms.292 Companies like Facebook rely on data as

286 See supra note 145 for the definition of data furnisher.
287 This standard came out of cases in the 1980s, such as Pinner v. Schmidt 805 F.2d 1258 (5th Cir. 1986) and Koropoulos v. Credit Bureau Inc. 734 F.2d 372 (D.C. Cir. 1984). Cited in, Wu, 128. Additionally, a report may be considered inaccurate, if it is misleading (despite being technically accurate), though this is subject to Court decision: “Numerous district and state courts have held that a consumer report can be inaccurate, even if the report is technically true in some narrow sense, but the report is overly general, incomplete, out of date, or misleading.”288 Since the scope of the FCRA is dependent on the definition of a consumer report, this expansion would require consideration of new data furnishers. See supra note 146.
289 Wu, 77: “If the information is collected and used exclusively for marketing purposes unrelated to credit, employment, insurance, legitimate business need, or other FCRA purposes, it is probably not a consumer report and is not covered by the FCRA.”
291 Yu et al.
part of their business, though they do not fall squarely under FCRA enforcement. Solove and Schwartz present a view that any effort to establish a consumer's worthiness, whether for credit or otherwise, should be considered a consumer report or pre-screened list, and fall under the FCRA. Going further, expanding the definition of consumer report would also expand the definition of data furnishers and include any company that shares consumer data with third parties. This would be a broad consumer data protection choice, but increasingly necessary given the way “alternative” processes are advancing. Information that may seem unrelated to credit scoring, such as whether an applicant fills out a form in capital letters, or their social network activity, is more valuable to “alternative” credit scoring than one could expect. An argument against expanding the FCRA could be an increased burden or compliance cost for companies, however, the reforms in the 1990s indicated that such changes do not have significant impact on business costs.

Encourage Model Explanations

As described in Chapter I, another source of unfairness can be in the models themselves. For example, ZestFinance discovered that their models assessed lower creditworthiness in part for credit applicants who filled out forms in capital letters. While not commonly used by creditors today, the data-driven consumer credit context is ripe for machine learning techniques. Such models may allow for more individualized outcomes, but they may also be assessing individuals based on similarity characteristics, which, as described in Chapter I, could be less than ideal for fairness, depending on how it is defined. While the ECOA in theory prevents the use of certain data for determination of credit scores, these data can be inferred or predicted from the combination of other data types and the type of machine learning technique applied. This begs the question as to how does one determine if a machine learning application is inferring or

293 Solove and Schwartz, 60: “For Mierzwinski and Chester, the law should view these ‘online scoring databases’ as ‘equivalent to pre-screened lists, which are consumer reports.’ These authors call on the FTC to determine whether companies that sell this information should fall under FCRA restrictions due to their ‘establishing the consumer’s eligibility for personal, family, or household purposes; employment purposes; or any other kind of purpose authorized under section 1681b…”

294 Byrnes: “In other words, algorithms fed by so much diverse data will be less prone to discrimination than traditional human-driven lending based on a much more limited number of factors. Among the insights discovered by ZestFinance’s algorithms: that income is not as good a predictor of creditworthiness as the combination of income, spending, and the cost of living in any given city. The algorithm also takes note of people who fill out a loan application in all capital letters, whom their model has found to be worse credit prospects than those who fill it out in all lowercase letters.”

295 Robinson + Yu, 15.


297 See supra note 294.

298 Barocas and Selbst. See also, Byrnes: “Could data with unrecognized biases, when fed into such systems, inadvertently turn them into discriminators? Fairness is one of the most important goals, says Merrill. To guard against discrimination, his company has built machine-learning tools to test its own results. But for consumers, unable to unpack the complexities of these secret and multifaceted programs, it will be hard to know whether they have been treated fairly.”
predicting prohibited data? Research suggests that model transparency and explanations may be one way to assist in such determinations.

Depending on their design, machine learning models can be more or less opaque, and difficult to analyze or draw human understandable conclusions from. They are perceived as proprietary “black boxes” that block causal analysis and cannot be investigated. As this perception perpetuates, trust in and accountability of these models diminishes. Building explainable and interpretable models is one way to attempt to ensure trust in the consumer credit system. As the related research indicates, explainability and interpretability are areas of research that are still under development, and in designing such systems for the consumer credit context, researchers and policymakers would need to coordinate so as to determine the requirements for the explanation narrative. Some example include:

- Are there some areas, such as the FCRA and ECOA compliance mechanisms, that need more explanation than others?
- Who is the target audience? Do explanations need to be modified for different uses?
- Should an explanation be normative or descriptive?
- How feasible is an explanation?
- What makes an explanation sufficient?
- How feasible is an explanation, both technically and financially? Would that information create prioritizations among explanations?
- How should explanation errors be addressed?

Not only does the increasingly complexity of credit scoring systems undermine consumers’ ability in the FCRA to dispute inaccurate information, there exist a variety of different consumer scores, some of which are not necessarily disclosed to consumers. Unlike the requirement for consumer credit reports, consumer reporting agencies are not required to give consumers free annual access to their credit score. Two are of the primary credit scores, VantageScore and FICO, are “meaningfully different.” If this is the case for traditional credit scoring, where scores are based on a more limited set of input data, then it will be less attainable to explain variance of “alternative” credit scores without some form of model explanation process. Given the FCRA requirement for consumers to dispute inaccuracies and the ECOA requirement for adverse information to be explained, it is unlikely that creditors would deploy models that were too complex to explain. “Alternative” model deployment could represent too much of a reputational and legal risk for incumbent lenders, though might be more attractive to new

---

299 Doshi-Velez and Kim. See also, O’Neil, 31: “There are three elements of a WMD: Opacity, Scale, and Damage.”
300 This question is based on a Google white paper. See, Google. 2019. “Perspectives on Issues in AI Governance.” https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf. See also, Doshi-Velez and Kim: “The need for interpretability comes from incompleteness, but not all systems require interpretability. Ethics is an example of where the problem formulation itself is not clear, so the interpretability may be more needed, but also more difficult to formulate what is needed.”
301 Wu, 718. Consumer reporting agencies could be developing and selling other types of specialized risk scores, but they are only required to disclose credit scores as part of regulatory requirements.
302 VantageScore is a joint venture among the three nationwide consumer reporting agencies.
303 Wu, 717 citing findings from 2012 CFPB report “Analysis of Differences Between Consumer- and Creditor-Purchased Credit Scores.”
entrants. Model explanations and interpretations could be useful to ensure that these new entrants are compliant with the existing structure.

**Substantive Fairness**

When proposing reforms to the consumer credit system in the late 1960s, policymakers exercised caution in data unfairness. Chapter III is a current reflection of the same debate about the inclusion of personal data in credit models. At the time, policymakers prescribed equality in the face of difference by prohibiting certain information from being used as the basis for credit decisions. In this choice was a recognition that there were differences between, for example, men and women, but that their sex should not be a factor in decisions.\(^{304}\) As described in the Introduction, prescribing equality through “blind” processes was a popular policy in other contexts that recognized the need to curb discrimination in decisions, such as orchestra auditions starting with the Boston Symphony Orchestra in the 1950s.\(^{305}\) The orchestra auditions turned out to be a successful case where “blindness” equated to fairness and has largely remained the same today.

In the consumer credit context, however, there are many more proxies and technological advances that allow for protected information to be inferred or predicted. As such, the “blindness” mechanism is no longer an effective policy in service of fair credit goals. A change in mechanism might be more effective to stand the test of time. Prohibited bases, if used, could improve credit scoring model design, testing, and auditing, and may help better determine discrimination in cases where there is a lack of legal clarity. There are some changes possible to substantive fairness mechanisms that modernize them while still maintaining the goals established in the ECOA.

According to the ECOA, creditors are prohibited from using any of protected class information in the credit determinations. In requiring creditors to be “blind” to the characteristics that were most vulnerable to discrimination, the ECOA represented a simple way to address fairness, and might have provided an elegant solution to discrimination brought to light by the Civil Rights Movement in the late 1960s. The logic was as follows: if the information was not collected, then it could not be used, and then it was not possible to discriminate on the basis of that information. The fallacy of equating “blindness” with fairness is that its seeming simplicity and elegance might actually, given today’s technology, be promoting ignorance instead of fairness, perhaps giving the impression that discrimination is rare in the credit system when it has in fact been hidden and left unsolved.\(^ {306}\)

---


\(^{305}\) Rice (2013).

\(^{306}\) See, Gonen and Goldberg.
The choice to equate “blindness” with fairness is now under pressure, and could be considered suboptimal, since there are two lines of reasoning for which more transparency and use of protected class information would be more fair. First, from an auditing perspective, regulators could require that tools be tested against protected class information in order to verify that these models are not inadvertently inferring protected class information. Second, from a policy perspective, regulators permit the use of protected class information to curb systemic discrimination. It might serve broader substantive fairness and financial inclusion goals to be able to use protected class information. Does maintaining the “blindness” policy represent an ignorant fallacy impeding on a fairer future?

Require Self-Testing Using Prohibited Information

To maintain prescriptive equality, one idea would be to focus more on the outcome of credit decisions, instead of maintaining prescriptive equality of the data themselves. While the current legal and regulatory framework, as noted in Chapter III, focuses on data, self-testing in an exception that focuses on both data and models. Robinson + Yu note, “fair lending law traditionally focuses on human bias [...], but the theory could be a powerful tool for investigating computer systems whose effect may reinforce existing bias.” Like some of the remedies to unfairness described in Chapter I, self-testing of models is a form of post-hoc de-biasing. One could argue that creditors, however, are not necessarily de-biasing models, since prohibited information is not generally collected and included. This argument does not account for proxies that correlate highly with prohibited information.

Self-testing is encouraged specifically to limit model unfairness by verifying that credit scoring models are not the cause, or having an effect of, discrimination on the basis of protected attributes. Self-testing is an elegant mechanism by which creditors can self-regulate and comply with existing fairness policies. Self-regulation assumes trust by allowing creditors to correct themselves. However, self-testing is currently voluntary and approximative, making it less efficient and trustworthy than it could be. Since creditors are granted privilege as an incentive for self-testing, the only guarantee that these self-tests are occurring is when the CFPB conducts its Baseline Review or an Examination, as outlined in Chapter III.

Currently, a credit decision cannot be made on “the basis of” any protected data. Even though the prohibition exists only at the decision stage, it is unlikely that creditors collect prohibited data so as to absolve themselves of any liability and to minimize their risk. While creditors have an exception for self-testing under Regulation B, they choose to be cautious and risk-averse: if the information is not collected, then it can not be used, and then it is not possible to

---

307 Robinson + Yu, 30.
308 See Chapter II for a list of exceptions to the ECOA prohibitions.
309 Official Interpretations of Reg. B, 12 C.F.R. pt. 1002.15: “To qualify for the privilege, appropriate corrective action is required when the results of a self-test show that it is more likely than not that there has been a violation of the ECOA or this part. A self-test is also privileged when it identifies no violations.” Cited in, Brattle, 349.
discriminate on the basis of that information. When engaging in the self-testing process, creditors resort to approximations, such as zip codes appended to census data for race, instead of the actual data. The number of proxies will increase with “alternative” data and models as will correlations to protected attributes. “Alternative” data and models facilitate inference or prediction of protected information. Testing with approximations for the actual data is suboptimal when the actual data can be collected.

Model self-testing is a way to verify that credit scoring systems are EDDSS and therefore authorized for use. For data to be included in a credit scoring system, there must be a “demonstrable relationship” between that data point and creditworthiness. It would seem, based on this rule, that there would be limited use of proxies for prohibited bases, assuming that there is not a “demonstrable relationship” between prohibited bases and creditworthiness. However, there is no evidence that these relationships have been tested, or that they were tested prior to instituting the “blindness” mechanism in the ECOA in the late 1960s. Zip codes, discussed earlier as an example of a proxy for race, highlight the discrepancy in evidence: regulators were concerned that zip codes had some relationship to creditworthiness, but declined to prohibit their use since there was no indication that zip codes were used in credit scoring models. Improved model testing will provide more information about the relationship between certain data and creditworthiness.

There are three ways in which to improve model-self testing, which can be implemented as gradient parts or, for full improvement, a whole:

1. **Require self-testing.** Instead of the current voluntary nature of self-testing, regulators could require that self-testing occur as often as commercially reasonable. While creditors claim to test their models often, as they should to remain competitive since technical debt can accumulate, there is still no guarantee to consumers or regulators that such testing is occurring when needed from a consumer protection perspective. There is little information about how often and how credit scoring systems are tested. A commercial reasonableness clause would allow for regular testing, at the discretion of the creditor, and is a mechanism that would keep pace with technological advancement.

2. **Test with prohibited information.** Models should be tested using prohibited information, since approximating information is less efficient than using prohibited

---

310 See, VantageScore’s “Testing Credit Scoring Models for Statistical Bias.”
311 See supra note 180.
313 Robinson + Yu, 17; “The FTC was concerned that zip codes might bear on creditworthiness, but declined to rule them as below the line because there was no evidence to establish that zip codes were actually used, or were expected to be used, as a credit eligibility factor in scoring.”
information directly. In 2009, the Government Accountability Office suggested that the lack of prohibited data impeded regulatory efforts. In fact, as noted earlier, the use of prohibited information is excepted for model self-testing, though it is unlikely that creditors use this information out of reputational and legal caution. This claim is also difficult to evaluate since creditors are granted privilege for model self-testing, and can lose such privilege in disclosing privileged information, such as test results.

3. **Increase transparency.** There should be some disclosure requirement accompanying the self-tests, either to the regulator only, or publicly. Knowing that models are being tested regularly would help ensure trust in the consumer credit system. Doing so does not have to violate the privilege provisions currently in place. Instead, creditors could disclose information that is not privileged, such as testing method and data sources (e.g., pre-application or compiled, etc.). This information is not privileged since it is used to determine whether the prerequisites for privilege have been met. Standardization for such disclosures could perhaps be modeled off of Gebru et al.’s “datasheets” or Mitchell et al.’s “model cards.”

The idea that models should be tested using prohibited information may be the most controversial, so it requires further elaboration. The FRB proposed and studied this idea over two decades ago, as described in Chapter II. At the time, the FRB concluded that retaining the prohibition was most effective against discrimination for two reasons. First, the FRB was concerned that voluntarily supplying information would create discrepancies and result in data quality issues. They did not, however, consider that mandatory collection might be abat those concerns. There was also precedent for the mandatory collection of race data in the Home Mortgage Disclosure Act. Second, the FRB was concerned that creditors, in having

---

315 Chen et al. show that it is less meaningful to conduct compliance assessments without the attributes of interest. See, Chen, Jiahao, Nathan Kallus, Xiaojie Mao, Geoffrey Svacha, and Madeleine Udell. 2019. “Fairness Under Unawareness: Assessing Disparity When Protected Class Is Unobserved.” [https://doi.org/10.1145/3287560.3287594](https://doi.org/10.1145/3287560.3287594). 316 Taylor, 249: “Pursuant to a congressional request for an overview of federal oversight and enforcement of fair lending laws, the GAO noted in a July 2009 report that Regulation B’s data collection ban “impedes federal oversight efforts” by prohibiting lenders from collecting personal characteristic data, such as applicants’ race, ethnicity and sex for nonmortgage loans.” 317 Privilege is granted when a self-test concludes that there are no ECOA violations, and when there are ECOA violations but the creditor has determined a plan of action to correct the error. The regulator may determine that the creditor’s plan is not sufficient, and if so, may investigate further, in which case privilege may no longer be upheld. In Reg. B, 12 C.F.R. § 1002.2.15 paragraph 15(d)(2)(ii), “The privilege is lost if the creditor discloses privileged information, such as the results of a self-test. The privilege is not lost if the creditor merely reveals or refers to the existence of the self-test.” 318 Public disclosure may go against trade secret protections. 319 See, Gebru et al. 320 See, Mitchell et al. 321 Taylor, 222. 322 Taylor, 224. There may be some First Amendment concerns with mandatory collection of prohibited information, which we do not elaborate on in this document. 323 Taylor, 216: “In 1989, Congress amended the [Home Mortgage Disclosure Act (HMDA)] and imposed a similar race data collection requirement on certain mortgage lenders. This requirement applies to home improvement loans in addition to home purchase loans. Also, in contrast to the ECOA and Regulation B, lenders are required to report the collected HMDA data to federal regulators and the general public.”
access to prohibited information, would use the information in a discriminatory manner. Today, given the proliferation of proxies, creditors arguably have this information already. Instead of deliberately using prohibited information in a discriminatory manner, which would likely fall under the disparate treatment doctrine, creditors might inadvertently and unintentionally be inferring this information. Allowing them to test such relationships directly would be more valuable. Engaging in such testing would also limit liability once a model is deployed.

Using this information does not necessarily solve the problem of what is considered *highly correlated*, as discussed in Chapter II, but it gives creditors and regulators more information to make that determination. One of the effects of added information and transparency in the use of prohibited information is that it could advance the disparate impact test since the verification of outcomes against protected data would indicate clearly whether a violation has occurred. It’s possible that this type of testing would create a third-prong of the disparate impact test, evaluating whether there exists a less discriminatory alternative. This could be valuable in the world of “alternative” data since there would be many different types of data to choose from and correlations would differ. Unless the regulator establishes a correlation threshold for proxies, companies would likely have to justify their choice of certain data over others.

Regulators should require that tools be tested against protected class information in order to verify that these models are not inadvertently inferring protected class information. A practical consideration is that self-testing of models using protected information is done to examine models but is not a credit decision, and thus an exception in the ECOA prohibitions, whereas the use of protected information in credit decisions is currently prohibited in the ECOA. The advantage of this approach is that it requires only a change in the implementing Regulation B, instead of the ECOA itself, though, based on the FRB’s past engagement on the issue, it may be prudent to engage Congress in such a decision. The following section outlines a more complex proposal to permit use of currently-prohibited information.

**Permit the Use of Currently-Prohibited Information**

Chapter I presented the view that using prohibited information in models could allow for more comprehensive model design. It would provide more context for the technology. Unlike the previous proposal to use prohibited information in model testing only, this proposal suggests incorporating information into models, where it could have an impact on credit decisions. The ECOA prohibits credit decisions “on the basis” of prohibited information. As discussed in Chapter II, “on the basis of” indicates cause, but there is no clear threshold for cause defined, which is similar to the problem with proxies, in which there is no defined threshold for “highly

---

325 Throughout the credit process, though especially in the application stage, the ECOA prohibits creditors from requesting prohibited information about applicants. It is possible that the creditor may receive information from consumer reporting agencies, for example, but the creditor is prohibited from using that information in as the basis for a credit decision. In terms of retention, creditors may retain prohibited information so long as it was received, not requested, and they may not use the information in credit decisions, except in certain circumstances.
correlated” information. Even if prohibited information were included in models, this information does not have to be the cause for a decision. Unlike traditional models which may have been designed only with a few factors, “alternative” models incorporate thousands of data points.\textsuperscript{326} Using one or all of the prohibited data points would not necessarily indicate cause in an “alternative” model. In fact, using these data may allow for creditors to be more fair and comprehensive in their decisions. Following a study of the issue of disparate impact in consumer credit scoring, the FRB concluded, in 2007, that while there were differences between groups (“the mean score of African Americans was half that of white non-Hispanics”\textsuperscript{327}), credit scores generally benefited consumers overall, even minorities.\textsuperscript{328} If creditors were permitted to use prohibited information, such as race, in their models, would there still be differences between groups, or could such differences be weighed or calibrated more fairly?

Certainly, permitting the use of prohibited information raises a number of legal concerns, which are not fully developed in this Chapter. These legal concerns increase the complexity of the proposal to permit use of prohibited information, but make it no less important. First, as experienced by the FRB in its attempts to change Regulation B, a change of an over fifty-year old “blindness” policy is significant, and might require Congressional action. Such a change, depending on how it would be enacted by a regulatory agency, could be challenged under the Constitution’s equal protection clause. Unlike the proposals in Chapter I, there is currently no agreed definition of fairness in the consumer credit context that can be easily applied or modified.\textsuperscript{329} Defining fairness mechanically would perhaps be unwise, given how the “blindness” mechanism is being critiqued in Chapter I and in throughout this thesis for its inability to stand the test of time and technological advances.

Perhaps less significant of a change would be for regulators to permit the use of protected classes, without further direction, leaving the private sector to engage should they choose. One way creditors could use this permission is to train credit scoring models to prevent bias, and then remove the information in deployment for credit decisions.\textsuperscript{330} Another option might be for creditors to equalize outcomes, which would likely require treating different people and different groups differently, possibly creating a tension between disparate impact and disparate treatment.\textsuperscript{331} In weighing outcomes, for example, companies would engage in disparate

\begin{itemize}
\item \textsuperscript{326} See, Hurley and Adebayo.
\item \textsuperscript{327} Wu, 735.
\item \textsuperscript{328} Wu, 736. See also, Wu, 730: “However, consumers living in minority and lower-income neighborhoods experienced errors or omissions in credit data more frequently.”
\item \textsuperscript{329} The ECOA’s substantive fairness mechanisms can be interpreted as demographic parity or equal opportunity, for example, depending on how they are interpreted.
\item \textsuperscript{331} Barocas and Selbst: “Ricci was the first indication at the Supreme Court that disparate impact doctrine could be in conflict with disparate treatment. The Court had previously ruled in essence that the anti-subordination principle could not motivate a constitutional decision, but it had not suggested that law effectuating that principle could itself be discriminatory against the dominant groups. That has now changed.”
\end{itemize}
treatment. However, perhaps, as with affirmative action, companies could justify their approach by claiming that it justifies a legitimate business purpose (e.g., financial inclusion). It is possible to interpret Upstart's fairness corrections as such, however it is difficult to do fairness calibration in models without the protected attributes, as described in Chapter I.

Going further, regulators could permit use of protected class information to curb systemic discrimination. This does not mean that creditors would all engage in such practices, but it could create the opportunity for targeted credit expansion programs, whether by private creditors, or in the context of public-private partnerships. There is perhaps some precedent for this idea in the corrective action provisions for self-testing in Regulation B. Following the results of a self-test that indicates discrimination, creditors are required to take corrective action, which may include the following.

- Extending credit to individuals who were improperly assessed (and compensating them for costs and damages due to that improper assessment);
- Correcting policies that led to the error;
- Developing outreach programs, marketing strategies, or loan products to more effectively serve groups that were likely discriminated against; and
- Improving audit and oversight to prevent recurrence.

Extending credit, remedies, and increasing outreach programs is one way to curb systemic discrimination, but it unfortunately occurs after harm has occurred. It may be possible to engage in curb discrimination prior to the occurrence of harm by developing more comprehensive and fair models. In its request for a no-action letter, Upstart suggested such action would be part of its business model: “Upstart may affirmatively solicit or encourage members of traditionally disadvantaged groups to apply for credit, especially groups that might not normally seek credit for Upstart.”

Unlike affirmative action in other contexts, it is not clear that fairer credit would create significant trade-offs in opportunities. Spaces at a private university might be limited, but every adult in the United States is scored for creditworthiness. The way credit has been democratized and expanded, it is theoretically unlimited, so fairer evaluations, meaning evaluations that are more comprehensive and take into account prohibited information, may be the best way to expand access to credit and promote financial inclusion.

---

332 There are ongoing debates about the legality of affirmative action, and a current undecided Supreme Court case on topic which could shape the view of affirmative action moving forward.
333 Upstart, 8.
334 Reg. B, 12 C.F.R. § 1002.15 paragraph 15(c)(2). See also, Brattle, 349, citing Official Interpretations of Reg. B, 12 C.F.R. pt. 1002.15: “A creditor’s determination about the type of corrective action needed, or a finding that no corrective action is required, is not conclusive in determining whether the requirements of this paragraph have been satisfied. If a creditor’s claim of privilege is challenged, an assessment of the need for corrective action or the type of corrective action that is appropriate must be based on a review of the self-testing results, which may require an in-camera inspection of the privileged documents.”
335 Upstart, 14.
336 Some argue that the quality of debt products suffered in favor of the quantity needed to democratize access. See, Olegario, 12: “In the latest phase, when credit was plentiful, reformers continued to tackle the challenge of providing emergency credit to the poor. How much credit should be available, and to whom, were among the questions that recurred in every phase.” See also, Di Maggio and Yao.
This Chapter suggests another option, to change the existing framework. These changes would not affect the fair credit goals, but rather their mechanisms which may not provide enough mitigation of risk procedurally or substantively, or could be preventing the expansion of credit in their limitations. The choice to have prohibitions on classification of individuals by certain categories represents an evolution in thinking on fairness, with formalization in law half a century ago. This Chapter does not discard this vision of fairness, but rather the mechanisms by which it is implemented. Prohibited bases are so protected that their existence promotes a form of “blind,” yet ignorant and ineffective, justice, especially in the wake of “alternative” credit processes.

Using protected information in self-testing is currently allowed, but not required. Self-testing of credit scoring systems using protected class information as correlatory benchmarks should be a requirement. Perhaps more radically, permitting the use of such information could also allow for more comprehensive model design, allowing policymakers to more directly assess discrimination and unfair credit access issues should they choose to allow, or perhaps encourage, some form of outcome calibration or affirmative action. The former suggestion to explicitly use protected information in decision-assistive models would have its own set of risks and may therefore be less feasible, but no less important, than the first recommendation to bolster self-testing. Given historical tensions, the use of sensitive information should always come with increased attention and study, especially when it is changing a fifty-year-old policy choice. Technological advances should allow us to better reveal disparate impact, which was more difficult to prove before. Does considering prohibited information in credit decisions change or evolve our view of fairness, or does the ability to measure and document simply provide more clarity to the same vision?

337 See, Heen.
Conclusion

“Big data is changing the world. But it is not changing Americans belief in the value of protecting personal privacy, of ensuring fairness, or preventing discrimination.”

This year, the Massachusetts Institute of Technology (MIT) launched a new College of Computing, which is its largest structural change since 1950. In a press release, the rationale for the new college highlighted the combined transformational effects of computing techniques and data: “[... ] many have spoken of how their fields are being transformed by modern computational methods — specifically, by access to large data sets and the tools to learn from them. Some of the most exciting new work in fields like political science, economics, linguistics, anthropology, and urban studies — as well as in various disciplines in science and engineering — is being made possible when advanced computational capabilities are brought to these fields.” MIT’s announcement is the newest and most local transformations I have witnessed, but it is not the first. It is a recognition of technology’s “disruption” of traditional systems, much like the effect of machine learning on the consumer credit system. The discourse on technology’s impact has been ongoing and will likely continue for much longer.

In the consumer credit context, the United States Congress coalesced around a vision for a fairness and passed the FCRA and the ECOA to solidify this vision almost half a century ago. These foundational laws were accompanied by implementing mechanisms, some of which have not kept pace with technology. New data and machine learning modeling techniques to determine consumer creditworthiness come with new risks not mitigated by existing mechanisms. Left unaddressed, these “alternative” processes will impede on a longstanding vision of fairness and diminish trust in the consumer credit system. Chapters I, II, and III serve as background to the case Chapter IV makes for modernized fair credit mechanisms. Chapter I presents a technical review of sources of unfairness in technology, and current research on remedies and trade-offs. Chapter II is a historical review of the legal and regulatory context, indicating a perhaps Hegelian view that many of the issues we think of as “new” today were also issues, some unsolved, of the past. Chapter III is an exercise is deploying “alternative” credit processes (using the knowledge from Chapter I) into the existing consumer credit context presented in Chapter II. Chapter III serves to highlight varying levels of compliance risk, depending on “alternative” data type, and that compliance mechanisms are outdated.

“Alternative” credit processes pose challenges to the mechanisms established in the FCRA and the ECOA and their vision for fairness. In particular, “blindness” to certain attributes hinders

consumer fairness more than it helps, since it limits the ability for regulators to determine whether wrongful discrimination has occurred, and for the industry to build better performing models for populations that have been historically underscored. Chapter IV recommends four modernized fairness mechanisms:

- Expand the definition of consumer report;
- Encourage model explanations;
- Require self-testing using prohibited information; and
- Permit the use of prohibited information to allow for more comprehensive models.

Together, these fair credit mechanisms ensure trust in the consumer credit context and usher in a new era of financial inclusion and credit expansion. The choice to have prohibitions on classification of individuals by certain categories represents an evolution in thinking on fairness, which was formalized in law. This thesis does not discard this vision of fairness, but rather the mechanisms by which it is implemented. Prohibited bases are so protected that their existence promotes a form of “blind,” yet ignorant and ineffective, justice, especially in the wake of “alternative” credit processes.

In addition to the choice not to discard present visions of fairness, this thesis makes some assumptions, that run up against other, broader, fairness questions. First, this thesis presumes a consumer credit system and the use of machine learning, though many question their very premise. Second, this thesis assumed that machine learning modeling techniques are being deployed in consumer credit and will become more widespread in the near future, though such predictions may not hold. Third, this thesis did not delve into questioning the fairness of

---

340 See, Heen. See also, Statement of Rep. Annunzio at 121 Cong. Rec. 16, 740 (1975). Cited in, Brattle, 83: “The essential concept of nondiscrimination in the extension of credit is that each individual has a right when he applies for credit, to be evaluated as an individual: to be evaluated on his individual creditworthiness, rather than based on some generalization or stereotypes about people who are similar to him in race, color, national origin, religion, age, sex or marital status. Bias is not creditworthiness. Impression is not creditworthiness. An individual's ability and willingness to repay an extension of credit is creditworthiness.”

341 Questions include: Is the existence of a credit system fair? Is the existence of a scoring system fair? See, as an example, Citron, Danielle Keats, and Frank Pasquale. 2014. “The Scored Society: Due Process for Automated Predictions.” Washington Law Review 89 (1): 33. https://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1318/89wl0001.pdf?sequence=1. The authors question, in part, whether the scoring of individuals is fair, and consider risks of an expanding scoring system (beyond credit). These concerns are also discussed in O'Neil, Eubanks, Olegario (148), Robinson + Yu, and others. See also, Barocas and Selbst: “This Essay is a call for caution in the use of data mining, not its abandonment. While far from a panacea, data mining can and should be part of a panoply of strategies for combating discrimination in the workplace and for promoting fair treatment and equality. Ideally, institutions can find ways to use data mining to generate new knowledge and improve decision making that serves the interests of both decision makers and protected classes. But where data mining is adopted and applied without care, it poses serious risks of reproducing many of the same troubling dynamics that have allowed discrimination to persist in society, even in the absence of conscious prejudice.”

342 Responses to the CFPB's 2017 Request for Information seem to indicate that the use of “alternative” data is more promising than the use of “alternative” models, but this trend may also be reflective of the unmet potential for such technologies. Is machine learning effective in credit scoring compared to more traditional models and methods? Perhaps creditors are not yet deploying machine learning since they are being cautious of incurring too much risk in systems that they cannot yet explain. See, Di Maggio. See also, VantageScore’s “Anything But Conventional” white paper: “While skeptics have claimed that these scores are merely a ‘race to the bottom’ since the data in these consumers reports do not ‘fit’ the traditional methods of scoring, the real bottoming out lies with older methods reliance on simple equations and outdated rules used to determine if a tradeline is scorable. These ‘older school’ calculations are driven by the historical reliance on bankcard type data and ignore other more complex relationships.
making credit more personalized and individualized. Finally, this thesis could further expand with data analysis, qualitative studies, and primary source information from creditors and regulators. It is possible that these actors are aware and engaged in parts of the recommendations already, however, from the public view, it is not clear that this is the case.

This thesis seeks to provide policymakers and technologists with an understanding of the consumer credit process, its vision of fairness, and its regulatory history for managing technological risks so that they may more thoughtfully consider their policy and design choices. Technologists ought to consider how their products, whether through data collection and management tools or through modeling techniques, can be inadvertent sources of unfairness in current frameworks. Mitigating these risks may require either a cautious approach to “alternative” data and modeling, or a wholesale review of regulatory imperatives, which may be outside the realm of control for many technologists to tackle alone. A coordinated effort will secure opportunities for machine learning applications to further an existing vision of fairness.

Modern data and modeling methods have allowed for more complex relationships to be quickly assessed and for various credit behaviors to also be addressed. Is there a limit to the more subtle and overt forms of discrimination that occurs in “perfect personalization”? See, White House Office of Science and Technology Policy (2014), 7. See also, Schauer. The extreme cases of personalized scoring can be a form of surveillance, which is possibly manifested in China today. The Chinese model depends on the idea that big data form a scoring system that “understands you, but does not know who you are” (Long, Chen. 2019. “Digital Inclusive Finance: the future that is happening.” Presentation to MIT 15.483 (Spring 2019)). While some suggest that the Chinese model is similar to American marketing scores, others are concerned that such unparalleled use of data creates a surveillance state and limits individual freedoms. See, Ahmed, Shazeda. 2019. “The Messy Truth About Social Credit” Logic. https://logicmag.io/07-the-messy-truth-about-social-credit/. See also, Reuters. “China to bar people with bad ‘social credit’ from planes, trains.” 2018. https://www.reuters.com/article/us-china-credit/china-to-bar-people-with-bad-social-credit-from-planes-trains-idUSKCN1GS10S. See also, Marr, Bernard. 2019. “Chinese Social Credit Score: Utopian Big Data Bliss Or Black Mirror On Steroids?” Forbes. https://www.forbes.com/sites/bernardmarr/2019/01/21/chinese-social-credit-score-utopian-big-data-bliss-or-black-mirror-on-steroids/#9e6e4db48b8. See also, Vanderklippe, Nathan. 2019. “Chinese courts have put on social-credit punishment list about 13.5 million people deemed untrustworthy.” The Globe and Mail. https://www.theglobeandmail.com/world/article-chinese-courts-have-put-on-social-credit-punishment-list-about-13/

One idea would be to measure perceptions of fairness, or changes in trust, in the context of credit, similar to work of Woodruff et al. In the findings of the FCRA, Congress described the following: “[...] unfair credit reporting methods undermine the public confidence which is essential to the continued functioning of the banking system.” (FCRA § 602 or Pub. L. No. 90-321, tit. VI, § 602.) In this sense, public trust and confidence depend on the public’s perception of fairness in the credit system.
References


Credit Builders Alliance. “CBA reporter.”
https://www.credithuildersalliance.org/reporter.

https://www.brookings.edu/blog/social-mobility-memos/2015/03/31/poverty-isolation-and-opportunity/.


77


Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Deborah Raji, and Timnit Gebru. 2019. “Model Cards for Model Reporting.” [link to article]


Moor, James H. 1985. “What Is Computer Ethics?” UC Davis. [link to article]


Oliver, John. 2016. “Credit Reports: Last Week Tonight with John Oliver” HBO via YouTube. [link to video]


Planet Money. 2017. “#798: Bad Credit Bureau.” National Public Radio. [link to transcript]


Reuters. “China to bar people with bad ‘social credit’ from planes, trains.” 2018. [link to article]

Rice, Curt. 2013. “How blind auditions help orchestras to eliminate gender bias”. The Guardian. [link to article]

Rice, Patricia. 2006. “Linguistic profiling: The sound of your voice may determine if you get that apartment or not” The Source at the Washington University in St. Louis. [link to article]


