Algorithmic Discrimination in Service

Kalinda Ukanwa
Roland T. Rust

March 21, 2020

Kalinda Ukanwa is Assistant Professor in the Marketing Department at Marshall School of Business, University of Southern California. Address: 701 Exposition Blvd, HOH321, University of Southern California, Los Angeles, CA 90089. Phone: 213-740-1421. Email: Kalinda.Ukanwa@marshall.usc.edu.

Roland T. Rust is Distinguished University Professor and David Bruce Smith Chair in Marketing, and Executive Director of the Center for Excellence in Service at the Robert H. Smith School of Business, University of Maryland. Address: 3451 Van Munching Hall, University of Maryland, College Park, MD 20742. Phone: 301-405-4300. Fax: 301-405-0146. Email: rrust@rhsmith.umd.edu.

This research is supported by grants from the Marketing Science Institute and the Robert H. Smith School of Business, University of Maryland. The authors thank the attendees of the Frontiers in Service Conference, faculty and students of the Marketing department at the Smith School of Business, and attendees of the PhD Project MDSA Conference for helpful comments.

Algorithmic Discrimination in Service

Abstract

The goals of this research are 1) to examine how service discrimination can emerge from algorithmic decision-making, 2) to investigate how service discrimination interacts with consumer word-of-mouth to affect demand and profits, and 3) to explore public policy and managerial implications of algorithmic discrimination in service outcomes. By employing a mixed-methods approach, the authors develop a theory demonstrating that discrimination can be profitable in the short-run, but can backfire in the long-run. An agent-based model shows the macro-level implications on demand and profits resulting from micro-level decisions based on biased algorithms. This research demonstrates that although there can be a short-term profit advantage from discrimination in service, non-discriminatory service providers can earn higher long-term profits, on average, than discriminatory service-providers when factoring in consumer word-of-mouth and competition. Large error in measuring consumer quality (value or profitability to the firm) exacerbates service discrimination, while large intra-group variation in consumer quality attenuates discrimination. This research emphasizes the long-term benefits of switching to a service policy that does not use group identity information in algorithmic decision-making, as well as incorporation of word-of-mouth considerations in the firm's objective function. However, for firms that must persist in using group identity information, this research recommends increasing investment in methods of measurement error reduction and increasing exposure to consumers of different populations. By doing so, a firm could reduce service discrimination while improving its long-term profits and societal well-being.

Keywords

discrimination, algorithms, algorithmic bias, agent-based modeling, word of mouth, service

Algorithmic Discrimination in Service

Statement of Intended Contribution

Discrimination is a sad reality that causes many consumers to endure a lower level of service. Although some companies have claimed that they are eliminating discrimination by using AI algorithms, researchers have shown that even algorithms can discriminate, and not just because the programmer's biases are accidentally transferred. We show that such algorithmic discrimination can be attractive to companies, because it can lead to higher profits in the short run. Our research shows that this result reverses if we do a long-term analysis that includes a sufficient level of word-of-mouth. These findings are of immediate public policy importance, given that governments in the US, Europe, and elsewhere are considering regulations regarding the use of algorithms that use group information in service decisions such as lending. The findings are also of profound importance to companies, because use of group information in service alogrithms, although attractive in the short run, can backfire in the long run.

Algorithmic Discrimination in Service

On November 7, 2019, tech entrepreneur David Hannemeier Hansson made a series of posts on Twitter where he accused Apple of discrimination in service against his wife. He complained about Apple's "black box algorithm" that denied his wife a credit line increase for the new Apple Card product. Hansson pointed out that he and his wife shared finances and that his wife had the better credit score. Yet, Hannson's credit limit was 20 times that of his wife. Two days later, Steve Wozniak, Apple's co-founder, replied to Hansson's tweet with his own claim that he received 10 times the credit limit his wife received, despite the fact that they also completely shared finances and assets (Harris 2019). Hannson's tweets went viral, generating substantial word-of-mouth and media coverage. At the time Hansson posted his tweets, he had over 350,000 followers and Wozniak had over 600,000 followers. As of January, 2020, Hansson's Apple Card tweets have been liked over 29,000 times and retweeted over 9,600 times (Twitter). A search by this paper's authors in the Lexis-Nexis database for the term "apple card discrimination" returned 4,970 news articles from around the world (LexisNexis). The tweets also attracted the attention of government regulators. The New York State Department of Financial Services launched an investigation into whether the algorithm used by Goldman Sachs, the financial institution that manages the Apple Card, is discriminatory (Vigdor 2019).

Apple and Goldman Sachs are not singular in facing challenges about discrimination in service. In February, 2018, California's capital city of Sacramento joined Miami, Oakland, Los Angeles, Baltimore, and Philadelphia in suing Wells Fargo & Co. for violating the U.S. Fair Housing Act. They claimed that the bank engaged in discriminatory lending practices, which resulted in excessive foreclosures of minority-owned properties in their cities (Koren 2018). Wells Fargo had already made a $175 million settlement with the U.S. Department of Justice in July 2012 for alleged discrimination against minority borrowers from 2004 to 2009 (Broadwater 2012). In the past two decades alone, other prominent corporations such as American Express, Toyota, and Ally Financial paid more than half a billion dollars in settlements and fines for discrimination in service cases. This amount is likely a lower bound on the total cost because it does not include additional losses due to word-of-mouth, bad publicity, or impaired reputation and brand. Indeed, prior research provides evidence that discrimination in service exists across a wide spectrum of markets (Bertrand and Duflo 2017; Yinger 1998), including rentals (Carpusor and Loges 2006; Ewens, Tomlin, and Wang 2013; Hanson and Hawley 2011), auto sales (Ayres and Siegelman 1995), retail (Leonard, Levine, and Giuliano 2010), healthcare (Williams and Mohammed 2009), sharing economy services (Edelman, Luca, and Svirsky 2017), financial lending (Blanchard, Zhao, and Yinger 2008; Ferguson and Peters 1995), and education (Milkman, Akinola, and Chugh 2012).

We define discrimination in service (a.k.a., service discrimination) as the difference in a firm's service treatment of consumers of equal quality (defined as value or profitability to the firm) who differ only in group membership (e.g., race/ethnicity, gender, education, social class, age, residential location, occupation, etc.). Consistent with the sociological literature, we distinguish discrimination from prejudice. Prejudice, along with stereotyping, bigotry, and racism, focuses on internally-held attitudes, beliefs, and ideologies. These are not the focus of our research. In contrast, discrimination can be independent of internally-held attitudes, and concerns decision outcomes that exhibit disparate impact: unequal treatment of people based on the category to which they belong. Discrimination is not necessarily driven by internally-held attitudes such as prejudice or bigotry (Pager and Shepherd 2008; Quillian 2006). Rational algorithms, in particular, do not typically suffer from bigotry, but that does not ensure that discrimination will not result.

Decades of media reports provide qualitative evidence that consumers experience discrimination in service based on race, ethnicity, and gender (Elliott 2003; Gutierrez 2015; Koren 2016). A substantial body of research provides not only empirical evidence of the prevalence of service discrimination against these U.S. protected groups (Pager and Shepherd 2008; Rodgers 2009), but also demonstrates its impact on group members (Bone, Christensen, and Williams 2014; Crockett, Grier, and Williams 2003). Although service discrimination based on race, ethnicity, and gender receives much attention, the media and literature document service discrimination based on other category designations such as age (Silver-Greenberg 2012), disabilities (Baldwin and Johnson 2006), residential location (Schroeder 2017), social class (Kugelmass 2016), and occupation (Addady 2016).

Some service providers are undoubtedly prejudiced. However, suppose a firm has no prejudiced intent and uses an objective algorithm to make service decisions? In fact, suppose a firm employs decision-making algorithms in an effort to minimize or eliminate the discriminatory tendencies found in human decision-making? How can discrimination in service still emerge and persist in a 21st century world where algorithms make service decisions? Our research examines this phenomenon and specifically investigates the following questions:

1. Can algorithmic discrimination pay in the short run? – we wish to test whether our theoretical formulation is consistent with prior literature in concluding that algorithmic discrimination in service can be profitable. We preview that we confirm the previous findings theoretically. Note that we will extend these findings in several important ways, including the fact that our theory shows conditions under which the traditional findings reverse.

2. Does discrimination pay in the long run? – We set up a realistic, long-run agent-based model simulation to demonstrate that word-of-mouth can drive the reversal of results found in prior literature.

Technological advancements have enabled firms' increased use of algorithms and artificial intelligence for service decision-making. These technological changes in conjunction with societal shifts motivate an increased need for more research on why discrimination in service still persists in the 21st century (Anderson and Ostrom 2015; Bone, Williams, and Christensen 2010; Hill and Stephens 2003) and how algorithmic decision-making can impact the trajectory. Social fissures created by service discrimination have a direct impact on consumer and societal well-being (Bone, Christensen, and Williams 2014; Crockett, Anderson, Bone, Roy, Wang, and Coble 2011).

Algorithmic discrimination in service has several policy implications, especially for anti-discrimination laws. Even if there is no prejudicial intent in an algorithm's design, and even if there is a concerted effort to avoid using data associated with protected classes, such algorithms can still produce a disparate impact– a disproportionate effect on certain groups of people. Consequently, claims of a firm's algorithm creating disparate impact on their consumers pose a legal risk, regardless of the presence or absence of intentional discrimination. However, recent activity among policymakers suggest that there is an urgent need now to shape policy and regulation of algorithms across applications in multiple domains.

For example, in August 2019 the U.S. Department of Housing and Urban Development (HUD) proposed a new rule shaping policy regarding the use of algorithms in housing decisions. The rule weakens the consumer's' ability to make disparate impact claims under the Fair Housing Act of 1968 to fight discriminatory housing practices. Currently, a consumer can hold a business liable, even if it is an unintended consequence, if the service practices produce disparate-impact – a disproportionate effect on certain groups of people. The proposed new HUD rule, however, allows businesses to escape liability if their service decisions are based on third-party algorithms or if their in-house algorithms do not use data that represent or proxy for protected classes (U.S Department of Housing and Urban Development 2019). Critics such as the Brookings Institute, the ACLU, the NAACP Legal Defense Fund, and the National Fair Housing Alliance have raised objections that this rule provides a loophole which enables businesses to make algorithmic-based discrimination in service decisions with impunity (Defend Civil Rights 2020).

The debate about policy regarding algorithmic bias applies to the U.S. financial and insurance sectors as well. In 2018 the U.S Federal government as well as California and New York state governments actively debated regulatory changes regarding firm use of group categories to make service decisions. What is particularly interesting is that these governmental entities moved in opposite policy directions. While the federal government took steps to reduce the Consumer Financial Protection Bureau's (CFPB) power to regulate and enforce restrictions on using consumer race/ethnicity information in auto lending service decisions (Haggerty 2018), California and New York increased the power to prevent insurance companies from using consumer gender (California) or education and occupation information (New York) in their insurance service decisions (CDIpress 2019; Loconte 2018).

Policy implications of biases in algorithmic decision-making are not limited to the U.S. In May, 2018, Article 22 of the European Union General Data Protection Regulation (GDPR) took effect, which prohibits the use of ”automated individual decision-making” algorithms from making decisions about an individual based on special categories, unless the individual gives explicit consent. The special categories, defined in Article 9, include race/ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, sexual orientation, and health status. Furthermore, the regulation provides individuals the right to an explanation of algorithmic decisions made about them, the right to contest the decision, and the right to request human intervention (Vollmer 2018). The impact of this regulation is far reaching, because it prohibits a broad class of algorithms currently used for financial, insurance, real estate, programmatic advertising, and recommendation system decisions around the world, including the United States. Compliance with this regulation required significant overhauls to pre-existing algorithms (Goodman and Flaxman 2017).

There are two primary theoretical streams of research that model firm decision-making and resultant discriminatory outcomes. The first one, the “taste for discrimination” literature, assumes prejudicial intent and asserts that firms who discriminate have included in their objective function a disutility for interacting with members of certain groups. Their objective function is not necessarily profit-maximizing (Becker 1957; Schelling 1969). A real-life example of this type of discrimination is the case of a Colorado bakery, which in 2012 refused to provide a wedding cake to a same-sex couple because of its religion-based service policy (Savage 2017). Our research assumptions instead align with a second stream, the statistical discrimination literature (Aigner and Cain 1977; Arrow 1973; Coate and Loury 1993; Phelps 1972). This literature models discrimination as an outcome of the firm’s problem of incomplete information about its consumers. The firm attempts to resolve this problem by using observable attributes such as group membership to draw inferences about individual consumers. The literature assumes firms are profit-maximizing and do not have an inherent disutility for interacting with certain groups. For example, List's (2004) study of the sports card trading market provides empirical support for this theory. This study finds that dealers who give inferior offers to minority traders do so because they have statistical distributional assumptions about minority traders that differ from their assumptions about non-minorities. Our study, consistent with statistical discrimination theory, assumes that the firm has no prejudicial intent and uses an algorithm to make service decisions. Our study demonstrates conditions under which discrimination in service can still emerge. We assert that our findings present a lower-bound on the emergent effects of algorithmic service discrimination. If the firm's algorithm is indeed driven by a "taste for discrimination" or prejudicial intent (e.g., the algorithm unwittingly reflected the designer's own prejudices,), it would likely intensify and accelerate the effects we present.

Our study contributes to four literatures. First, we contribute to the relatively new literature on biases in algorithmic decision-making. Some notable examples in the literature include Lambrecht and Tucker (2019)'s study of unintentional algorithmic bias in the delivery of STEM job ads–more ads were delivered to men than women. Cowgill (2018) shows conditions where algorithms can reduce hiring biases. Mullainathan and Spiess (2017) provide an overview of the strengths and weaknesses of using machine learning for economic decision-making. Corbett-Davies et al. (2017) and Kleinberg et al. (2018) show that algorithm decision-making can reduce biases and improve pre-criminal trial release decisions when compared to decision-making by judges. Our study differs from these in that it demonstrates conditions where algorithmic biases can have a negative long-term impact on profits and demand for services when consumer word-of-mouth is activated by biased algorithmic service decisions.

Our second contribution is to the literature on differential service treatment of consumers (Haenlein, Kaplan, and Schoder 2006; Haenlein and Kaplan 2010, 2012). Some notable examples include Homburg, Droll, and Totzek (2008), which shows that prioritizing selected groups of customers can be profitable and produce positive effects on customer relationships. Lepthien et al.

(2017) finds that demarketing to deprioritized customers has negative effects on customer relationships via the spread of negative word of mouth. In contrast, our work examines a different part of the consumer journey by investigating the decision of whether or not to provide service to the prospective consumer *before* she or he becomes a customer.

Our research also contributes to the marketing literature on discrimination. There is a prior literature in transformative consumer research (Mick 2006; Pettigrew 2001) and transformative service research (Anderson et al. 2010) that studies service discrimination's impact on the consumer. The impact can be in terms of self-concept, choice, consumption behavior, and decision-making with regards to domains such as financing options (Bone, Christensen, and Williams 2014), and consumer racial profiling (Crockett et al. 2003; Evett et al. 2013; Harris et al. 2005). There is another line of work that examines consumers discriminating against the service providers, including a study of businesses selling "ethnic" French products in English-dominated Canada (Ouellet 2007) and a study of consumer spending at retail outlets with demographically dissimilar employees (Leonard et al. 2010). Our study differs from these in that we investigate service discrimination from the viewpoint of, and impact on, the firm.

Our fourth contribution is to the statistical discrimination literature. We provide a novel theory on how consumer word-of-mouth can arise from service discrimination and can consequently impact long-term demand and profits. To the best of our knowledge, our study may be the first to do so. Furthermore, our model is a dynamic model of statistical discrimination, which are relatively sparse in the literature (Fang and Moro 2011). Extant dynamic discrimination research examine the cost-based, supply-side impact (labor and employment) of discrimination on the profit function (Antonovics 2006; Bjerk 2008; Blume 2005, 2006; Bohren et al. 2017; Craig and Fryer 2017; Fryer 2007). Our model of the demand-side dynamics of discrimination differentiates it from prior dynamic models of statistical discrimination.

Our research employs a mixed-methods approach. Our integrated analytical/agent-based model of service discrimination models the decision-making algorithm a service provider uses to determine whether or not to provide service to a prospective consumer. Our findings apply to service contexts that meet four criteria: 1) consumers can be segmented based on an observable attribute into distinct groups; 2) firms have uncertainty about the quality (e.g., profitability) of the prospective consumer before providing service to the consumer; 3) to form an expectation of the consumer's quality, firms use information about both the individual consumer and the group of which she or he is a member (what we define as Group-Aware firms–in contrast, we define Group-Blind firms as those that do not use group information); 4) firms screen prospective consumers based on expected quality and ultimately select which consumers to serve. For example, our framework could apply to a scenario in which a property manager uses an algorithm to decide whether or not to rent an apartment to a prospective tenant, a manager uses an algorithm to assess whether to allow a prospective member to join an exclusive club, or a bank uses an algorithm to determine whether to offer a loan to an applicant. Some of the many examples of additional services in which providers may decision-making algorithms to screen prospective consumers include law enforcement and judicial services, healthcare, and educational services.

Our research findings suggest that while Group-Aware service algorithms are perhaps more profitable in the short-run, they can backfire over the long-run, due to the effects of word-of-mouth and competition. Our research provides insight into how variance in consumer quality and measurement error can drive the emergence of discrimination, even if the firm's algorithms are not inherently prejudiced. Large measurement error in detecting consumer quality exacerbates service discrimination, while large variance in consumer quality attenuates it. This research suggests remedies that involve reducing error in measuring consumer quality and/or reducing the role that group membership information plays in assessing a consumer's quality (profitability). We elaborate on these themes in the remainder of the paper as follows: first we present our theoretical model and propose a formal definition of discrimination in service. Then, we present an agent-based model (ABM) to go beyond the limitations of the analytical model: the ABM enables investigation into emergent macro-phenomena from the micro implications of the analytical model. With the ABM, we investigate the long-term impact of service discrimination and its interaction with competition and word-of-mouth. Finally, we conclude with a discussion of the policy, managerial, and consumer implications of our findings.

DOES ALGORITHMIC DISCRIMINATION PAY IN THE SHORT RUN?

The analytical portion of our model extends work by Nobel Prize winner Edmund Phelps (1972). We refer the reader to that paper and Aigner and Cain (1977) for details of that model and its derivations. Our model could apply to business contexts such as rentals, insurance, or club memberships, but we choose to ground it in a bank lending context because of the substantial empirical evidence of service discrimination in this domain (Blanchard et al. 2008; Ladd 1998; Ferguson and Peters 1995). However, we intend the bank lending context to be only an illustrative example used to facilitate understanding of the model, as our research is more general and does not focus on the financial lending industry, per se.

We model a bank's decision of whether to offer a loan to each applicant i. We assume that the bank uses an algorithm to determine which applicants receive loans. It is important to emphasize that we do not examine the separate decision of what price to charge for the service. For this reason, we assume that the loan amount and interest rate are exogenous. We acknowledge that a service provider may choose to provide service at varying prices to prospective consumers (such as a bank algorithm offering different interest rates to different loan consumers that vary in risk profiles). In such cases, price discrimination becomes relevant. That is not the scenario we assume here. For those interested in the topic of price discrimination, we refer readers to Bergemann et al. (2015), Narasimhan (1984), and Varian (1989) for excellent insights into this area.

We assume that each applicant i is a member of one and only one of two groups $j \in \{H, L\}$, a high (H) or low (L) quality group. Groups are determined by some attribute that is observable to the algorithm (e.g., education, occupation, ethnicity, gender, location, etc.). Quality (Q_{ij}) is a latent attribute of the consumer that is directly related to achieving the firm's objective function.

Quality, as we define it here, only pertains to the individual's value to the firm, not to the person's inherent worth as an individual. In for-profit contexts such as this, quality can be interpreted as the applicant's true profitability to the bank. At the group level, the mean quality (A_j) of the applicants within a group define whether the group is high or low. However, individual applicant quality can vary within a group. We assume there is inequality between the groups, where

$$A_H > A_L > 0 \quad (1)$$

Examples of differences in quality between groups in some contexts could be upper versus lower social class, men versus women, racial, ethnic, or religious majorities versus minorities, college-educated versus less-educated, east-side of town vs. west side of town, etc. We define the difference between two groups as $Inequality = A_H - A_L$.

We assume that the algorithm uses locally available information (i.e., data on applicants at the focal bank only), is not forward-looking, and does not have knowledge of the bank's competitors' information or beliefs. We believe these are reasonable assumptions based on what we learned from interviews with loan analysts at financial institutions. Their primary sources of information used in loan decisions include the applicant's credit score, credit history, current income and assets. Interviews with loan analysts consistently supported these assumptions in stating that they used their institutional historical data from past applicants and loan recipients to evaluate current applicants in the decision-making process. They did not use information from competing financial institutions, nor did they look at future trends of applicant groups in making loan decisions.

We assume that the algorithm uses a Bayesian learning approach (Ching et al. 2013; Erdem and Keane 1996; Roberts and Urban 1988) to learn from bank applicant data about the quality of individual consumers as well as their groups over time. We assume that the algorithm's objective is to maximize profit by offering loans only to applicants whose quality, Q_{ij}, exceeds a threshold Q^{min}. The threshold Q^{min}, assumed to be exogenous, is the quality of the marginal consumer whose true profitability is 0 to the firm. We assume that Q_{ij} is normally distributed[1] around group j's mean quality, A_j. Since the algorithm cannot observe Q_{ij} before making a service decision, the loan algorithm forms an expectation of Q_{ij} by using available information about the applicant (in the banking context, this can be credit history, net worth, income, debt, employment history, etc.). We assume all applicant information, outside of group membership, is summarized in a single score, S_{ij}, which is a noisy measure of Q_{ij}. The relationships between A_j, Q_{ij}, and S_{ij} are as follows:

(2a) $$Q_{ij} = A_j + \upsilon_{ij}, \quad \upsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{q_j})$$

(2b) $$S_{ij} \mid Q_{ij} = Q_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\varepsilon_j}), \text{ where } \upsilon_{ij} \perp\!\!\!\perp \varepsilon_{ij}$$

(2c) $$S_{ij} \sim \mathcal{N}(A_j, \sigma^2_{q_j} + \sigma^2_{\varepsilon_j})$$

Because S_{ij} has error, the algorithm may supplement the score with information about the group of which applicant i is a member. Although each group's true mean (A_j) and variance ($\sigma^2_{q_j}$) of quality are unknown, we assume that their distributions are known and that the algorithm uses that information to form a prior distribution. The prior could be uninformative, or it could be based on bank historical data, research conducted on the consumer market, or managerial experience.

As a result, the algorithm's expectation of $Q_{ij} \mid S_{ij}$ is a weighted combination of information about the individual applicant (S_{ij}) and the expectation of the mean quality of the group that applicant i belongs to ($\hat{A}_j$):

(3) $$E(Q_{ij} \mid S_{ij}) = \gamma_j S_{ij} + (1-\gamma_j)\hat{A}_j \\ \text{where } \gamma_j = \frac{\hat{\sigma}^2_{q_j}}{\hat{\sigma}^2_{q_j} + \sigma^2_{\varepsilon_j}}$$

The quantity γ_j is known as the reliability of a measurement in classical score theory (Novick 1965). It indicates how much reliance is placed on information about the individual applicant,

[1] We find that our results are robust to other distribution specifications. See Web Appendix B for more detail.

captured in score S_{ij}, as opposed to information about the group, captured in mean quality A_{ij}. The score reliability has important properties that highlight the impact that variation in quality ($\hat{\sigma}^2_{q_j}$) and score measurement error ($\sigma^2_{\varepsilon_j}$) have on the algorithm's expectations of consumer quality. Increasing variation in consumer quality or decreasing variation in score measurement error increases the score reliability (γ_j). As score reliability increases, the algorithm places increasing weight on the consumer's individual information (S_{ij}) and less on group information ($\hat{A}_j$).

The graph in Figure 1 visually displays an example of the model using a range of 450 - 650 for S_{ij} (score) on the x-axis and a matching range on the y-axis for expected quality values, conditional on score: $E(Q_{ij} \mid S_{ij})$. The solid and dashed parallel lines are graphs of Equation (3): the algorithm's expectation of quality of H and L applicants respectively. In this example, the two groups have the same score reliability, $\gamma = \gamma_H = \gamma_L = .5.$, where $\hat{A}_H = 723$, and $\hat{A}_L = 640$.

[Insert Figure 1 about here.]

The gray line at the 45° arc has a slope of $\gamma = 1$. This is where S_{ij} perfectly measures Q_{ij} without error. At this value, the algorithm has no need for group information $\hat{A}_j$ to form an expectation of consumer i's quality. As measurement error is introduced, however, γ_j decreases. As $\gamma_j \to 0$, the regression representing the algorithm's expectation of applicant quality rotates clockwise towards a horizontal line with intercept $\hat{A}_j$. At its limit, $\gamma = 0$ and the consumer's score S_{ij} no longer has weight–the algorithm has a monolithic belief about group j's members: $E(Q_{ij} \mid S_{ij}) = \hat{A}_j$ for all i.

The algorithm's expected profit from a single loan is

(4)
$$E(\pi_{ij} \mid S_{ij}) = \begin{cases} E(Q_{ij} \mid S_{ij}) - Q^{min} & \text{when } E(Q_{ij} \mid S_{ij}) > Q^{min} \\ 0 & \text{otherwise} \end{cases}$$

Consequently, the realized profit from a loan is

(5)
$$\pi_{ij} = Q_{ij} - Q^{min}$$

The realized profit function is driven by the assumption that the true quality (e.g., value to firm or profitability) of consumer *i* is realized after she consumes the service. Note that the realization of profit can be negative. It is possible that after the consumer consumes the service, her revealed true quality falls short of the Q^{min} the firm needs to be profitable.

If the algorithm uses group information as well as the consumer's score to form expectations about each applicant, then the algorithm uses a service policy where it offers a loan to applicants whose score, S_{ij}, exceeds a minimum score criterion for their group. We subsequently refer to this service policy as the *Group-Aware* policy. Alternatively, we define a *Group-Blind* service policy as one where the algorithm ignores group information and uses a single score criterion, regardless of group membership. We derive the Group-Aware minimum score criterion for each group, S_j^{min}, by setting Equation (3) equal to Q^{min} and rearranging terms.

$$S_j^{min} = Q^{min} + (Q^{min} - \hat{A}_j)\left(\frac{1-\gamma_j}{\gamma_j}\right) \tag{6}$$

In Figure 1, the Group-Aware minimum score criteria of the example model are located at the vertical dotted lines labeled "**H** Min. Score (S_H^{min})" and "**L** Min. Score (S_L^{min})". Note that these vertical lines intersect with a horizontal dashed line labeled "Profit Threshold (Q^{min})" at the top right corner of the graph. Each intersection point is precisely where the expected quality of a member of the given group, conditional on score, is equal to the Q^{min} that represents the marginally profitable consumer.

In contrast, the Group-Blind algorithm ignores group membership and aggregates all applicant information to form expectations of quality and a single minimum score criterion, S_{all}^{min} (not shown on the graph). Because we assume that the errors associated with Q_{ij} and $S_{ij} \mid Q_{ij}$ are independent of each other (see Equation (2b)), aggregation of the two groups has no impact on σ_ε^2. However, aggregation does impact the mean and variation in consumer quality. The overall mean ($\hat{A}_{all}$) and variance ($\hat{\sigma}_{q_{all}}^2$) of applicant quality are a combination of the mean and variance of quality of the L-group and H-group members. They are weighted by the proportion of total applicants represented by each group. Consequently, the minimum score criterion of applicants under the Group-Blind policy is as follows:

$$S_{all}^{min} = Q^{min} + \left(Q^{min} - \hat{A}_{all}\right)\left(\frac{1-\gamma_{all}}{\gamma_{all}}\right) \tag{7}$$

Thus far, the algorithm we have described has taken a profit-maximizing, non-prejudiced approach to forming a service policy. So where is the discrimination in service? We formalize our definition of discrimination in service (D_i) as follows:

Definition *Algorithmic discrimination in service occurs when the decision-making service algorithm treats two consumers with the same score and who are equal in value to the firm (i.e., consumers with the same quality) differently just because they are members of different groups. It is equivalently defined as the service algorithm's change in treatment of consumer i if consumer i changes group membership, conditional on maintaining the same quality and score. Discrimination (D_i) is defined as*

$$\begin{aligned} D_i &= E(Q_{i,H} \mid S_i^*, Q_i^*) - E(Q_{i,L} \mid S_i^*, Q_i^*) \\ &= (\gamma_H - \gamma_L)Q_i^* + \left[(1-\gamma_H)\hat{A}_H - (1-\gamma_L)\hat{A}_L\right] \\ &\text{where } S_i^* = S_{i,H} = S_{i,L} \text{ and } Q_i^* = Q_{i,H} = Q_{i,L} \end{aligned} \tag{8}$$

The graph in Figure 1 shows by example the magnitude of discrimination for consumers with a score $S_i^* = 550$. When $\gamma_H \neq \gamma_L$, there is a critical quality level Q_{D0}[2] where consumers of that quality do not experience discrimination. any consumer with quality level $Q_i^* = Q_{D0}$ will experience no discrimination. Their quality level is $Q_i^* = Q_{D0} = \frac{(1-\gamma_L)\hat{A}_L-(1-\gamma_H)\hat{A}_H}{(\gamma_H-\gamma_L)}$. However, other consumers with quality levels higher or lower than Q_{D0} experience discrimination at magnitudes that increase in absolute value the further away quality is from Q_{D0}. If $\gamma_H = \gamma_L = \gamma$, however, then the magnitude of discrimination is constant across all consumers. Under this

[2] $Q_{D0} = \frac{(1-\gamma_L)\hat{A}_L-(1-\gamma_H)\hat{A}_H}{(\gamma_H-\gamma_L)}$

condition, Equation (8) simplifies to $D_i = (1-\gamma)\left(\hat{A}_H - \hat{A}_L\right)$.

While Phelps (1972) implies that each group will have its own service threshold to receive service, we extend the Phelps model in the following findings that provide additional insights (see Web Appendix A for proof):

Proposition 1 *Let S_j^{min} be the minimum service criterion for prospective consumers from group $j \in \{H,L\}$. Typically, the L-group consumers must meet a higher minimum service criterion than H-group consumers to receive the same service at Group-Aware firms. However, when both groups are expected, on average, to be profitable or both groups are expected to be unprofitable, then the ordinal relationship could reverse, where the H-group has the higher service criterion, under the following conditions:*

1. *If both groups, on average, are expected to be profitable, and if consumer heterogeneity of the L-group is sufficiently low compared to the heterogeneity of the H-group, then H-group consumers must meet a higher service criterion than L-group consumers to receive the same service at Group-Aware firms (i.e., $S_L^{min} < S_H^{min}$).*

2. *If both groups, on average, are expected to be unprofitable, and if consumer heterogeneity of the L-group is sufficiently high compared to the heterogeneity of the H-group, then H-group consumers must meet a higher service criterion than L-group consumers to receive the same service at Group-Aware firms.*

Consistent with prior statistical discrimination models, in most cases the service provider has a higher service criterion for L-group consumers than for H-group consumers. However, we bring additional insight to this assertion beyond Phelps (1972) and subsequent literature. We highlight conditions where the reverse is true: the H-group has the higher service criterion ($S_L^{min} < S_H^{min}$). This occurs when both groups, on average, are profitable to the firm and $\hat{\sigma}_{q_L}^2 < \sigma^* = \hat{\sigma}_{q_H}^2 \left(\dfrac{\hat{A}_L - Q^{min}}{\hat{A}_H - Q^{min}}\right)$. It also occurs when both groups, on average, are unprofitable to the firm and $\hat{\sigma}_{q_L}^2 > \sigma^*$.

The intuition behind these results is as follows: in the case where both groups, on average, are profitable (i.e., $Q^{min} < \hat{A}_L < \hat{A}_H$), if heterogeneity in the L-group is sufficiently low relative to the heterogeneity of the H-group, then the firm trades off the greater profitability of the H-group in favor of the lower uncertainty of the L-group (i.e., $\hat{\sigma}^2_{q_L} < \sigma^*$). Analogously, in the case where both groups are unprofitable (i.e., $\hat{A}_L < \hat{A}_H < Q^{min}$), the firm trades off smaller losses of the H-group for greater uncertainty of the L-group (i.e., $\hat{\sigma}^2_{q_L} > \sigma^*$). It is important to note that in the third possible case where the H-group is profitable and the L-group is unprofitable, (i.e., $\hat{A}_L \leq Q^{min} < \hat{A}_H$) , the L-group's minimum service criterion is always higher than the H-group. Illustrative examples of regions of the criterion state space where these relationships hold can be found in Web Appendix C.

Consequently, we find that it can be profitable, in the short-run, to discriminate. The average per period (short-term) profits are greater from a Group-Aware service policy than from a Group-Blind one (see Web Appendix A for proof):

Proposition 2 *On average, it is more profitable for a service provider to have a service criterion for each consumer group (Group-Aware service policy) than to have a single service criterion for all consumers (Group-Blind service policy).*

It is important to note here that this finding is conditional on the assumption that the focal firm is a monopoly. As prior literature has suggested, allowance for competition can reverse this effect (Becker 1957). We further support this assertion under a novel competitive context involving consumer word-of-mouth in our agent-based model, which we further elaborate upon later in this paper.

Taking the derivative of D_i in Equation (8) with respect to σ^2_ε, $\hat{\sigma}^2_q$, and γ yields the following proposition:

Proposition 3 *Let each group $j \in \{H, L\}$ have the same magnitude of consumer heterogeneity (σ^2_q) and measurement error (σ^2_ε). Then the following is true:*

1. *Discrimination (D_i) in service increases as the magnitude of measurement error of*

consumer quality (σ_ε^2) increases.

2. *Discrimination in service decreases as consumer heterogeneity within each group (σ_q^2) increases.*

The intuition behind this proposition is that the greater the inequality between two groups, the greater the difference between the algorithm's expectations of the quality of two consumers who are equal in true quality from each group. However, the greater the reliability of individual consumer information in assessing quality, the less the algorithm will rely on group information to form an expectation. Reliability of individual information improves when there is more information about members within a group (i.e., more intra-group variation in consumer quality) and when there is decreased error in measuring the quality of individuals. Increased reliance on the consumer's score/decreased reliance on group information leads to decreased discrimination.

Table 1 presents a summary of the findings from our analytical model.

[Insert Table 1 about here.]

The analytical model provides insight on how variation in consumer quality and measurement error can have a direct impact on the magnitude of discrimination in service. It also highlights how changes in group quality over time can change not only the variation in the quality of the group, but also the magnitude of discrimination that they experience. However, what the analytical model does not address is the potential response of customers via word-of-mouth to the algorithm-driven service policies. Furthermore, we want to examine this in the context of competition, which provides an outside option to the consumer. This is central to our research question. To investigate this, we next examine the impact of dynamic consumer word-of-mouth (WOM) and its interaction with competition as an outside option for those consumers.

Although the dynamics of competitive effects could be potentially captured in an extension of the analytical model, competition's interaction with the social dynamics of consumers in a network is harder to analyze analytically. To investigate these, we turn to agent-based modeling (ABM), which is well-suited to handle this modeling challenge because of its ability to simulate interacting individual agents who autonomously behave based on their own rule-based system. The ABM will enable us to more easily model the dynamics of consumer WOM in a network structure and their interactions with competing banks with competing service policies.

DOES ALGORITHMIC DISCRIMINATION PAY IN THE LONG RUN?

Agent-based modeling (ABM) is a research tool that enables the researcher to simulate the behavior and interactions of autonomous individual agents (people, organizations, etc.) to analyze emergent macro phenomena. It is often used to understand the dynamics of collective patterns in a complex system (Delre, Broekhuizen, and Bijmolt 2016a; Goldenberg, Libai, and Muller 2001b, 2010; Rand and Rust 2011). By using both ABM and analytical modeling in a complimentary fashion, we leverage the strengths of each (full parameter space exploration for analytical modeling, modeling of complex interactions for ABM) to answer our research questions more fully than by using one or the other alone (Peres and Van den Bulte 2014). ABMs can be used for two different purposes. One purpose is to use an ABM as an extension of an econometric model. In such applications, careful validation of all the input parameters is essential (e.g., see Libai, Muller, and Peres 2013). An alternative use, however, is to use an ABM as an extension of an analytical model in order to show directional results of how variables affect outcomes (Delre, Panico, and Wierenga 2016b). This reflects our purpose. However, we still strive to use realistic, data-justified values where possible. In that spirit, in the next section we discuss our use of ABM to investigate the long-run implications from our findings.

To analyze the dynamics of algorithmic discrimination in service, we employ a 2^8 full factorial design (256 separate simulations) in the agent-based model (ABM). The ABM models supply and demand for loans in a simulated city. The city contains four competing banks and a population of 200 consumers comprised of people from an H-group or L-group. Banks and consumers are randomly distributed throughout the geographic area. Based on the distributional assumptions used in the analytical model, the ABM randomly assigns quality and credit score attributes to consumers. Each bank has one algorithm. Two randomly-determined banks have a

Group-Aware service policy (a minimum score criterion for each population group) while the other two have a Group-Blind service policy (a single minimum score criterion). This allows us to examine competition and its impact on consumer demand and firm profits over time in the ecosystem. Becker (1957) theorized that market forces can ultimately drive out firm discriminatory behavior if non-discriminatory competitors exist. We test the spirit of this theory by including firms in the ABM ecosystem that employ a group-blind minimum score criterion. In each time period in the ABM, a random selection of consumers applies for a loans. These applicants select one and only one bank in any given period based on their utility from the bank (to be elaborated on shortly). Subsequently, each bank algorithm offers loans to applicants with scores exceeding the minimum score criterion determined by bank service policy. Algorithms use historical data of past applicants to update their beliefs about group mean quality levels and to set new minimum score criteria in each period. Each applicant retains a history of loan applications and rejection/acceptance outcomes. Banks cannot observe each applicant's history, but consumers can observe the application history of other consumers in their network.

In the primary ABM, we implement a complete network (i.e., a fully connected network), a network structure where all consumers are connected to all other consumers. This network structure has been used extensively in the marketing literature (Bass 1969; Goldenberg et al. 2010; Rand and Rust 2011). To test robustness of the model with respect to network structure, we also ran ABM simulations based on an Erdös - Rényi random network Erdös and Rényi (1959) and a Barabasi-Albert preferential attachment network Barabási and Albert (1999). For a random network, the probability that two consumers are connected is equally likely across all consumers. For a preferential attachment network, some consumers are disproportionately more likely to be connected than other consumers. Both alternative network structures are widely used in the literature (Rand and Rust 2011; Wilensky and Rand 2015). We find that all three network structures produce qualitatively similar results in our investigation. The remainder of the paper reports results based on the complete network structure. However, the interested reader can find more detail on the robustness analysis of network structure in Web Appendix D.

The ABM uses combinations of high and low values for each of the eight factors. Three of the eight factors come directly from the analytical model: intra-group quality variance, measurement error variance, and degree of inequality ($\hat{\sigma}^2_{qt}$, $\sigma^2_{\varepsilon t}$, and $Inequality_t = \hat{A}_{H,t} - \hat{A}_{L,t}$). We test both static and dynamic inequality conditions over time. In simulations with dynamic inequality, we allow $\hat{A}_L$ to grow at a rate of .16% per period[3] while holding $\hat{A}_H$ fixed.

The remaining five factors are assimilation, population mix, number of applicants, and two aspects of word-of-mouth (WOM). Assimilation can be thought of as adopting observable characteristics or cultural practices associated with the H-group. We expect that greater degrees of assimilation reduce discrimination. Assimilation reduces the chance that an L-group member is identifiable as L-group because the person possesses attributes of both the H-group and L-group. For example, a religious minority who attends a bank loan interview dressed in a business suit (characteristic of the H-group majority) may experience less discrimination than if he attends in traditional religious garb. We operationalize assimilation in the ABM model by varying the proportion of characteristics (in terms of mean quality) that the L-group shares with the H-group (0% vs. 50%).

Varying the population mix of the applicant pool allows us to test whether the frequency of exposure to applicants impacts discrimination in service. An increased balance in population mix – a 50/50 split in two populations represents perfect balance – increases the algorithm's exposure to members of both groups. More exposure provides the algorithm with more information. We operationalize population mix in the ABM by varying the percentage of population that is H-group (9% vs. 63%)[4]. The lower percentage of 9% represents a less balanced population. We predict that the magnitude of discrimination will be lower when the H-group population represents 63% of the population mix. This is because a 63/37 population mix is much closer to a balanced population than a 9/91 split. Discrimination decreases because the algorithm has more information from both groups about consumer quality.

[3]Based on the annual growth rate of average Black wealth relative to Whites from 1967 to 2010 in the U.S. Source– Pew Research Center

[4]Based on the percentage of the population that is White in South Africa and U.S respectively. Source: South African National Census of 2011 and 2011 Pew Research Center Report

Varying the intensity of demand allows us to test how demand for service impacts service discrimination. We operationalize this by varying the percentage of the city population that applies for a loan in each ABM time period (20% vs. 80%). We posit that a greater frequency of applications would lead to less service discrimination. A greater frequency of applications provides banks with more information. More information should improve variation in quality over time and thus decrease discrimination. This scenario reflects potential differences between highly trafficked banks (e.g. city banks) versus less trafficked banks (e.g. rural banks), even after controlling for other factors like population mix.

We investigate how the final factor, consumer word-of-mouth (WOM), affects demand for services over time. Prior literature has established that WOM can have strong influence on consumer choice (Goldenberg, Libai, and Muller 2001a; Libai, Muller, and Peres 2013; Trusov, Bucklin, and Pauwels 2009). Our model assumes that loan applicants are utility-maximizing. Utility for bank b has an inverse relationship with distance ($Dist_{ib}$) between applicant i and bank b, and it increases with i's assessment of her probability of receiving a loan from the bank. The inclusion of distance as a factor in the utility function is consistent with models in the consumer store choice literature (e.g., Huff 1964; Rust and Donthu 1995). We account for additional unobservable factors that influence an applicant's utility with an extreme-value distributed error term, ε_{ibt}.

WOM about banks is an important factor in each consumer's bank selection. Each consumer in the ABM "talks" to other consumers in her network to find out who has received loans and from which banks. We operationalize WOM through each consumer's ability to access the application history of other consumers in their network. WOM utility that applicant i has for applying to bank b is driven by applicant i's assessed probability that she will be offered a loan from bank b at time t ($P_{ibt}^{WOM} = Pr(Loan_{ibt} \mid \alpha, \mathbf{w_i})$). The probability is equal to the proportion of the applicant's social ties that has received loan offers from bank b weighted by the strength of the social connection between i and each social tie k. Consistent with prior research, strong ties have a greater probability of affecting an individual's choice than weak ties (Brown and Reingen

1987). The strength of the social connection is measured as the inverse of the distance (Soc_{ik}) between i and k in the simulated city. WOM is also weighted by whether the source of WOM is an in-group vs. out-group member. For example, if i is a member of the squares group in the ABM, then i considers other squares as in-group sources of WOM and triangles as out-group sources. Extant literature has shown that consumers give consideration to in-group versus out-group sources of WOM (Podoshen 2006; Lam et al. 2009; Uslu et al. 2013).

We vary $\alpha \geq 1$, the weight that consumers place on WOM received from in-group relative to out-group sources, with input values of 1 vs 3 (based on Brown and Reingen (1987); Podoshen (2006); Zhao and Xie (2011) findings). When $\alpha = 1$, applicant i equally weights in-group and out-sources of WOM. An $\alpha > 1$ implies that i places greater weight on WOM from other in-group ties. We also vary β, the weight that consumers place on WOM about bank b relative to the weight placed on the distance to the bank $Dist_{ib}$, with values 2 vs. 20 (based on Trusov, Bucklin, and Pauwels (2009)). The utility that i has for applying to bank b at time t is as follows:

$$
\begin{aligned}
&U_{ibt} = \beta P_{ibt}^{WOM} - Dist_{ib} + \varepsilon_{it} \text{ , where} \\
&P_{ibt}^{WOM} = \frac{\sum_k w_{ik} \ \mathbf{1}(\text{if } b \text{ has ever offered a loan to } k \text{ as of time } t)}{\sum_k w_{ik}} \qquad (9) \\
&w_{ik} = \frac{1 + \alpha \, \mathbf{1}(i, k \in j)}{Soc_{ik}}
\end{aligned}
$$

Each replication of the bank-applicant ecosystem runs for 300 time periods. Developed in the NetLogo programming language (Wilensky 1999), the ABM generated over 15.7 million records of data.

ABM Analysis and Results

We ran simulations that represent our primary ABM with the 2^8 full factorial design. This ABM includes increases in the number of competing banks to two Group-Aware banks versus two Group-Blind banks and additional variables mentioned in the prior section to examine the robustness of our results in the face of realistic factors in a complex system. Consistent with

Proposition 1.1, for example, the Group-Aware banks in the ABM are significantly more likely to offer loans to H-group applicants than their L-group counterparts (Static Mean Quality: -1.616, $p < .001$; Dynamic Mean Quality: $-.801$, $p < .001$). Consistent with Proposition 3, decreases in measurement error decreases service discrimination (Static Mean Quality: -24.118, $p < .001$; Dynamic Mean Quality: -14.720, $p < .001$)[5]. Lower intra-group variance in quality increases the magnitude of discrimination (Static Mean Quality: 13.774, $p < .001$; Dynamic Mean Quality: 16.706, $p < .001$). These results provide added confidence that the ABM is appropriately simulating the micro-results from the analytical model.

Consistent with our prediction, the ABM results suggest that increases in the proportion of the population that is H-group (moving from an imbalanced to a balanced, integrated society) decrease discrimination (Static Mean Quality: -1.982, $p < .001$; Dynamic Mean Quality: -9.214, $p < .001$). Recall that discrimination is measured as a difference in expected quality, conditional on two consumers from two groups having the same quality and score. A greater percentage of the population applying for loans increases discrimination (Static Mean Quality: 19.433, $p < .001$; Dynamic Mean Quality: 13.377, $p < .001$). Increased assimilation also has the significant effect of decreasing discrimination (Static Mean Quality: 46.990, $p < .001$; Dynamic Mean Quality: 33.980, $p < .001$). Recall that the degree of assimilation relates to the proportion of characteristics, and thus mean quality level, that the L-group shares with the H-group. The ABM results support the expectation that the more assimilated the L-group is, the less the group is discriminated against in receiving service.

We find that WOM and competition can drive loss of applicant market share and long-term profits. On average, Group-Blind banks have a significantly greater share of all applicants in the market (Static Mean Quality: 52.4% vs. 47.6%, $p < .001$; Dynamic Mean Quality: 52.8% vs. 47.2%, $p < .001$). WOM also can have a large impact on long-term profits. We regressed cumulative profits on *Group*, α (weight placed on in-group sources of WOM), β (weight placed on WOM in general in the applicant's utility function), and their interactions. We also included

[5]The dependent variable is an exact measure of discrimination based on Equation 8. Data has been 1% trimmed to reduce the effect of extreme outliers of discrimination values.

controls for other ABM simulation factors (*ScoreLowVar*, *QualityLowVar*, assimilation, population mix, number of applicants). Consistent with findings from prior WOM literature (Trusov et al. 2009; Libai et al. 2013), we find that WOM in general (β) has a positive impact on long-term profits (Static Mean Quality: \$2,333.27, $p < .001$; Dynamic Mean Quality: \$3,537.74, $p < .001$). However, the interaction of WOM parameters with the *Group* dummy reveals that the greater the weight consumers place on WOM in general, the more negative its impact on the long-term profits of Group-Aware banks relative to Group-Blind banks (Static Mean Quality: \$-4,487.35, $p < .001$; Static Mean Quality: \$-6,431.03, $p < .001$). Regarding the weight placed on in-group sources of WOM (α), we find mixed statistical support for its impact on profits. Overall, the weight on in-group sourced WOM has a directionally positive but not statistically significant impact on long-term profits (Static Mean Quality: \$1,460.26, $p = .170$; Dynamic Mean Quality: \$2,241.08, $p < .091$). However, its effect on Group-Aware banks' long-term profits is negative and statistically significant (Static Mean Quality: \$-2,597.01, $p = .085$; Dynamic Mean Quality: \$-4,176.76, $p = .026$)

Comparing Group-Aware and Group-Blind banks' average short-term profits across all ABM conditions, we find that the Group-Aware banks have, on average, higher profits per loan under static mean quality conditions (Static Mean Quality: \$72.83 Group-Aware vs. \$69.05 Group-Blind, $p < .001$). This is consistent with our hypothesis in Proposition 2 which suggests that discrimination is profitable in the short-run. However, under dynamic mean quality conditions, we find directional but not statistically significant support (Dynamic Mean Quality: \$100.54 Group-Aware vs. \$100.20 Group-Blind, $p = .83$). This is probably because the L-group grows in mean quality throughout the simulation to eventually equal the H-group population by the end of the simulation.

We have shown that short term profits are better when the service adopts a Group-Aware policy. However, when we compare average Group-Aware and Group-Blind banks' long-term profits across all ABM conditions, we find a reversal. Figure 2 (see after Reference) shows average cumulative profits of each type of bank across ABM conditions.

[Insert Figure 2 about here.]

On average, Group-Blind banks have sizably greater cumulative profits than Group-Aware banks (Static Mean Quality: \$255,437.73 Group-Blind vs. \$240,966.49 Group-Aware, $p < .001$; Dynamic Mean Quality: \$339,956.23 Group-Blind vs. \$313,239.71 Group-Aware, $p < .001$). By regressing cumulative profits on *Group*, *time*, $time^2$, and their interactions, we find that while the main effect on *Group* (representing Group-Aware banks) is negative but not significant, its interaction with *time* indicates that Group-Aware bank profits substantially erode over time (Static Mean Quality: \$-134.73, $p < .05$; Dynamic Mean Quality: \$-196.93, $p < .01$).

We also ran a series of additional ABM simulations and analyses to check the robustness of our results and to further understand the impact of WOM and competition on the outcome. We have included details of the analysis in Web Appendix F. These simulations provide additional support for our findings: the interaction of WOM and competition causes the reversal of profit in favor of Group-Blind banks. Table 2 summarizes the key results from the ABM analysis.

[Insert Table 2 about here.]

First, we simulated a city with only one bank (a monopoly scenario). Comparing the monopoly scenario with a Group-Aware bank versus one with a Group-Blind bank, we find that the Group-Aware bank is significantly more profitable (first row in Table 2). This is consistent with Proposition 2's assertion that Group-Aware banks are more profitable. Then we simulated a city with two banks who compete with the same service policies (i.e., Group-Aware vs Group-Aware banks and Group-Blind vs. Group-Blind banks). There is no WOM among the consumers in this scenario. We find no significant difference in profit between the two banks in this scenario. We next simulated a city with two banks competing with alternative service policies. Although the Group-Aware bank is directionally more profitable, we find no significant differences in profit (third row in Table 2). Next, we ran all competitive scenarios mentioned with consumer WOM. In all cases we find no statistical difference in profits except in the scenario of competing alternative service policies and the presence of consumer WOM. In this scenario, we find that Group-Blind banks earn significantly more than Group-Aware banks. This outcome supports the conclusion that in the long-run, myopically profitable, rationally-based algorithmic discrimination does not pay.

DISCUSSION

Summary

Our study shows how discrimination in service can emerge from algorithmic decision-making designed with no prejudicial intent. Such decision-making implies that discrimination can be profitable in the short-run, but we show that it can become unprofitable in the long-run. This is especially true if consumer word-of-mouth is extensive, as is increasingly the case, given modern social media. Although many associate discrimination with race, ethnicity, and gender, our theory and findings should apply equally to contexts beyond these categories. They should apply to any service scenario where the decision-making algorithm 1) can segment consumers based on an observable attribute into distinct groups; 2) has uncertainty about the quality (e.g., profitability) of the prospective consumer before providing service to the consumer; 3) forms an expectation of the consumer's quality by using information about both the individual consumer and the group of which she or he is a member; 4) screens prospective consumers based on expected quality, and ultimately selects which consumers to serve. For example, consider how our theory applies to the scenario of the algorithm that makes recommendations on whether to rent the luxury apartment to a 18-year old man versus a 65-year old man. Or perhaps the apartment rental algorithm decides between a 65-year old garbage man versus a 65-year old business man who have equal net worth. Service decisions such as these, in isolation, may seem to have little impact on firm profits. But the macro social patterns that can emerge from service decisions that rely on group information can produce discriminatory outcomes with negative long-term profit implications.

Theoretical Contributions

Our research provides the following theoretical contributions to the literature: First, we examine the critical role that variance plays in the emergence and persistence of service discrimination. Our research shows that service discrimination can arise from low intra-group variation in consumer quality and high measurement error of consumer quality. Second, our findings demonstrate that temporal changes in group mean quality level can potentially improve or exacerbate service discrimination. We find conditions where an L-group can experience increasing discrimination despite its improving mean quality levels over time. This is of concern because historically L-groups have been improving in quality over time in the U.S. Third, we show that if word-of-mouth is influential, and if competition can provide outside options to consumers, then a Group-Aware service policy using a minimum score criterion can be more profitable in the short-run, but less profitable in the long run compared to a Group-blind service policy. This matters because a myopic firm can be led down a damaging path by short-term profitability when using group information in its service decisions.

Implications for Policymakers

Our findings suggest that potential discriminatory outcomes from decision-making algorithms can generate word-of-mouth that impacts long-term consumer demand and profits for firms. Consequently, our research has multiple public policy implications. First, it has implications for methods to detect potential cases where decision-making algorithms are producing disparate impact on affected groups of consumers. A salient example is the case of the Apple Card, which is the story we began this paper with. As aforementioned, the volume of digital word-of-mouth (tweets) attracted the attention of New York financial regulators and motivated them to launch an investigation into whether the Apple Card algorithm is discriminatory in its credit decisions. Determining where to look for potentially discriminatory outcomes of algorithmic decisions is in itself a difficult task. Our research suggests that word-of-mouth could be a potential bellweather for regulators in the interest of detecting potential incidents of algorithmic discrimination in service, such as was the case in the Apple Card.

Furthermore, U.S. laws such as the Civil Rights Act of 1964, the Equal Credit Opportunity Act, and the Fair Housing Act strive to protect consumers from discrimination in service. However, the task of identifying and proving existence of discrimination when enforcing these laws has been a difficult and controversial one. For example, in May, 2018 the U.S. President signed Congress' joint resolution to strike down the Consumer Financial Protection Bureau (CFPB) agency's Bulletin, "Indirect Auto Lending and Compliance with the Equal Credit Opportunity Act". Deregulation of auto lending practices was the primary motivation of the resolution. This bulletin, designed to curb discriminatory auto lending practices, provided guidance to auto lenders regarding compliance with the Equal Credit Opportunity Act Consumer Financial Protection Bureau (2020). Congress struck down the bulletin because of the belief that the guidance actually impeded auto lenders' abilities to provide auto lending service. Striking down the bulletin significantly restricted the CFPB's ability to enforce ECOA and provided auto lenders the leeway to use different score cutoffs for different groups (Merle 2018). This is a public policy debate for which our research is relevant. One reason given for repealing the rule concerns the controversy regarding the Consumer Financial Protection Bureau's method of determining whether discrimination exists in the first place (Hayashi 2018). Our research theory, definition of discrimination, and our findings can provide a framework for developing tools to detect and measure discrimination. Furthermore, the same framework could be the basis of measurement for litigation cases of consumer discrimination.

It also has implications for the push by some policymakers for algorithmic transparency. The notion behind this concept is that if algorithm code is available for public examination and commentary, the algorithm owner will have the incentive to ensure the fairness of the algorithm itself. Furthermore, those that examine the code will be in a stronger position to challenge the algorithm owner if its found to be unfair, thereby increasing the owner's accountability. However, our research suggests that it is not clear whether such a policy would have prevented the disparate impact outcomes and subsequent demand shifts as a result of the algorithm. This is because public investigation may have found that the algorithm seems reasonable since it is based on the logic that the algorithm seeks to maximize profit and appears to be rational in its attempt to minimize risk: both reasonable pursuits for business.

Implications for Managers

These findings have important managerial implications for firms that meet the conditions of our research. In cases where group information is not part of the objective function, but is merely used to minimize information uncertainty, then our findings may apply. Furthermore, our findings apply to markets where 1) consumers do place value on word-of-mouth in their decision-making about services and 2) there is an outside option for the consumer in the form of an alternative service provider that ignores group information. If the necessary conditions are relevant for a firm, then our findings suggest that the firm may want to consider the following recommendations. This is especially true in the current digital age, in which word-of-mouth has a strong and pervasive presence through digital media.

First, we recommend that firms that use a Group-Aware policy and meet the conditions outlined should consider the long-term benefits of switching to a Group-Blind service policy where group membership information is ignored. We have shown that employing a Group-Blind service policy can provide a competitive advantage. It may initially seem that a Group-Aware service policy is optimal because such a service policy provides the decision-making service algorithm an effective device to screen out risky consumers and identify profitable ones. However, we have shown that such a policy can produce discrimination in service that erodes profits and market share over time. Because of strong word-of-mouth effects, consumers can learn from each other which firms are unlikely to provide favorable service conditions to them. If services with Group-Blind policies are available as competitive alternatives, L-group consumers in macro may switch their preferences for these services over time. If sufficient numbers of H-group consumers also patronize Group-Blind services as well, this trajectory could produce conditions where Group-Blind firms enjoy a profit advantage. Although discriminatory practices may seem profitable in the short term, they can damage service demand and profits in the long-run.

However, if the firm must use a Group-Aware policy, then we recommend that the firm consider measuring and monitoring the degree to which there is service discrimination, as well as its impact on profits. Furthermore, we recommend that Group-Aware firms invest in methods of measurement error reduction such as developing advanced methods of measuring consumer quality or more sophisticated predictive models that improve accuracy in predicting consumer quality. This suggestion is attenuated by the fact that investment can engender a cost that may alter the profit outcomes of our results, so the impact of such costs would have to be considered carefully. The Group-Aware firm could also increase its exposure to consumer populations, which could improve information about the mean and variance of group quality. For example, the firm could purchase outside data about target markets to supplement its internal data. This could be a way to reduce service discrimination by increasing the algorithm's exposure to a potentially wider range of consumer quality (i.e., increase data heterogeneity). By doing so, a firm can put itself on the path to reducing algorithmic discrimination in service and increasing its profits over time.

Implications for Consumers

These findings also have consumer implications. Our findings imply that consumers seeking less discriminatory experiences in service may seek services that are, by nature, Group-Blind. For example, the auto insurer Root Insurance uses a Group-Blind approach to market its services to consumers. In fact, a tagline on their website states "we use an app to rate drivers based on how they actually drive—not just their demographics" (Root 2019). Another consumer implication directly results from the knowledge that Group-Aware services are likely to have different service criteria for groups that differ in mean quality. With this knowledge, if a consumer must seek service from a Group-Aware service provider, he or she may choose to mask or omit his or her information regarding group membership. Alternatively, the consumer could seek the provider that has the most favorable minimum score criterion for his or her group. The consumer could also improve her outcome by acquiring attributes of the H-group (assimilation) when seeking service. For example, the man seeking to rent the luxury apartment may have a better service experience by providing more information associated with H-group attributes (e.g, type of car he drives) than he otherwise would, regardless of his age or occupation.

Limitations and Opportunities for Future Research

There are limitations to this research which suggest many ways that researchers can broaden our knowledge on this topic. For example, our theory assumes that consumers are members of only one population group. In reality, a consumer can be a member of multiple groups, some of which may be high quality groups while others may not (e.g., a wealthy entrepreneur who has no high-school or college degree). It would be interesting to explore the boundaries of our theory under conditions where consumers may have two or more group memberships with varying levels of mean quality. A second limitation is that we assume in our theory that algorithms continuously update expectations using historical information available about consumers who have sought their service. Although we have found qualitative support in our interviews that this can happen in loan services, this may not be true in all service contexts. A promising avenue for future research is investigating how varying the frequency of updates and varying the historical window of data about consumers can affect decision-making algorithm's expectations. A great deal of work is still needed to fully understand the nature and boundaries of service discrimination, but we believe that the theoretical framework created here serves as a starting point to exploring these and many more questions about the effects of algorithmic discrimination in service.

Conclusion

We had three goals at the outset of the research discussed in this paper: 1) to examine how service discrimination can emerge from algorithmic decision-making; 2) to investigate how service discrimination interacts with consumer word-of-mouth to affect demand and profits; and 3) to explore public policy and managerial implications of algorithmic discrimination in service outcomes. We did so by developing a theoretical model that illuminates the critical roles that variation in consumer quality and measurement error in detecting quality play in the emergence and magnitude of discrimination in service. With our agent-based model, we showed the long-term macro effects on profits when firm competition and consumer word-of-mouth embedded in a complex system are taken into consideration. We found that although Group-Blind service algorithms, that do not use consumer group membership information in its service decisions, are less profitable than their Group-Aware competitors in the short-run, Group-Blind service algorithms are more profitable in the long-run. This is because consumer word-of-mouth drives consumers to select the most service-friendly alternatives among competitive options.

We provide managerial recommendations on reducing service discrimination's profit-damaging effects. This research emphasizes the long-term benefits of switching to a service policy that does not use group identity information. However, for firms that must persist in using group identity information, this research has additional recommendations which include increasing investment in methods of measurement error reduction and increasing exposure to consumers of different populations. By doing so, a firm could reduce algorithmic service discrimination while improving its long-term profits and societal well-being.

References

Addady Michal (2016) 3 More Fast Food Workers Who Refused to Serve Law Enforcement.

Aigner Dennis J, Cain Glen G (1977) Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relations Review* 30(2):175.

Anderson Laurel, Ostrom Amy L (2015) Transformative Service Research: Advancing Our Knowledge About Service and Well-Being. *Journal of Service Research* 18(3):243–249.

Anderson Laurel, Ostrom Amy L, Bitner Mary Jo, Brown Stephen W, Burkhard Kevin A, Goul Michael, Smith-Daniels Vicki, Demirkan Haluk, Rabinovich Elliot (2010) Improving Well-Being through Transformative Services. *Journal of Service Research* 13(1):4–36.

Antonovics Kate (2006) Statistical Discrimination and Intergenerational Income Mobility. *Working Paper* 18.

Arrow Kenneth J (1973) The theory of discrimination. *In Discrimination in Labor Markets*, 3–33 (Princeton University Press).

Ayres Ian, Siegelman Peter (1995) Race and gender discrimination in bargaining for a new car. *The American Economic Review* 304–321.

Baldwin Marjorie L, Johnson William G (2006) A critical review of studies of discrimination against workers with disabilities. *Handbook on the Economics of Discrimination* 119–160.

Barabási Albert-Laszlo, Albert Reka (1999) Emergence of Scaling in Random Networks. *Science* 286:5.

Bass Frank M (1969) A new product growth for model consumer durables. *Management Science* 15(5):215–227, URL `http://dx.doi.org/10.1287/mnsc.15.5.215`.

Becker Gary S (1957) *The Economics of Discrimination* (University of Chicago Press), 1 edition.

Bergemann Dirk, Brooks Benjamin, Morris Stephen (2015) The limits of price discrimination. *American Economic Review* 105(3):921–57.

Bertrand M, Duflo E (2017) Field Experiments on Discriminationan. Banerjee Abhijit Vinayak, Duflo Esther, eds., *Handbook of Economic Field Experiments*, volume 1, 309–393 (North-Holland).

Bjerk David (2008) Glass ceilings or sticky floors? Statistical discrimination in a dynamic model of hiring and promotion. *The Economic Journal* 118(530):961–982.

Blanchard Lloyd, Zhao Bo, Yinger John (2008) Do lenders discriminate against minority and woman entrepreneurs? *Journal of Urban Economics* 63(2):467–497, ISSN 00941190, URL `http://dx.doi.org/10.1016/j.jue.2007.03.001`.

Blume Lawrence E (2005) Learning and statistical discrimination. *The American Economic Review* 95(2):118–121.

Blume Lawrence E (2006) The dynamics of statistical discrimination. *The Economic Journal* 116(515).

Bohren J Aislinn, Imas Alex, Rosenberg Michael (2017) The Dynamics of Discrimination: Theory and Evidence. *Working Paper* .

Bone Sterling A, Christensen Glenn L, Williams Jerome D (2014) Rejected, Shackled, and Alone: The Impact of Systemic Restricted Choice on Minority Consumers' Construction of Self. *Journal of Consumer Research* 41(2):451–474.

Bone Sterling A, Williams Jerome D, Christensen Glenn L (2010) When Consumer Well-Being Meets Small Business Ownership: Transforming Financial Service Systems to Eradicate Disparate Treatment and Discrimination. *Journal of Macromarketing* 30(4):408–422.

Broadwater Luke (2012) Wells Fargo agrees to $175 million settlement in pricing discrimination suit. *The Baltimore Sun* .

Brown Jacqueline Johnson, Reingen Peter H (1987) Social ties and word-of-mouth referral behavior. *Journal of Consumer research* 14(3):350–362.

Carpusor Adrian G, Loges William E (2006) Rental Discrimination and Ethnicity in Names1. *Journal of Applied Social Psychology* 36(4):934–952, ISSN 1559-1816, URL `http://dx.doi.org/10.1111/j.0021-9029.2006.00050.x`.

CDIpress (2019) Commissioner issues regulations prohibiting gender discrimination in automobile insurance rates. URL `http://www.insurance.ca.gov/0400-news/0100-press-releases/2019/release003-19.cfm`, [Accessed May 10, 2019].

Ching Andrew T, Erdem Tülin, Keane Michael P (2013) Learning Models: An Assessment of Progress, Challenges, and New Developments. *Marketing Science* 32(6):913–938, ISSN 0732-2399, 1526-548X.

Coate Stephen, Loury Glenn C (1993) Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review* 83(5):1220–1240.

Consumer Financial Protection Bureau (2020) CFPB to Hold Auto Lenders Accountable for Illegal Discriminatory Markup. URL `https://www.consumerfinance.gov/about-us/newsroom/consumer-financial-protection-bureau-to-hold-auto-lenders-accountable-for-illegal-d` library Catalog: www.consumerfinance.gov.

Corbett-Davies Sam, Pierson Emma, Feller Avi, Goel Sharad, Huq Aziz (2017) Algorithmic decision making and the cost of fairness. *arXiv:1701.08230 [cs, stat]* URL `http://dx.doi.org/10.1145/3097983.309809`, arXiv: 1701.08230.

Cowgill Bo (2018) The Impact of Algorithms on Judicial Discretion: Evidence from Regression Discontinuities. *Working Paper* 24.

Craig Ashley C, Fryer Roland G (2017) Complementary Bias: A Model of Two-Sided Statistical Discrimination. *NBER Working Paper* .

Crockett David, Anderson Laurel, Bone Sterling A, Roy Abhijit, Wang Jeff Jianfeng, Coble Garrett (2011) Immigration, Culture, and Ethnicity in Transformative Consumer Research. *Journal of Public Policy & Marketing* 30(1):47–54.

Crockett David, Grier Sonya A, Williams Jacqueline A (2003) Coping with marketplace discrimination: An exploration of the experiences of black men. *Academy of Marketing Science Review* 2003:1.

Defend Civil Rights (2020) Who We Are. URL `https://www.defendcivilrights.org/who-we-are`, accessed: 2020-03-16.

Delre Sebastiano A, Broekhuizen Thijs LJ, Bijmolt Tammo HA (2016a) The effects of shared consumption on product life cycles and advertising effectiveness: The case of the motion picture market. *Journal of Marketing Research* 53(4):608–627.

Delre Sebastiano A, Panico Claudio, Wierenga Berend (2016b) Competitive strategies in the motion picture industry: An abm to study investment decisions. *International Journal of Research in Marketing* 34(1):69 – 99.

Edelman Benjamin, Luca Michael, Svirsky Dan (2017) Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics* 9(2):1–22, ISSN 1945-7782, URL `http://dx.doi.org/10.1257/app.20160213`.

Elliott Andrea (2003) Suit Accuses Macy's of Bias Against Minority Shoppers. *The New York Times* ISSN 0362-4331.

Erdem Tülin, Keane Michael P (1996) Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science* 15(1):1–20.

Erdös Paul, Rényi Alfréd (1959) On random graphs, i. *Publicationes Mathematicae (Debrecen)* 6:290–297.

Evett Sophia R, Hakstian Anne-Marie G, Williams Jerome D, Henderson Geraldine R (2013) What's Race Got to Do with It? Responses to Consumer Discrimination: Responses to Consumer Discrimination. *Analyses of Social Issues and Public Policy* 13(1):165–185.

Ewens Michael, Tomlin Bryan, Wang Liang Choon (2013) Statistical Discrimination or Prejudice? A Large Sample Field Experiment. *The Review of Economics and Statistics* 96(1):119–134, ISSN 0034-6535, URL `http://dx.doi.org/10.1162/REST_a_00365`.

Fader Peter S, Hardie Bruce GS, Lee Ka Lok (2005) RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. *Journal of Marketing Research* 42(4):415–430, ISSN 0022-2437, 1547-7193, URL http://dx.doi.org/10.1509/jmkr.2005.42.4.415.

Fang Hanming, Moro Andrea (2011) Theories of statistical discrimination and affirmative action: A survey. Benhabib Jess, ed., *Handbook of Social Economics. Vol. 1A* (Elsevier, North-Holland).

Ferguson Michael F, Peters Stephen R (1995) What Constitutes Evidence of Discrimination in Lending? *The Journal of Finance* 50(2):739, ISSN 00221082, URL http://dx.doi.org/10.2307/2329427.

Fryer Roland G (2007) Belief flipping in a dynamic model of statistical discrimination. *Journal of Public Economics* 91(5-6):1151–1166.

Gill Jeff (2007) *Bayesian Methods: A Social and Behavioral Sciences Approach, Second Edition* (CRC Press), ISBN 978-1-58488-562-7, google-Books-ID: Iq_epk4mtM4C.

Goldenberg Jacob, Libai Barak, Muller Eitan (2001a) Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters* 12(3):211–223.

Goldenberg Jacob, Libai Barak, Muller Eitan (2001b) Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review* 2001(9):1.

Goldenberg Jacob, Libai Barak, Muller Eitan (2010) The chilling effects of network externalities. *International Journal of Research in Marketing* 27(1):4–15.

Goodman Bryce, Flaxman Seth (2017) European Union regulations on algorithmic decision-making and a ”right to explanation”. *AI Magazine* 38(3):50–57, ISSN 2371-9621, 0738-4602, URL http://dx.doi.org/10.1609/aimag.v38i3.2741, arXiv: 1606.08813 version: 2.

Granovetter Mark (1983) The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory* 1:201, ISSN 07352751, URL http://dx.doi.org/10.2307/202051.

Granovetter Mark S (1973) The strength of weak ties. *American journal of sociology* 78(6):1360–1380.

Gutierrez Lisa (2015) Shoppers fighting back against ‘shopping while black’ profiling with social media, lawsuits. *miamiherald* .

Haenlein Michael, Kaplan Andreas M (2010) An Empirical Analysis of Attitudinal and Behavioral Reactions Toward the Abandonment of Unprofitable Customer Relationships. *Journal of Relationship Marketing* 9(4):200–228, ISSN 1533-2667, 1533-2675, URL http://dx.doi.org/10.1080/15332667.2010.522474.

Haenlein Michael, Kaplan Andreas M (2012) The impact of unprofitable customer abandonment on current customers' exit, voice, and loyalty intentions: an empirical analysis. *Journal of Services Marketing* 26(6):15.

Haenlein Michael, Kaplan Andreas M, Schoder Detlef (2006) Valuing the Real Option of Abandoning Unprofitable Customers When Calculating Customer Lifetime Value. *Journal of Marketing* 70(3):16.

Haenlein Michael, Libai Barak (2013) Targeting Revenue Leaders for a New Product. *Journal of Marketing* 77(3):65–80, ISSN 0022-2429, 1547-7185, URL `http://dx.doi.org/10.1509/jm.11.0428`.

Haggerty Neil (2018) Trump makes repeal of CFPB auto lending rule official. URL `https://www.americanbanker.com/news/trump-makes-repeal-of-cfpb-auto-lending-rule-official`.

Hanson Andrew, Hawley Zackary (2011) Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in US cities. *Journal of Urban Economics* 70(2):99–114, ISSN 0094-1190, URL `http://dx.doi.org/10.1016/j.jue.2011.02.003`.

Harris Anne-Marie G, Henderson Geraldine R, Williams Jerome D (2005) Courting Customers: Assessing Consumer Racial Profiling and Other Marketplace Discrimination. *Journal of Public Policy & Marketing* 24(1):163–171.

Harris Diane (2019) The Apple Card's supposed gender bias? Don't assume its discrimination, experts warn. URL `https://www.newsweek.com/apple-card-gender-bias-credit-limit-goldman-sachs-1471146`.

Hayashi Yuka (2018) Senate Votes to End Auto-Loan Anti-Discrimination Rule. *Wall Street Journal* .

Hill Ronald Paul, Stephens Debra Lynn (2003) The Compassionate Organization in the 21st Century. *Organizational Dynamics* 32(4):331–341.

Homburg Christian, Droll Mathias, Totzek Dirk (2008) Customer Prioritization: Does it Pay off, and how Should it be Implemented? *Journal of Marketing* 72(5):110–130, ISSN 0022-2429, 1547-7185, URL `http://dx.doi.org/10.1509/jmkg.72.5.110`.

Huff David L (1964) Defining and estimating a trading area. *The Journal of Marketing* 34–38.

Kleinberg Jon, Lakkaraju Himabindu, Leskovec Jure, Ludwig Jens, Mullainathan Sendhil (2018) Human decisions and machine predictions. *The Quarterly Journal of Economics* 133(1):237–293.

Koren James Rufus (2016) Toyota will compensate black and Asian borrowers to settle bias investigation. *Los Angeles Times* ISSN 0458-3035.

Koren James Rufus (2018) Sacramento becomes latest city to allege mortgage discrimination by Wells Fargo. URL https://www.latimes.com/business/la-fi-wells-fargo-sacramento-20180228-story.html.

Kugelmass Heather (2016) "sorry, i'm not accepting new patients": An audit study of access to mental health care. *Journal of Health and Social Behavior* 57(2):168–183.

Ladd Helen F (1998) Evidence on Discrimination in Mortgage Lending. *The Journal of Economic Perspectives* 12(2,):41–62, URL http://www.jstor.org/stable/2646961.

Lam Desmond, Lee Alvin, Mizerski Richard (2009) The effects of cultural values in word-of-mouth communication. *Journal of international marketing* 17(3):55–70.

Lambrecht Anja, Tucker Catherine (2019) Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science* 65(7):2966–2981, ISSN 0025-1909, 1526-5501, URL http://dx.doi.org/10.1287/mnsc.2018.3093.

Leonard Jonathan S, Levine David I, Giuliano Laura (2010) Customer Discrimination. *Review of Economics and Statistics* 92(3):670–678, ISSN 0034-6535, 1530-9142, URL http://dx.doi.org/10.1162/REST_a_00018.

Lepthien Anke, Papies Dominik, Clement Michel, Melnyk Valentyna (2017) The ugly side of customer management – Consumer reactions to firm-initiated contract terminations. *International Journal of Research in Marketing* 34(4):829–850, ISSN 01678116, URL http://dx.doi.org/10.1016/j.ijresmar.2017.02.001.

LexisNexis (1970) 4,970 results for apple card discrimination. URL https://advance-lexis-com.libproxy2.usc.edu/search/?pdmfid=1516831&crid=1c8b8779-bdba-4868-8252-da67ebeb7e6e&pdsearchterms=apple+card+discrimination&pdstartin=hlct%3A1%3A1&pdtypeofsearch=searchboxclick&pdsearchtype=SearchBox&pdqttype=and&pdpsf=&pdquerytemplateid=&ecomp=b7qgkgk&earg=pdpsf&prid=c30fcaf8-8bdc-4b34-9c8b-2f83342a4f4c, [Accessed Jan. 11, 2020].

Libai Barak, Muller Eitan, Peres Renana (2013) Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. *Journal of Marketing Research* 50(2):161–176.

List J A (2004) The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field. *The Quarterly Journal of Economics* 119(1):49–89, ISSN 0033-5533, 1531-4650, URL http://dx.doi.org/10.1162/003355304772839524.

Loconte Richard (2018) Press Release - March 12, 2018: DFS Announces New Agreement with Geico to Protect New York Drivers from Unfairly Discriminatory Auto Insurance Rates. URL https://www.dfs.ny.gov/reports_and_publications/press_releases/pr1803131, [Accessed May 10, 2019].

Merle Renae (2018) Senate poised to roll back watchdog effort to prevent discrimination in auto lending market. *Washington Post* .

Mick DG (2006) Meaning and Mattering Through Transformative Consumer Research. *Advances in Consumer Research* 33(1):1–4.

Milkman Katherine L, Akinola Modupe, Chugh Dolly (2012) Temporal Distance and Discrimination: An Audit Study in Academia. *Psychol Sci* 23(7):710–717, ISSN 0956-7976, URL `http://dx.doi.org/10.1177/0956797611434539`.

Mullainathan Sendhil, Spiess Jann (2017) Machine learning: an applied econometric approach. *Journal of Economic Perspectives* 31(2):87–106.

Narasimhan Chakravarthi (1984) A price discrimination theory of coupons. *Marketing Science* 3(2):128–147.

Novick Melvin R (1965) The Axioms and Principal Results of Classical Test Theory. *ETS Research Bulletin Series* 1965(1):i–31.

Ouellet Jean-François (2007) Consumer racism and its effects on domestic cross-ethnic product purchase: An empirical test in the United States, Canada, and France. *Journal of Marketing* 71(1):113–128.

Pager Devah, Shepherd Hana (2008) The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. *Annual Review of Sociology* 34(1):181–209.

Peres Renana, Van den Bulte Christophe (2014) When to take or forgo new product exclusivity: Balancing protection from competition against word-of-mouth spillover. *Journal of Marketing* 78(2):83–100.

Pettigrew S (2001) Why a Journal of Research for Consumers. *Journal of Research for Consumers* 33(1):1–4.

Phelps Edmund S (1972) The statistical theory of racism and sexism. *The american economic review* 62(4):659–661.

Podoshen Jeffrey Steven (2006) Word of mouth, brand loyalty, acculturation and the American Jewish consumer. *Journal of Consumer Marketing* 23(5):266–282.

Quillian Lincoln (2006) New Approaches to Understanding Racial Prejudice and Discrimination. *Annual Review of Sociology* 32(1):299–328.

Rand William, Rust Roland T (2011) Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing* 28(3):181–193.

Roberts John H, Urban Glen L (1988) Modeling Multiattribute Utility, Risk, and Belief Dynamics for New Consumer Durable Brand Choice. *Management Science* 34(2):167–185, ISSN 0025-1909, 1526-5501, URL `http://dx.doi.org/10.1287/mnsc.34.2.167`.

Rodgers William M (2009) *Handbook on the Economics of Discrimination* (Edward Elgar Publishing).

Root Insurance Co (2019) FAQ and customer support | Answers to Root Car Insurance questions. URL https://www.joinroot.com/help-t20/.

Rust Roland T, Donthu Naveen (1995) Capturing geographically localized misspecification error in retail store choice models. *Journal of Marketing Research* 32(1):103–110.

Savage David G (2017) A baker refused to make a cake for a gay couple due to religious beliefs. Supreme Court will rule on the case in fall. *Los Angeles Times* ISSN 0458-3035.

Schelling Thomas C (1969) Models of segregation. *The American Economic Review* 59(2):488–493.

Schmittlein David C, Peterson Robert A (1994) Customer Base Analysis: An Industrial Purchase Process Application. *Marketing Science* 13(1):41–67, ISSN 0732-2399, 1526-548X, URL http://dx.doi.org/10.1287/mksc.13.1.41.

Schroeder Pete (2017) American Express to pay $96 million over discriminatory card terms. *Reuters* .

Silver-Greenberg Jessica (2012) American Express Agrees to Refund $85 Million. *The New York Times* .

Trusov Michael, Bucklin Randolph E, Pauwels Koen (2009) Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. *Journal of Marketing* 73(5):90–102.

Twitter (2006) Twitter. URL https://twitter.com/home, [Accessed Jan. 11, 2020].

US Department of Housing and Urban Development (2019) HUD's Implementation of the Fair Housing Act's Disparate Impact Standard. URL https://www.federalregister.gov/documents/2019/08/19/2019-17542/huds-implementation-of-the-fair-housing-acts-disparate-impact-standard, accessed: 2020-03-16.

Uslu Aypar, Durmuş Beril, Taşdemir Sina (2013) Word of Mouth, Brand Loyalty, Acculturation and the Turkish Ethnic Minority Group in Germany. *Procedia - Social and Behavioral Sciences* 99:455–464.

Varian Hal R (1989) Price discrimination. *Handbook of industrial organization* 1:597–654.

Vigdor Neil (2019) Apple Card Investigated After Gender Discrimination Complaints. URL https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html.

Vollmer Nicholas (2018) Article 22 EU General Data Protection Regulation (EU-GDPR). URL `https://www.privacy-regulation.eu/en/article-22-automated-individual-decision-making-including-profiling-GDPR.htm`, library Catalog: www.privacy-regulation.eu Publisher: SecureDataService.

Wilensky Uri (1999) *NetLogo*, volume 4952 (Center for Connected Learning and Computer-based Modeling).

Wilensky Uri, Rand William (2015) *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo* (Cambridge, Massachusetts: The MIT Press), ISBN 978-0-262-73189-8.

Williams David R, Mohammed Selina A (2009) Discrimination and racial disparities in health: evidence and needed research. *J Behav Med* 32(1):20–47, ISSN 1573-3521, URL `http://dx.doi.org/10.1007/s10865-008-9185-0`.

Yinger John (1998) Evidence on Discrimination in Consumer Markets. *Journal of Economic Perspectives* 12:18.

Zhao Min, Xie Jinhong (2011) Effects of social and temporal distance on consumers' responses to peer recommendations. *Journal of Marketing Research* 48(3):486–496.

Table 1: Summary of Analytical Model Results

Proposition	Result	Implication
1.1	Under admissible conditions, L-group consumers must meet a higher service criterion than H-group consumers to receive the same service.	The greater the inequality $(A_H - A_L)$ between the H-group and the L-group, the more likely L-group members will have to meet a higher service criterion to receive service.
1.2	Under admissible conditions, H-group consumers must meet a higher service criterion than L-group consumers to receive the same service.	If the firm has uncertainty about the average quality of the L-group is low enough relative to the H-group, then the firm will trade off the profitability of the H-group in favor of the lower uncertainty of the L-group when setting service criteria for groups.
2	It is more profitable for a service provider to have a service criterion for each consumer group (Group-Aware service policy) than to have a single service criterion for all consumers (Group-Blind service policy).	If the firm is a monopoly, it is more profitable to discriminate.
3.1	Discrimination increases as the magnitude of measurement error of consumer quality increases.	Greater accuracy in assessing individuals results in less discrimination.
3.2	Discrimination in service decreases as consumer heterogeneity within each group increases.	More variation within a group results in less discrimination.

Note: H = High mean quality; L = Low mean quality; quality = the customer's value to the firm.

Table 2: Comparison of Avg. Long-Run Profits from ABM Results

Competitive Scenario	WOM Present?	More Profitable?	Significant Difference?
Monopoly	No	Group-Aware	Yes
Same Service Policies	No	NA	No
Alternate Service Policies	No	Group-Aware	No
Same Service Policies	Yes	NA	No
Alternate Service Policies	Yes	Group-Blind	Yes

Figure 1: Analytical Model of Algorithmic Discrimination in Service

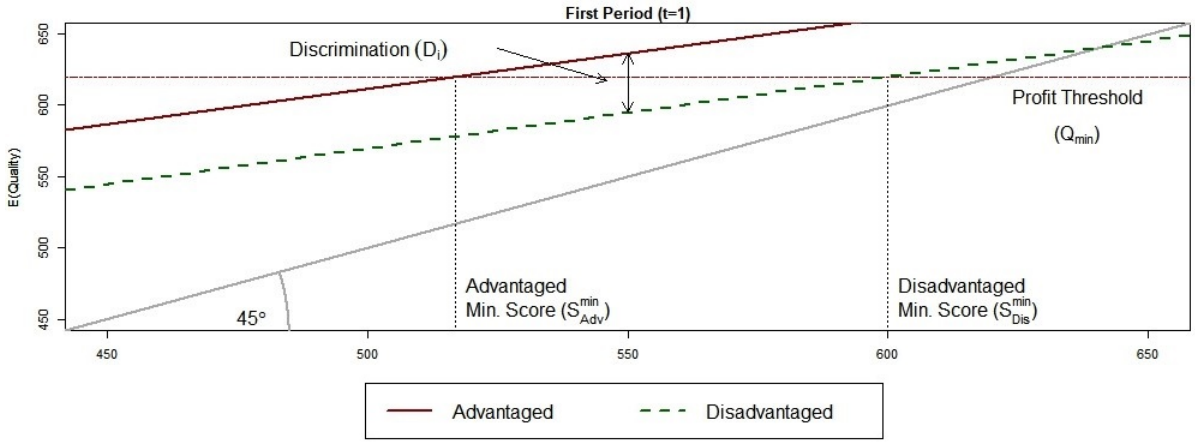
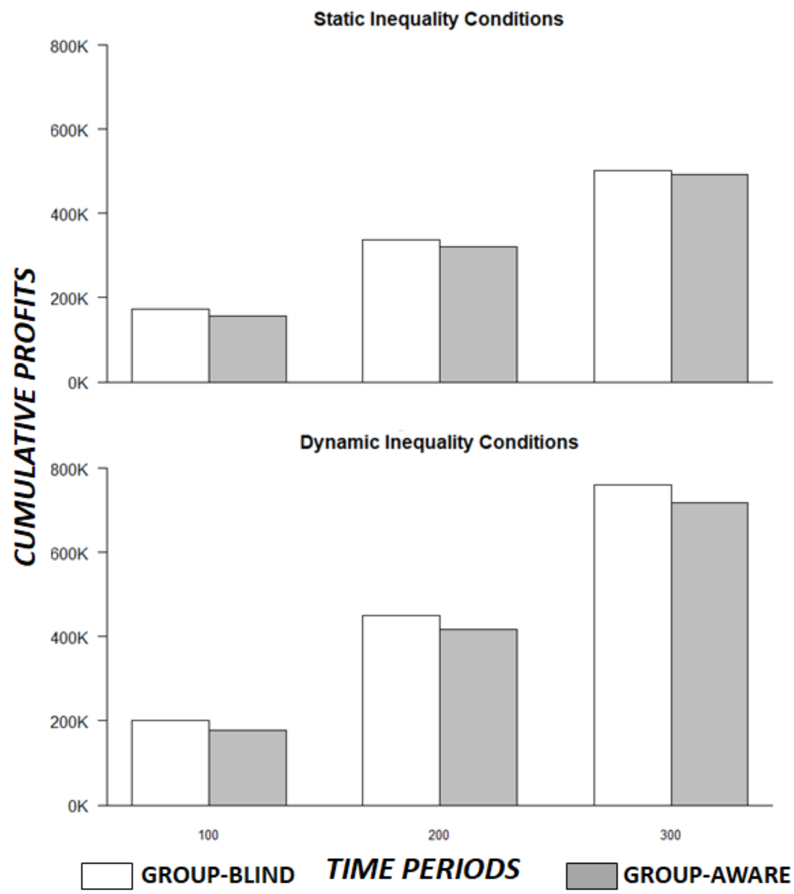


Figure 2: Long-Term Profits: Group-Blind vs. Group-Aware Banks from ABM Analysis



Web Appendix A: Analytical Proofs

Because S_{ij} has error, the loan algorithm may supplement the score with information about the group of which applicant i is a member. Although each group's true mean (A_j) and variance ($\sigma^2_{q_j}$) of quality are unknown, we assume that their distributions are known and that there is a prior: a normal distribution for the mean, an inverse-gamma distribution for the variance, and a normal-inverse-gamma joint distribution prior on the mean and variance. These assumptions are consistent with normally distributed Bayesian updating models with unknown mean and variance (Gill 2007), which results in the following Bayesian posteriors:

$$\begin{aligned}
&P(A_j \mid \sigma^2_{q_j}, S_j) \sim \mathscr{N}\left(\frac{n_0 A_{j0} + n_j \bar{S}_j}{n_0 + n_j}, \frac{\sigma^2_{q_j}}{n_0 + n_j}\right) \\
&P(\sigma^2_{q_j} \mid S_j) \sim \mathscr{IG}\left(\frac{n_0 + n_j}{2}, \frac{n_0 \sigma^2_{q_{j0}} + n_j \bar{\sigma}^2_{q_j} + \frac{n_0 n_j}{n_0 + n_j}(A_{j0} - \bar{S}_j)^2}{2}\right) \\
&\text{where } \bar{S}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} S_{ij} \text{ and } \bar{\sigma}^2_{q_j} = \frac{1}{n_j}\sum_{i=1}^{n_j}(S_{ij} - \bar{S}_j)^2 \\
&\{A_{j_0}, \sigma^2_{q_{j0}}, n_0\} = \{\text{priors on } A_j, \text{ and } \sigma^2_{q_j}, \text{ and } n_j \text{ (number of data points), respectively}\}
\end{aligned} \tag{10}$$

Consequently, the means of the posterior distributions in Equations 10 are the loan officer's expectations of the mean and variance of the quality of group j:

$$\begin{aligned}
E(A_j \mid \sigma^2_{q_j}, S_j) &= \hat{A}_j = \frac{n_0 A_{j0} + n_j \bar{S}_j}{n_0 + n_j} \\
E(\sigma^2_{q_j} \mid S_j) &= \hat{\sigma}^2_{q_j} = \frac{n_0 \sigma^2_{q_{j0}} + n_j \bar{\sigma}^2_{q_j} + \frac{n_0 n_j}{n_0 + n_j}(A_{j0} - \bar{S}_j)^2}{n_0 + n_j}
\end{aligned} \tag{11}$$

The score reliability has the following important properties:

$$0 < \gamma_j < 1, \quad \frac{\partial \gamma_j}{\partial \hat{\sigma}^2_{q_j}} = \frac{\sigma^2_{\varepsilon_j}}{(\hat{\sigma}^2_{q_j} + \sigma^2_{\varepsilon_j})^2} > 0, \text{ and } \frac{\partial \gamma_j}{\partial \sigma^2_{\varepsilon_j}} = \frac{-\hat{\sigma}^2_{q_j}}{(\hat{\sigma}^2_{q_j} + \sigma^2_{\varepsilon_j})^2} < 0 \tag{12}$$

Let p_H and $p_L = 1 - p_H$ represent the proportion of all applicants that are members of the H and L groups respectively. Using the equations for pooled mean and variance, the mean quality, variance in quality, score reliability, and

(13a) $$\hat{A}_{all} = p_H \hat{A}_H + (1 - p_H)\hat{A}_L$$

(13b) $$\hat{\sigma}^2_{q_{all}} = \hat{\sigma}^2_q + p_H(1 - p_H)(\hat{A}_H - \hat{A}_L)^2 > \hat{\sigma}^2_q$$

(13c) $$\gamma_{all} = \frac{\hat{\sigma}^2_{q_{all}}}{\hat{\sigma}^2_{q_{all}} + \sigma^2_\varepsilon} > \gamma$$

(13d) $$S^{min}_{all} = Q^{min} + \left(Q^{min} - \hat{A}_{all}\right)\left(\frac{1 - \gamma_{all}}{\gamma_{all}}\right)$$

Let $f_j(S)$ and $F_j(S)$ respectively represent the probability density function and cumulative distribution function of group j scores. The loan officer's expected profits (Π) under the Group-Aware and Group-Blind policies are respectively:

(14a) $$E(\Pi \mid S^{min}_{j \in \{L,H\}}) = \sum_{j \in \{L,H\}} \int_{S^{min}_j}^{\infty} \frac{p_j\, E(Q_{ij} \mid S_{ij}) f_j(S) dS}{2 - F_H(S^{min}_H) - F_L(S^{min}_L)}$$

(14b) $$E(\Pi \mid S^{min}_{all}) = \sum_{j \in \{L,H\}} \int_{S^{min}_{all}}^{\infty} \frac{p_j\, E(Q_i \mid S_i) f_j(S) dS}{2 - F_H(S^{min}_{all}) - F_L(S^{min}_{all})}$$

Using Equations 10, the following represents the loan officer's posterior beliefs of j's mean quality level, variation in quality, score reliability, and the minimum score criterion (where subscript c represent current beliefs as of the end of $t = 2$ in a two-period model):

(15a) $$\begin{aligned}\hat{A}_{jc}(g_j) &= p_{j1}\hat{A}_{j1} + p_{j2}\hat{A}_{j2} \\ &= \hat{A}_{j1}\left[p_{j1} + g_j(1 - p_{j1})\right]\end{aligned}$$

(15b) $$\begin{aligned}\hat{\sigma}^2_{q_{jc}}(g_j) &= p_{j1}\hat{\sigma}^2_{q_{j1}} + p_{j2}\hat{\sigma}^2_{q_{j2}} + p_{j1}p_{j2}\left[\hat{A}_{j1} - \hat{A}_{j2}\right]^2 \\ &= \hat{\sigma}^2_{q_{j1}} + p_{j1}(1 - p_{j1})\left[\hat{A}_{j1}(1 - g_j)\right]^2 \geq \hat{\sigma}^2_{q_{j1}}\end{aligned}$$

(15c) $$\gamma_{jc}(g_j) = \frac{\hat{\sigma}^2_{q_{jc}}(g_j)}{\hat{\sigma}^2_{q_{jc}}(g_j) + \sigma^2_\varepsilon} \geq \gamma_{j1}$$

(15d) $$S^{min}_{jc} = Q^{min} + \left(Q^{min} - \hat{A}_{jc}(g_j)\right)\left(\frac{1 - \gamma_{jc}(g_j)}{\gamma_{jc}(g_j)}\right)$$

For all proofs, we assume the following:

1. Each loan applicant i is a member of one of two population groups $j \in \{H, L\}$. Initially, the H-group has a mean quality level that is greater than the L group ($A_H > A_L$). The groups are initially equal in intra-group variation in quality ($\sigma_q^2 = \sigma_{q_H}^2 = \sigma_{q_L}^2$). We also assume that the bank's ability to measure quality is unaffected by changes in composition of the groups (thus σ_ε^2 is constant across groups and across time).

Proof of Proposition 1

Let the given assumptions and findings from Equations (6) and (13) hold. Then, the following is true:

$$\begin{gathered}
S_L^{min} < S_H^{min} \\
Q^{min} + \left(Q^{min} - \hat{A}_L\right)\left(\frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_{q_L}^2}\right) < Q^{min} + \left(Q^{min} - \hat{A}_H\right)\left(\frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_{q_H}^2}\right) \\
\left(\frac{\hat{A}_H - Q^{min}}{\hat{\sigma}_{q_H}^2}\right) < \left(\frac{\hat{A}_L - Q^{min}}{\hat{\sigma}_{q_L}^2}\right) \\
\frac{\hat{\sigma}_{q_L}^2}{\hat{\sigma}_{q_H}^2} < \frac{\hat{A}_L - Q^{min}}{\hat{A}_H - Q^{min}} \\
\hat{\sigma}_{q_L}^2 < \hat{\sigma}_{q_H}^2\left(\frac{\hat{A}_L - Q^{min}}{\hat{A}_H - Q^{min}}\right) = \sigma^* \\
\therefore S_L^{min} < S_H^{min} \text{ when} \\
\hat{\sigma}_{q_L}^2 < \sigma^* \text{ and } \hat{A}_H > \hat{A}_L > Q^{min} \text{ or } \hat{\sigma}_{q_L}^2 > \sigma^* \text{ and } Q^{min} > \hat{A}_H > \hat{A}_L
\end{gathered} \tag{16}$$

Proof of Proposition 2

Before we show the proof for Proposition 2, we first must establish the ordinal relationship of S_{all}^{min}, S_L^{min}, and S_H^{min}. Proposition 1 establishes that the minimum score criteria derived from a Group-Aware policy have the ordinal relationship $S_L^{min} > S_H^{min}$ as long as $\hat{\sigma}_{q_L}^2 > \sigma^*$. We establish the ordinal relationship of the Group-Aware firm's minimum score criteria with respect to S_{all}^{min} in the following Lemma:

Lemma 1 *Under admissible conditions, the Group-Aware policy minimum score criterion for the L-group is greater than the minimum score criterion of a Group-Blind policy (i.e., $S_{all}^{min} < S_L^{min}$).*

Proof by contradiction: Let us suppose the contrary, that $S_{all}^{min} > S_L^{min}$. Drawing from Equations (6) and (13), that implies:

$$S_{all}^{min} > S_L^{min}$$

$$Q^{min} + \left(Q^{min} - A_{all}\right)\left(\frac{\sigma_\varepsilon^2}{\sigma_{q_{all}}^2}\right) > Q^{min} + \left(Q^{min} - A_L\right)\left(\frac{\sigma_\varepsilon^2}{\sigma_{q_L}^2}\right)$$

$$\frac{Q^{min} - A_{all}}{Q^{min} - A_L} > \frac{\sigma_{q_{all}}^2}{\sigma_{q_L}^2} \tag{17}$$

Since $(Q^{min} - A_{all}) < (Q^{min} - A_L)$ and $\sigma_{q_{all}}^2 > \sigma_{q_L}^2$, then

$\dfrac{Q^{min} - A_{all}}{Q^{min} - A_L} < 1 < \dfrac{\sigma_{q_{all}}^2}{\sigma_{q_L}^2}$, which is a contradiction.

$$\therefore S_{all}^{min} < S_L^{min} \ \forall \ \sigma_q^2, \sigma_\varepsilon^2, Q^{min}, A_j$$

We wish to establish the conditions where $E(\Pi \mid S_j^{min}) \geq E(\Pi \mid S_{all}^{min})$: the average per period profit resulting from a Group-Aware service policy is greater than that of a Group-blind service policy. Based on the equations in (14), we can expand this inequality and rearrange terms as follows:

$$E(\Pi \mid S_{j\in Adv,L}^{min}) > E(\Pi \mid S_{all}^{min})$$

$$\sum_{j\in Adv,L} \int_{S_j^{min}}^{\infty} \frac{p_j\, E(Q_{ij} \mid S_{ij}) f_j(S) dS}{2 - F_H(S_H^{min}) - F_L(S_L^{min})} > \sum_{j\in Adv,L} \int_{S_{all}^{min}}^{\infty} \frac{p_j\, E(Q_i \mid S_i) f_j(S) dS}{2 - F_H(S_{all}^{min}) - F_L(S_{all}^{min})}$$

$$\int_{S_H^{min}}^{S_{all}^{min}} \frac{E(Q_{i,Adv} \mid S_{i,Adv}) f_H(S) dS}{F_H(S_{all}^{min}) - F_H(S_H^{min})} + \int_{S_{all}^{min}}^{\infty} \frac{p_H\, [E(Q_{i,Adv} \mid S_{i,Adv}) - E(Q_i \mid S_i)] f_H(S) dS}{2 - F_H(S_H^{min}) - F_L(S_L^{min})}$$

$$> \int_{S_L^{min}}^{\infty} \frac{p_L\, [E(Q_i \mid S_i) - E(Q_{i,L} \mid S_{i,L})] f_L(S) dS}{2 - F_H(S_{all}^{min}) - F_L(S_{all}^{min})} + \int_{S_{all}^{min}}^{S_L^{min}} \frac{E(Q_i \mid S_i) f_L(S) dS}{F_L(S_L^{min}) - F_L(S_{all}^{min})}$$

$$\therefore E(\Pi \mid S_{j\in Adv,L}^{min}) > E(\Pi \mid S_{all}^{min})$$

Web Appendix B: Sensitivity Analysis of Distributional Assumptions

To examine whether the analytical model is robust to different specifications of distributional assumptions, we conducted an analysis of a set of numerical simulations of the analytical model. Although we assume that Q_{ij}, the true quality of a consumer seeking service, is normally distributed in the original model (consistent with prior models on statistical discrimination), other prior empirical research has shown that the distribution of customer revenue, one potential realization of consumer quality, can be right-skewed (Fader et al. 2005; Schmittlein and Peterson 1994). These findings motivate the testing of a lognormal distributional assumption for Q_{ij}, an assumption consistent with Haenlein and Libai (2013). We also ran a second set of simulations using an uninformative distributional assumption – the uniform distribution. We constructed both alternative distributions to have the same mean (A_j) and standard deviation ($\sigma^2_{q_j}$) as the baseline normal distribution to which we compare. Table 3 displays a comparison of the results of these simulations to those of the original model. As indicated in the table, all three models are qualitatively consistent with each other, which suggests that the analytical model is robust to other distributional assumptions. Furthermore, all three models show statistically significant effects on the Group-Aware Bank dummy, the customer heterogeneity, and the measurement error variables. The direction of their signs on these effects are consistent with the expectations proposed by Propositions 1, 2, and 3 in the analytical model.

[Insert Table 3 about here.]

Web Appendix C: Minimum Service Criterion State Space

The three charts in Figure 3 provide examples of the relationship between S_L and S_H in the state space.

[Insert Figure 3 about here.]

On the x-axis is a range of possible Q^{min} values. The y-axis represents values for γ_H. Each chart is distinguished by low (0.24), medium (0.50), and high (0.76) values of γ_L for top, mid, and bottom charts respectively. The dark blue regions in the charts represent conditions where the service provider would set a higher criterion for the L group than the H group. The white regions represent conditions where the service provider would set the reverse: a higher minimum score criterion for the H group. The black curves on the boundaries between the dark and white regions represent values where the groups have equal minimum service criteria. The light dashed vertical lines mark the values of $\hat{A}_L = 640$ and $\hat{A}_H = 723$ in these examples.

Note that $S_H < S_L$ for all γ_j when $\hat{A}_L < Q^{min} < \hat{A}_H$. The intuition for this region is that a loan to the average L consumer is unprofitable while a loan to the average H customer is profitable, hence it stands to reason that the minimum score criteria for the L consumer is always higher. However, $S_L < S_H$ only if $Q^{min} < \hat{A}_L$ or $Q^{min} > \hat{A}_H$ and there is a sufficient difference in reliability ratios between the groups. In these regions, consumers from both groups are on average profitable to the bank or are unprofitable to the bank ($Q^{min} > \hat{A}_H$).

Web Appendix D: Sensitivity Analysis of Network Structure Assumptions

To examine whether the agent-based model (ABM) is robust to different specifications of social networks, we conducted an analysis of a set of numerical simulations of the ABM with alternative social network structures. Although we assume in our primary model that the social network of the applicants has the structure of a complete network, other network structures may produce results that differ in outcome. For this reason, we test robustness of the ABM by running a set of simulations with an Erdös - Rényi random network (Erdös and Rényi 1959) and a Barabasi-Albert preferential attachment network (Barabási and Albert 1999). We selected these networks structures because of their wide use in graph theory, social network analysis, and marketing literatures (Wilensky and Rand 2015). Table 4 presents results from this analysis. As indicated in the table, the ABM is generally robust to network structure specification. All three models are qualitatively consistent with each other. Furthermore, all three models show statistically significant effects with expected signs on the WOM parameter β, customer heterogeneity for the L-group variable, the measurement error variable, and parameter (g) for the growth rate in the mean quality of the L-group. This analysis suggests that different specifications for network structure produced the consistent outcome that as the weight of WOM increases in the utility function of the consumer, the greater the long-run profits of the Group-Blind exceeds that of the Group-Aware banks.

[Insert Table 4 about here.]

Web Appendix E: ABM Rules of Engagement

The following is a detailed description of the setup of the agent-based model which was developed and implemented in NetLogo (Wilensky 1999):

A Group-Blind bank has a loan strategy that does not use demographic group membership information. Instead, it uses the historical performance of all customers collectively as well as information about the individual applicant in its decision of whether or not to offer a loan.

A Group-Aware bank has a loan strategy that does use demographic group membership information. It uses the historical performance of all customers collectively by group as well as information about the individual applicant, conditional on group membership, in its decision of whether or not to offer a loan.

Customers who are interested in applying for a loan have a strategy where they will choose the bank that maximizes the customer's utility function. Their choice strategy, modeled by a multinomial logit choice function, has a utility that includes the likelihood a bank accepts their application, the distance between the customer's home and the bank, and unobservables captured by an error term. They assess the probability of acceptance through two information sources: advertising and word-of-mouth (WOM). The customer's personal assessment of the probability her application will be accepted depends on the information she gathers via WOM about acceptance rates of each bank. She differentially weights WOM information based on the characteristics of its source (strong vs. weak ties, in-group vs. out-group sources). Literature has demonstrated that people place greater weight on WOM from strong ties versus weak ties (Goldenberg, Libai, and Muller 2001a; Granovetter 1973, 1983). Literature has also shown that people place greater weight on in-group sources of WOM than out-group sources (Podoshen 2006; Uslu, Durmuş, and Taşdemir 2013). If the customer decides to apply for a loan, she selects the option that maximizes her utility. Her choice set is either one of four banks in her world or the outside option (no loan application).

Timeline of events during setup of the world in the ABM simulation (this happens once in simulation)

1. Two Colorblind Banks and two Full Information Banks are randomly placed in geographic locations in the ABM world
2. The banks are endowed with rules for their respective strategies to assess loan applicants (see next list for more detail)
3. A large population of people are randomly distributed throughout geography of world. They are randomly "born" to be either green or red people.
4. Each person is endowed with randomly provided characteristics (e.g., credit score, group membership) as well as rules governing the strategy for applicant's choice selection of a bank to apply for a loan

Timeline of events within one period in the ABM simulation

1. A randomly selected percentage of the population decides to apply for a bank loan.
2. These applicants select from a choice set of 4 possible banks to apply for the loan based on utility maximization
    (a) Applicant gathers WOM information about each bank in choice set from strong and weak ties. Within each category of ties, applicants also use in-group and out-group WOM information. WOM is regarding the percentage of past ties whose loan was accepted by each bank.
    (b) Applicant computes her own likelihood of her loan being accepted at each bank in choice set based on gathered WOM. Strong ties and in-group WOM gets greater weight than weak ties and out-group WOM.
    (c) Applicant computes her own utility for each bank in choice set, which is based on loan acceptance probability, distance to bank, and error.
    (d) Applicant selects bank based on multinomial logit choice model.

3. The banks review each loan application and forms expectation of applicant quality (interpreted as ability to repay loan) based on its own policy

   (a) Colorblind Banks' policy is to form expectation based on credit score of applicant and historical scores of all past applicants.

   (b) Full Information Banks have policy of forming expectation based on credit score of applicant and historical scores of all past applicants from applicant's group.

4. Banks offer loan to applicant if the expected quality of applicant meets or exceeds banks minimum threshold of quality (Q_{min}). All other applicants are rejected. NOTE: All banks have same minimum threshold.

5. Banks update historical information with applicant's information.

   (a) Colorblind Banks update historical information on scores of all past applicants

      i. Update historical mean of scores S

      ii. Update minimum accepted score S_{min}.

   (b) Full Information Banks update historical information on scores of all past applicants by group membership

      i. Update historical mean of scores S_j by group

      ii. Update minimum accepted score S_{min_j} by group.

6. All applicants who have applied for loan update their own historical information about banks (success/no success at applying for loan at each bank).

   (a) If applicant is rejected by a bank, she updates her probability of loan acceptance by that bank to 0. This remains in effect indefinitely.

   (b) If applicant is accepted by a bank, she updates her probability of loan acceptance by that bank to 1. This remains in effect indefinitely.

(c) For any bank where the applicant has never applied for a loan, the applicant will update probability of loan acceptance based on gathered WOM information in the next period.

7. Simulation clock proceeds to next period. Entire process starts again.

Web Appendix F: ABM Analysis of Monopolistic vs. Competitive Market

The ABM results provide additional insight into the dynamics of discrimination in service. First we conducted an analysis of the ABM simulating a Group-Aware and separately a Group-Blind bank monopoly in a city where consumers are not influenced by WOM. We found that the regression results, displayed in Table 5, are consistent with the results of the analytical model which provided assurance that the ABM and the analytical model are aligned. The first regression has realized bank profits as the DV. We control for consumer heterogeneity, measurement error, population mix of the two consumer groups (*Perc_H_Group* represents the percentage of the population that are H-group members) and for change in mean quality of the L-group (*Changing_Quality_L*). The regression shows the expected positive and significant effect on the Group-Aware dummy variable. This indicates that the Group-Aware monopoly is more profitable than the Group-Blind alternative in the monopoly-without-WOM scenario, consistent with Proposition 2. The DV for the second regression is D_{it}, the discrimination measure we have established in this paper. Regressing on the IVs for the Group-Aware bank in the simulation (because the Group-Blind bank has no discrimination, by definition), we find that customer heterogeneity and measurement error have statistically significant results with the expected signs. The ABM is again consistent with the results of the analytical model in this regard and supports Proposition 3.

[Insert Table 5 about here.]

The next table displays results from the analysis of ABM scenarios that added in competition and subsequently WOM as factors. We ran two types of competitive scenarios in the ABM: 1) two competing banks with the same service policy; 2) two competing banks with alternative service polices. We did this to be able to separate the impact of competition from the impact of competing service policies. Furthermore, we ran each of these scenarios with and without consumer WOM present. This enabled us to understand the impact of WOM on demand and profits. Table 6 displays these results.

[Insert Table 6 about here.]

In the table, *GAvsGB* is a dummy variable with a value of 1 when there is an ABM scenario where Bank 1 is a Group-Aware bank (GA) that competes against Bank 2, a Group-Blind bank (GB). Analogously, *GBvsGB* is a dummy variable representing a scenario where Banks 1 and 2 are two competing Group-Blind banks. The scenario not shown, *GAvsGA*, is the baseline for the regression. WOM is a dummy variable with a value of 1 if consumer WOM is present in the ABM simulation. All other main variables are the same as described in Table 5 as well as key interactions. The DV is the difference between Bank 2 and Bank 1 profits (i.e., Regression DV = Bank 2 profits - Bank 1 profits). The regression indicates that there are no significant differences in profits between the competing banks in each scenario when WOM is absent. This is interesting to note because the profit advantage that the Group-Aware bank has in a monopolistic scenario goes away when competition enters the picture. Furthermore, there are no significant differences in profit of competing banks when WOM is present and the banks use the same type of service policy. However, the interaction WOM : *GAvsGB* has a positive and significant effect. This indicates that when a Group-Aware bank competes against a Group-Blind bank in the presence of consumer WOM, the Group-Blind bank's profit is greater.

Table 3: Comparison of ABM Results on Distributional Assumptions for Q_{ij}

	Dependent variable:		
	Bank Profits		
	Normal	Lognormal	Uniform
Intercept	$1,448,713.00^{***}$	$1,515,150.00^{***}$	$1,498,005.00^{***}$
	(95,161.48)	(94,659.96)	(95,397.08)
Group-Aware Bank	$23,860.90^{**}$	$27,579.05^{***}$	$32,829.97^{***}$
	(9,480.61)	(9,430.65)	(9,504.08)
Heterogeneity_H	$2,352.93^{**}$	1,387.14	531.39
	(1,026.74)	(1,021.33)	(1,029.28)
Measure_Error_H	446.39	123.63	−711.72
	(1,026.74)	(1,021.33)	(1,029.28)
Heterogeneity_L	$13,143.94^{***}$	$12,261.85^{***}$	$15,066.44^{***}$
	(1,026.74)	(1,021.33)	(1,029.28)
Measure_Error_L	37.53	71.98	901.29
	(1,026.74)	(1,021.33)	(1,029.28)
Changing_Quality_L	$2,973,439,979.00^{***}$	$2,968,743,662.00^{***}$	$2,879,528,935.00^{***}$
	(23,293,881.00)	(23,171,118.00)	(23,351,550.00)
Percent_H_Group	$3,400,573.00^{***}$	$3,412,248.00^{***}$	$3,319,099.00^{***}$
	(20,599.21)	(20,490.65)	(20,650.20)
Heterogeneity_H:Measure_Error_H	−17.28	−10.69	−0.81
	(15.75)	(15.66)	(15.79)
Heterogeneity_L:Measure_Error_L	-52.01^{***}	-51.02^{***}	-65.79^{***}
	(15.75)	(15.66)	(15.79)
Observations	2,880	2,880	2,880
R^2	0.94	0.94	0.94
Adjusted R^2	0.94	0.94	0.94
Residual Std. Error (df = 2870)	254,391.50	253,050.80	255,021.20
F Statistic (df = 9; 2870)	$5,004.24^{***}$	$5,046.72^{***}$	$4,756.93^{***}$

Note: ABMs simulated a monopoly bank scenario. *p<0.1; **p<0.05; ***p<0.01

Table 4: Comparison of ABM Results on Network Structure Assumptions

	Dependent variable:		
	(Group Blind - Group Aware Profits)		
	Complete	ER Random	BA Pref. Attachment
Intercept	$932,950.10^{*}$	$136,166.00$	$221,947.90$
	$(505,579.50)$	$(104,920.00)$	$(136,052.30)$
WOM α	$-27,097.74$	$-5,944.26$	737.57
	$(25,045.32)$	$(5,197.51)$	$(6,739.74)$
WOM β	$36,697.86^{***}$	$9,244.79^{***}$	$2,399.73^{***}$
	$(2,782.81)$	(577.50)	(748.86)
Heterogeneity_H	$-6,017.28$	$1,602.35$	776.70
	$(5,424.76)$	$(1,125.77)$	$(1,459.81)$
Measure_Error_H	$13,437.34^{**}$	$6,432.54^{***}$	$7,327.35^{***}$
	$(5,424.76)$	$(1,125.77)$	$(1,459.81)$
Heterogeneity_L	$-19,155.21^{***}$	$-5,206.58^{***}$	$-5,473.65^{***}$
	$(5,424.76)$	$(1,125.77)$	$(1,459.81)$
Measure_Error_L	$-24,294.25^{***}$	$-9,021.49^{***}$	$-9,618.35^{***}$
	$(5,424.76)$	$(1,125.77)$	$(1,459.81)$
Changing_Quality_L	$257,622,564.00^{**}$	$123,079,027.00^{***}$	$128,302,904.00^{***}$
	$(123,072,816.00)$	$(25,540,595.00)$	$(33,119,112.00)$
Percent_H_Group	$511,214.20^{***}$	$120,462.90^{***}$	$5,453.29$
	$(108,835.60)$	$(22,586.02)$	$(29,287.84)$
Heterogeneity_H:Measure_Error_H	15.50	-44.02^{**}	-50.88^{**}
	(83.20)	(17.27)	(22.39)
Heterogeneity_L:Measure_Error_L	247.67^{***}	84.57^{***}	94.46^{***}
	(83.20)	(17.27)	(22.39)
Observations	5,760	5,760	5,760
R^2	0.06	0.10	0.05
Adjusted R^2	0.06	0.10	0.05
Residual Std. Error (df = 5749)	1,900,806.00	394,463.40	511,510.20
F Statistic (df = 10; 5749)	36.50^{***}	65.78^{***}	29.14^{***}

Note: ABMs simulated a competitive scenario. *p<0.1; **p<0.05; ***p<0.01

Table 5: ABM Analysis Results: Monopoly Bank with no WOM

	Dependent variable:	
	Realized Profits	Avg. Discrimination (D_{it})
	(1)	(2)
Intercept	1,448,713.00***	41.09***
	(95,161.48)	(1.80)
Group-Aware Bank	23,860.90**	
	(9,480.61)	
Heterogeneity_H	2,352.93**	−0.28***
	(1,026.74)	(0.02)
Measure_Error_H	446.39	0.54***
	(1,026.74)	(0.02)
Heterogeneity_L	13,143.94***	−0.26***
	(1,026.74)	(0.02)
Measure_Error_L	37.53	0.0004
	(1,026.74)	(0.02)
Changing_Quality_L	2,973,439,979.00***	−25,884.63***
	(23,293,881.00)	(441.36)
Perc_H_Group	3,400,573.00***	−0.12
	(20,599.21)	(0.39)
Heterogeneity_H:Measure_Error_H	−17.28	−0.00000
	(15.75)	(0.0003)
Heterogeneity_L:Measure_Error_L	−52.01***	0.00001
	(15.75)	(0.0003)
Observations	2,880	1,440
R^2	0.94	0.93
Adjusted R^2	0.94	0.93
Residual Std. Error	254,391.50 (df = 2870)	3.41 (df = 1431)
F Statistic	5,004.24*** (df = 9; 2870)	2,447.60*** (df = 8; 1431)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 6: ABM Analysis Results: Bank Competition and Consumer WOM

	Dependent variable:
	(Bank 2 - Bank 1) Realized Profits
Intercept	$-70,823.26$
	$(79,392.65)$
WOM	$28,649.14^{*}$
	$(16,855.12)$
GAvsGB	$-12,843.96$
	$(21,320.23)$
GBvsGB	982.23
	$(21,320.23)$
Changing_Quality_L	$21,428,042.00$
	$(19,127,881.00)$
Perc_H_Group	$165,882.90^{***}$
	$(16,915.14)$
Heterogeneity_H	$2,335.80^{***}$
	(843.11)
Measure_Error_H	$5,709.48^{***}$
	(843.11)
Heterogeneity_L	$-3,022.89^{***}$
	(843.11)
Measure_Error_L	$-3,389.76^{***}$
	(843.11)
WOM:GAvsGB	$51,983.20^{**}$
	$(23,836.74)$
WOM:GBvsGB	$33,548.02$
	$(23,836.74)$
Heterogeneity_H:Measure_Error_H	-61.11^{***}
	(12.93)
Heterogeneity_L:Measure_Error_L	35.19^{***}
	(12.93)
Observations	21,600
R^2	0.02
Adjusted R^2	0.01
Residual Std. Error	572,081.80 (df = 21586)
F Statistic	25.64^{***} (df = 13; 21586)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Figure 3: Minimum Service Criteria State Space