Algorithmic Bias in Service

KALINDA UKANWA
ROLAND T. RUST

November 22, 2021

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@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 departments of University of Maryland Smith School of Business, Wharton University of Pennsylvania, Duke University, University of Colorado, Rutgers University, Harvard Business School Crossing Disciplines Workshop, and attendees of the PhD Project MDSA Conference for helpful comments. The authors also wish to thank Dr. Renana Peres and Dr. Eitan Muller for their help in securing social network data. The authors thank Zachary Cummings for research assistance on this study.

Algorithmic Bias in Service

Abstract

Research shows that algorithms using sociodemographic data (e.g., race, gender, education, etc.) can produce biased outcomes that cause many consumers to be excluded from or endure lower levels of service. Though research suggests that these algorithms are more profitable than unbiased algorithms that do not use sociodemographic data, prior findings do not consider potential social effects of these algorithms on consumer demand. This research investigates the dynamic outcomes of competition between biased and unbiased algorithms in a market where word-of-mouth influences consumer choice behavior. Relative to unbiased algorithms, this research demonstrates that biased algorithms can be more profitable in the short run but less profitable in the long run, due to consumer word-of-mouth. Models and simulations show that word-of-mouth leads marginalized consumers to gravitate towards easier-to-access unbiased algorithmic services. Non-marginalized consumers, on the other hand, learn they have a relatively easier time accessing services anywhere. When sufficient numbers of marginalized and non-marginalized consumers learn from each other via word-of-mouth, long run demand is greater for unbiased algorithmic services. This research demonstrates that firms that use unbiased algorithms and account for social effects (e.g., word-of-mouth) in the algorithm's design can reduce algorithmic bias while improving both long-term profits and societal well-being.

Keywords: algorithms, algorithmic bias, algorithmic fairness, discrimination, word of mouth, agent-based modeling

Algorithmic Bias in Service

In November 2019, tech entrepreneurs David Heinemeier Hansson and Steve Wozniak posted a series of accusations on Twitter that Apple Card's "black box algorithm" discriminated against women. They claimed that even though they and their wives shared the same financial histories, the algorithm had granted them favorable credit terms that were denied to their wives. At the time, Hansson and Wozniak had over one million followers combined. The tweets went viral, generating substantial word-of-mouth (WOM) and media coverage. As of May 2021, their viral posts have been liked over 31,000 times, retweeted over 13,000 times, and commented on more than 1,400 times (Heinemeier Hansson and Wozniak 2019).

Apple is not alone in facing WOM challenges due to algorithmic bias, defined as systematically unfair outcomes of an algorithm which arbitrarily disadvantages some sociodemographic groups relative to others (Barocas and Selbst 2016; Friedman and Nissenbaum 1996; Lambrecht and Tucker 2019; O'Neil 2016). In 2020, the organizations behind the International Baccalaureate (IB) and United Kingdom's A-level standardized tests faced WOM challenges which led to protests. Their testing algorithms appeared to advantage students from wealthier communities over students from poorer communities (Adam 2020; Simonite 2020). Also in 2020, Black TikTok creators generated WOM about their experiences of bias from TikTok's recommendation and content monitoring algorithms. Consequently, many Black TikTok creators are now planning to leave TikTok for perceived fairer platforms such as Fanbase and Clapper (Contreras and Martinez 2021).

Biased algorithmic decisions can have significant repercussions for consumers. Consider the case of high school senior Isabel Castañeda, one of the top-ranking students at her school. Isabel received a spot at Colorado State University, conditional on a sufficient score on the IB exam. The IB exam also held the promise of saving Isabel thousands of dollars in tuition via early college credits. Because the IB program canceled exams due to the coronavirus pandemic, Isabel was unable to take the IB exam in-person. Instead, received an exam score based on a prediction by

the IB progam's algorithm. The algorithm used Isabel's high school course performance, her teacher's prediction, and historical data on past performance of students from her school. Despite being a native Spanish speaker, earning top grade in her Spanish classes, and her teacher's positive predictions, Isabel was shocked to receive a failing score on the IB Spanish test. "I come from a low-income family - and my entire last two years were driven by the goal of getting as many college credits as I could [through the IB test] to save money on school. When I saw those scores, my heart sank." Despite indicators of her own merits, the IB algorithm assigned Isabel a lower score than expected because her school, which was predominantly minority and lower-income, had a history of lower scores on the exam (Asher-Schapiro 2020).

Suppose firms like Apple, TikTok, and International Baccalaureate had no prejudiced intent when they deployed their algorithms. In fact, many organizations claim that algorithms help them to eliminate bias. Yet, algorithmic bias research has shown that algorithms can discriminate and produce unfair outcomes (Buolamwini and Gebru 2018; Lambrecht and Tucker 2019; Obermeyer et al. 2019). Furthermore, algorithms can embed or reinforce existing systemic discrimination in meso- and macro-level organizational and societal structures which influence to whom the flow of critical resources like higher education go. Although often unintentional, this creates systematic disparate impact on marginalized groups (Crockett 2021). Nevertheless, prior research on algorithmic decision-making and statistical discrimination suggests that algorithms using sociodemographic information or its proxies (e.g., race, gender, education-level, etc.) are more accurate and optimal than those without (Fu, Huang, and Singh 2021; Kleinberg, Ludwig, Mullainathan, and Rambachan 2018; Zhang, Mehta, Singh, and Srinivasan 2021). However, the cases of the Apple Card, TikTok, the International Baccalaureate, and UK A-level standardized tests show that algorithmic bias can generate WOM.

A substantial amount of research in WOM and electronic (digital or online) WOM has shown that WOM influences consumer choices, thereby impacting demand, profits, and the optimality of the algorithms over time (Brown and Reingen 1987; Chevalier and Mayzlin 2006; Gopinath, Thomas, and Krishnamurthi 2014; Herr, Kardes, and Kim 1991). We argue that social

effects are critical to the impact of algorithmic bias on consumer demand because social learning helps potential consumers reduce risk and allocate effort when applying for services that pre-screen consumers and may reject them. Prior research has demonstrated that social effects contribute substantially to consumer demand (Bass 1969; Iyengar et al. 2011). Yet, social effects are often insufficiently accounted for in consumer demand models (Hogan et al. 2003). To the best of our knowledge, there is no research on the potential combined effects of algorithmic bias and consumer WOM on subsequent consumer demand over time. Our research seeks to fill this gap. In contrast to prior research, this research also demonstrates conditions under which algorithms that use sociodemographic information can be *less* optimal and profitable than algorithms that do not.

To investigate the impact of algorithmic bias on consumer WOM, we take an approach similar to Watts and Dodds (2007). We model and simulate the social interactions of consumers in a market where algorithms that use sociodemographic information compete against algorithms that do not. Calibrated with empirical data, this two-sided model of algorithmic service provision (supply-side) and WOM-driven consumer demand (demand-side) enables us to examine the dynamics of algorithmic bias over time in an endogenous system. This research finds that when non-marginalized and marginalized consumers do not generate WOM, firms employing algorithms that use sociodemographic group data are more profitable. This is consistent with prior research findings. However, algorithms using sociodemographic data can activate WOM that facilitates consumer social learning. When marginalized and non-marginalized groups communicate and influence each other via WOM, it can shift the composition of demand and can reverse profitability outcomes. Results show that within five years, firms employing algorithms that use sociodemographic group data attract *less* demand than firms employing algorithms that do not use sociodemographic information. At the rate of customer loss that we observe in our study, TikTok, which currently has over one billion monthly visitors (TikTok 2021), would lose more than 70 million visitors each month.

Our findings have implications for consumers because we theoretically show conditions

where algorithmic bias could influence consumers' consumption choices via their social networks. Furthermore, word-of-mouth is a collective consumer power that facilitates social learning and can help protect consumers from the potential harm of algorithmic bias. For example, algorithmic bias-generated WOM about the Apple Card attracted the attention of New York government regulators, who launched an investigation into whether the algorithm was discriminatory (Vigdor 2019). This research also has implications for public policy makers and firms. Findings could influence active legislation being proposed in the US and Europe on regulations regarding the monitoring of algorithmic bias. For firms, employing algorithms that use group information can be attractive in the short run but can backfire in the long run.

Our study answers recent calls for more research into systemic and structural sources of marketplace discrimination (Arsel, Crockett, and Scott 2021; Bradford and Perry 2021; Bruce, Cutright, Gosline, Thomas, and White 2020; Crockett 2021; Johnson, Thomas, Harrison, and Grier 2019; Poole, Grier, Thomas, Sobande, Ekpo, Torres, Addington, Weekes-Laidlow, and Henderson 2021) as well as the societal harms and ethics of technology and AI (Ekpo, DeBerry-Spence, Henderson, and Cherian 2018; Thomaz, Efremova, Mazzi, Clark, Macdonald, Hadi, Bell, and Stephen 2021). Advances in artificial intelligence, data science, and analytics, in conjunction with world-wide focus on social movements regarding discrimination (e.g., MeToo and International Women's Day, Black Lives Matter, Trans Lives Matter, Ninety-nine percenter, etc.) motivate a need for research on the impact of algorithmic bias on consumers (Anderson and Ostrom 2015; Hill and Stephens 2003). In other words, consumer advocacy, which focuses on mitigating the factors that harm consumers, should also be an important area of consumer research. Social fissures created by algorithmic bias could have a direct impact on consumer and societal well-being (Bone, Christensen, and Williams 2014; Crockett, Grier, and Williams 2003). We elaborate on these themes in the remainder of the article. We begin with an overview of our contribution to the literature. We next describe our conceptual framework and propose a definition of algorithmic bias. Then, we discuss our model and findings of the long-term impact of algorithmic bias and its interaction with competition and WOM. Finally, we conclude with a

discussion of the managerial, consumer, and policy implications of our findings.

MARKETPLACE ALGORITHMIC BIAS, DISCRIMINATION, AND DIFFERENTIAL SERVICE TREATMENT

This research sits at the intersection of algorithmic bias, marketplace discrimination, and differential service treatment. Although algorithmic bias research is a relatively new domain, there is a rich body of work on computational strategies to reduce or identify algorithmic bias, measures of algorithmic bias, and algorithm governance and accountability in terms of fairness. Given that our study focuses on the effect of algorithmic bias on social interactions, the preceding research areas are mostly outside the scope of our study. Corbett-Davies and Goel (2018), Khalil, Ahmed, Khattak, and Al-Qirim (2020), and Wieringa (2020) provide reviews of this literature. Similarly, extant research in marketplace bias and differential service treatment that do not focus on dynamic social effects of discrimination, bias, or differential service treatment are outside the scope of this study. Arsel, Crockett, and Scott (2021), Henderson, Hakstian, and Williams (2016), Johnson, Thomas, Harrison, and Grier (2019), Pager and Shepherd (2008), and Fang and Moro (2011) provide great overviews of this literature.

Our main intended contribution is to show the dynamic social effects of algorithmic bias on WOM and consumer demand. Table 1 summarizes how our study differs from papers we position this research against. For each article included, we describe the study's context, key findings, and whether the study examined 1) algorithms, 2) discrimination/bias, 3) consumer WOM, 4) dynamics, and 5) consumer demand. We highlight algorithms because algorithmic decision-making is fundamentally different from human decision-making. We highlight discrimination and bias because some forms of differential service treatment are based on willingness-to-pay or profitability differences that are independent from sociodemographic group membership. We highlight consumer WOM because, unlike consumer responses that impact only the self, this form of consumer response to algorithmic bias or differential service treatment has a

social element that invariably affects other consumers via social learning. We highlight dynamics because many of the papers show static effects but not effects over time. Finally, we highlight consumer demand because our research examines the macro implications of algorithmic bias's effects on consumer WOM. This is novel to the literature.

Algorithmic bias in marketplace contexts is a special type of marketplace discrimination–the differential treatment of consumers in the marketplace based on group membership (Crockett et al. 2003; Ekpo et al. 2018; Johnson et al. 2019). Because there is relatively little research on the behavioral and social implications of algorithmic bias, there is a call for more social science research in this area (Kordzadeh and Ghasemaghaei 2021), and in particular for algorithmic bias in the marketplace. Notable algorithmic bias papers that show impact on consumers in marketplace contexts include Sweeney (2013), Noble (2018), and Lambrecht and Tucker (2019), who show that algorithmic bias can diminish the quantity or quality of online ads and search results for female and Black consumers. Obermeyer et al. (2019) show that a biased medical algorithm compromised the health of sick Black patients. Srinivasan and Sarial-Abi (2021) show that consumers respond less negatively to a brand if its algorithmic errors (which include bias) are caused by an algorithm rather than a human. Zhang, Mehta, Singh, and Srinivasan (2021) demonstrate that AirBnB's pricing algorithm generates prices that were more suboptimal for Black than White hosts, yet narrowed the revenue gap between them. Fu et al. (2021) find that even though an investment algorithm produces more accurate predictions than human investors, it also produces biased gender- and race-based outcomes. Our research differs from these papers because we demonstrate that dynamic algorithmic bias can impact not only the immediate focal consumer but can also indirectly impact other consumers via social interactions (WOM) over time.

The marketplace discrimination literature intersects with multiple fields of research, including consumer behavioral research, transformative consumer and service research (Anderson et al. 2010), sociological research, and the economics of discrimination research (Fang and Moro 2011). Much of that body of work across these fields focuses on consumer discrimination

experiences, interactions,and processes at a micro (individual transaction) level rather than on a system-wide, structural level in the marketplace. Furthermore, the research focuses on human-driven discrimination. In contrast, we examine the impact of a non-human source of discrimination–algorithms–that impacts large populations of consumers from meso- and macro-level systems and processes. Meso-level (organizational processes) and macro-level (societal norms) sources of marketplace discrimination is under-researched in the literature (Crockett 2021; Poole et al. 2021). Given the breadth of prior related research (although there is a relative dearth of research in marketing), we cover notable papers most relevant to our study. Research into the impact of human service providers discriminating against consumers has shown that discrimination exacts high psychological costs on the consumer (Crockett et al. 2003), contributes to self-concept harm, and restricts the consumer sense of agency (Bone, Christensen, and Williams 2014). Furthermore, marketplace discrimination diminishes the level of service to the consumer (Harris, Henderson, and Williams 2005). Consumers from marginalized (non-marginalized) groups are more likely to have a system-challenging (system-supporting) response to the discriminatory service (Evett, Hakstian, Williams, and Henderson 2013). Although a preponderance of economics of discrimination research examines bias in firm employment decisions, ”taste-based” and statistical discrimination theories have been applied to marketplace settings as well. The ”taste-based” theory of discrimination assumes that discriminatory firms intend to discriminate because of a disutility for marginalized consumers, which is not necessarily profit-maximizing (Becker 1957). In contrast, statistical discrimination theory assumes that firms are rational, profit-maximizing actors who do not intentionally discriminate. Firms use sociodemographic group information to reduce uncertainty, and hence statistical discrimination is profit-maximizing. Discrimination that arises is a byproduct of statistical error (Arrow 1973; Phelps 1972).

The aforementioned literature addresses the direct impact of discrimination on consumers who interact with the discriminating firm, but not the additional social effects on other consumers who have not interacted with the firm. Furthermore, there is very little research that shows the

dynamic impact of marketplace discrimination. One exception is Bjerk (2008), who shows that discrimination can emerge if there are differences between sociodemographic groups in the frequency and precision of signaling quality level. Other exceptions are Blume (2006) and Fryer (2007), who show that dynamics in learning about the quality of people can influence beliefs about marginalized consumers in the long run. Our research differs from these in two ways. First, we demonstrate social effects of WOM over time that result from algorithmic discrimination. Second, we demonstrate conditions where use of sociodemographic information is not profit-maximizing, but in fact is less profitable than not using sociodemographic information.

In the differential service treatment literature, Rust, Zeithaml, and Lemon (2000) and Homburg, Droll, and Totzek (2008) argue that prioritizing selected groups of customers can produce profitable and positive effects on customer relationships. In contrast, Lepthien et al. (2017) and Haenlein and Kaplan (2012) show that demarketing to or abandoning unprofitable customers has negative effects on customer relationships via negative WOM. Our research differs from theirs in that they show static effects based on survey and experimental results. In contrast, we model the dynamic effects of WOM on consumer choice behavior over time. The study that our research is closest to is Hogan et al. (2003). They show the dynamic effects of consumer WOM when a consumer disadopts a product. We differ from their work in multiple significant ways. First, they examine dynamic social effects generated by a consumer who is already a customer to the firm and then decides to disadopt the firm's product. We examine dynamic social effects generated by consumers who apply to be a customer of the firm. Second, they assume WOM diffuses through a social network with some probability of influencing other consumers, but they do not provide theoretical insight into the mechanism that makes WOM influential. In contrast, our research provides a theoretical mechanism that explains how consumers evaluate WOM to assess their chances of being accepted to receive service. Third, Hogan et al. (2003) assumes the firm has human service providers and that the firm does not consider sociodemographic characteristics of the consumers in determination of their value to the firm. We examine non-human algorithmic service providers who used sociodemographic data to make

value determinations about prospective customers.

Although the central contribution of this study is to show the dynamic social effects of algorithmic bias on WOM and consumer demand, our model of consumer WOM is a contribution to the word-of-mouth (WOM) literature. Extant research on WOM has found that weak ties facilitate transportation of WOM between distinct groups. Meanwhile, strong and homophilous ties were more influential sources of information about goods and services (Brown and Reingen 1987; Granovetter 1973). Consumers engage in WOM for impression management, emotional regulation, social bonding, persuasion, or information acquisition purposes (Berger 2014). Consumers judge the moral hazards and social context of WOM they have received before passing it on to their social contacts (Frenzen and Nakamoto 1993). Early studies found that face-to-face WOM is more persuasive than print (Herr et al. 1991). However, the rise of internet and mobile technologies has increased the influence of e-WOM (electronic, social media, and other digital forms of WOM). Scholars have found that e-WOM is more effective than traditional marketing techniques and plays a significant role in shaping a wide-range of consumption decisions (Berger and Milkman 2012; Chevalier and Mayzlin 2006; Chintagunta et al. 2010; Godes and Mayzlin 2009; Trusov et al. 2009). Within the scope of this study, we focus on the consumer's use of WOM to acquire information about services. A common theme among this literature is that consumers in search of information use WOM to evaluate firms or their goods and services. Our paper differs from this literature in that we provide a theoretical model of how consumers use WOM to evaluate their chances of acceptance to receive service from a firm. In the next section, we present our conceptual framework about the mechanisms of algorithmic bias and WOM generated by it.

Table 1: Selected Research: Algorithmic Bias, Marketplace Discrimination, and Differential Service Treatment

Article	Context	Main Finding	Algorithms	Discrimination and Bias	Word-of-Mouth	Dynamics	Consumer Demand
Lambrecht and Tucker (2019)	Online Ad Delivery	Higher ad serving prices for women resulted in algorithmic ad delivery bias against women.	✓	✓			
Sweeney (2013)	Online Search Ads	Online search using Black-sounding (White-sounding) names are more (less) likely to produce online search ads associated with incarceration or arrest.	✓	✓			
Noble (2018)	Online Organic Search	Online organic search terms related to Black women or girls produces results from search engine algorithms with negative associations that are biased against Black females.	✓	✓			
Obermeyer, Powers, Vogeli, and Mullainathan (2019)	Healthcare	Lower healthcare spending rates among Black patients resulted in algorithmic healthcare bias against Black patients.	✓	✓			
Srinivasan and Sarial-Abi (2021)	Brands	Consumer assessment of the brand is less negative when an algorithmic (vs. human) error harms the brand.		✓			
Zhang et al. (2021)	Sharing Economy	AirBnB's Smart Pricing algorithm narrowed the revenue gap between Black and White hosts, although its output is more suboptimal for Black hosts.		✓			
Fu et al. (2021)	FinanceInvesting	ML algorithms outperform human investors on crowd lending site, but they exhibit race and gender bias.		✓			
Becker (1957)	Employment/Economic selection	Firms exhibiting discrimination in the marketplace have included a disutility for association with minority groups in their objective functions, even if it is not profit-maximizing.		✓			
Arrow (1973); Phelps (1972)	Employment/Economic selection	Firms exhibiting discrimination in the marketplace are profit maximizing, but produce discrimination due to statistical error under uncertainty.		✓			
Fryer (2007)	Employment/Economic selection	Firms initially discriminate against marginalized job candidates because of because of belief that members are of lower quality. However, those they do hire are believed to be of higher quality than non-marginalized counterparts.		✓		✓	
Bjerk (2008)	Employment/Economic selection	Equally skilled workers from different groups will have different likelihoods of making it to top jobs in the economy if groups differ in average skill level and the frequency and precision with which they signal skill level.		✓		✓	
Blume (2006)	Employment/Economic selection	Learning dynamics influence revisions of belief and outcomes of statistical discrimination.		✓		✓	
Crockett, Grier, and Williams (2003)	Retail	Marketplace discrimination against Black men extracts high psychogical costs. To cope, Black men often change behaviors in marketplace contexts.		✓			
Bone, Christensen, and Williams (2014)	Financial services	Financial service firm poor treatment of minority (vs. White) consumers contributes to self-concept harm and restrictions on agency and self-esteem for minority consumers.		✓			
Evett, Hakstian, Williams, and Henderson (2013)	Retail	A consumer's race and level of perceived societal discrimination influence response to marketplace discrimination. Black (vs. White) consumers are more likely to attribute marketplace discrimination to systemic (individual-level) sources and support a system-challenging (system-reproducing) response. High levels of perception of societal discrimination attenuated White responses.		✓			
Harris, Henderson, and Williams (2005)	Service	Analysis of court decisions regarding consumer allegations of racial/ethnic discrimination from services yields three dimensions of market discrimination: type of discrimination, service level, and existence of criminal suspicion.		✓			
Homburg, Droll, and Totzek (2008)	Customer Relationship Management	Customer prioritization ultimately leads to higher average customer profitability and a higher return on sales because it (1) affects relationships with top-tier customers positively but does not affect relationships with bottom-tier customers and (2) reduces marketing and sales costs.			✓		
Hogan, Lemon, and Libai (2003)	Service	The authors incorporate social effects in a profitability model to show that the value of a lost customer changes throughout the product life cycle. The loss of an early adopter costs the firm much more than the loss of a later adopter.			✓	✓	✓
Lepthien, Papies, Clement, and Melnyk (2017)	Customer Relationship Management	Consumers disapprove of customer demarketing, regardless of whether they experience it themselves or only observe it, regardless of the responsibility for the cause of the contract termination, and regardless of the social proximity to the dismissed customers.			✓		
Haenlein and Kaplan (2012)	Customer Relationship Management	Current customers are significantly more likely to respond actively to unprofitable customer abandonment (exit/voice) than passively through silence and loyalty. Additionally, it shows that increasing satisfaction or switching cost among current customers are unlikely to limit the potential negative consequences of unprofitable customer abandonment.			✓		
This Research	Service	Algorithmic bias can activate consumer WOM which shifts the composition of demand. When WOM is activated, biased algorithms are profitable in the short run but unprofitable in the long run compared to unbiased algorithms.	✓	✓	✓	✓	✓

CONCEPTUAL FRAMEWORK: HOW DOES ALGORITHMIC BIAS OCCUR AND ACTIVATE WORD-OF-MOUTH?

In our conceptual framework, we assume that consumer sociodemographic groups are defined by an observable attribute that all group members share (usually physical or socio-economic in nature). Examples include U.S. legally protected classes (e.g., race/ethnicity, gender, age, etc.) as well as unprotected classes (e.g., education-level, social class, residential location, etc.). Note that some defined groups can proxy for other groups. For example, research has shown that residential location, an unprotected class, can proxy for race/ethnicity, and even age, which are protected classes (Elliott et al. 2009). Consistent with the sociological literature and with US law, we distinguish between two types of discrimination or bias – disparate treatment and disparate impact. Disparate treatment is intentional discrimination against members of a group. It is often, but not always, driven by prejudice, stereotyping, bigotry, or racism – internally-held attitudes, beliefs, and ideologies. Disparate impact, on the other hand, is the unequal (and often unintentional) treatment of groups as an outcome of policies, rules, or decisions which appear non-prejudiced in design (Barocas and Selbst 2016; Pager and Shepherd 2008; Quillian 2006). Disparate impact can be independent of internally-held attitudes. Our study uses statistical discrimination assumptions (Phelps 1972) which presume that biased algorithms produce disparate impact. We believe this assumption represents a conservative lower-bound on the emergent effects of algorithmic bias. If the firm's algorithm is also driven by prejudicial intent or biased data collection, it would intensify and exacerbate the effects we observe.

This research applies to contexts that meet three criteria: 1) consumers can be segmented into sociodemographic groups; 2) the firm screens prospective consumers for the decision of whom to serve by using a classification algorithm (supervised learning algorithm that classifies new observations into categories or classes (Russell and Norvig 2020)); 3) the algorithm is trained on data about both the individual consumer and the consumer's group. For example, our

framework could apply to rental decisions, school admissions decisions, social media content acceptance decisions, or lending or credit card decisions. Note that our research examines only the decision of whether or not to offer service to a prospective consumer. We do not examine pricing of the service, which could incorporate pricing discrimination. For those interested in the topic of pricing discrimination, we refer readers to Bergemann et al. (2015), Narasimhan (1984), and Varian (1989) for excellent insights into this area. In the following section, we model how algorithmic bias can arise from an algorithmic decision, even though there is no intent to discriminate.

How Does Bias Arise from an Algorithm's Decision?

Algorithmic bias arises when consumers from different sociodemographic groups are systematically treated differently by an algorithm, even though they possess similar characteristics relevant to the algorithm's primary objective. For example, if Black and White content creators with similar content histories post the same content to TikTok, yet the TikTok algorithm systematically removes the Black creator's content while leaving the White creator's untouched, then this is algorithmic bias. Our definition of algorithmic bias is consistent with what is known as *individual fairness* in the algorithmic bias literature. Individual fairness is the concept that any two individuals who are similar with respect to the algorithm's objective should be treated similarly (Dwork et al. 2012). We now describe our model of an algorithm that uses sociodemographic data and how it can produce biased outcomes.

Imagine a population of consumers who are members of one and only one of two groups $j \in \{H, L\}$, an H-group (high advantaged type) or L-group (low advantage type) (e.g., upper- versus lower-income, men versus women, majorities versus minorities, etc.). Each consumer has a quality Q_{ij} that is an unobservable, latent attribute representing the consumer's value to a firm b. This could be thought of as the consumer's true profitability (value) to the for-profit (non-profit) firm, not the person's inherent worth as a human being. For simplicity of exposition, we will refer

to both for-profits and non-profits as firms, and consumer quality as their profitability to the firm. Group j members vary in quality levels that are normally distributed around a group mean, A_j and variance $\sigma^2_{q_j}$ (a measure of within-group consumer heterogeneity). H-group consumers have, on average, higher mean quality levels than those in L-group.

Each consumer generates information (data) from behavior that signals the consumer's quality (e.g., bill repayment history, standardized test performance history, content posting history, etc.). Consumer i's information, which is conditional on the consumer's quality, is captured in consumer i's score S_{ij}. The score, an error-prone measure of the consumer's latent quality, is assumed to be normally distributed with a mean Q_{ij} and variance σ^2_ε (a measure of error in measuring quality). For example, rental and financial services use credit scores to measure true credit-worthiness. Education institutions use standardized test scores like International Baccalaureate or A-1 Levels to measure a student's true likelihood of successfully completing college. Note that the score does not include information about the consumer's group membership. We assume this because 1) legal restrictions often prohibit considering information about protected groups, 2) we want to investigate the effect of considering that information, and 3) often group membership is separable from the individual's behavior.

The relationships between consumer variables A_j, Q_{ij}, and S_{ij} are summarized as follows:

$$Q_{ij} = A_j + \upsilon_{ij}, \quad \upsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{q_j}) \tag{1a}$$

$$S_{ij} \mid Q_{ij} = Q_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\varepsilon_j}) \tag{1b}$$

$$S_{ij} \sim \mathcal{N}(A_j, \sigma^2_{q_j} + \sigma^2_{\varepsilon_j}) \tag{1c}$$

$$\text{where } A_H > A_L > 0 \text{ and } \upsilon_{ij} \perp\!\!\!\perp \varepsilon_{ij} \tag{1d}$$

We assume that quality, score, and mean group quality do not change over time. Each group could have different levels of customer heterogeneity ($\sigma^2_{q_j}$), but measurement error (σ^2_ε) is the same for both groups and independent of the distribution of quality within groups. Thus, the errors for quality ($\sigma^2_{q_j}$) and score, conditional on quality (σ^2_ε), are independent. Importantly, relaxing this

assumption about independence of errors does not qualitatively change our results. Note that the two groups could overlap in quality and score attributes. It is possible that some H-group customers have the same quality and scores as some L-group customers, and some L-group customers may be higher quality than some in the H-group. Algorithmic bias arises if an algorithm treats individuals from the two groups differently, even when they have the same quality.

Firm b's expected and realized profit from each service exchange given to consumer i from group j is as follows:

$$E[\pi_{ij}] = E[Q_{ij} \mid S_{ij}] - Q^{min} \quad \text{expected profit} \tag{2a}$$

$$\pi_{ij} = Q_{ij} - Q^{min} \quad \text{realized profit} \tag{2b}$$

Imagine that firm b uses an algorithm to make decisions about whether or not to offer service to prospective consumer i (e.g.school admission, accept content, lend money, etc.). Because firm b is uncertain of consumer i's quality (Q_{ij}), the firm does not know whether service to consumer i will be profitable until after she consumes the service. Hence, firm b's objective is to offer service to consumers whose expected quality, conditional on her data (score) ($E[Q_{ij} \mid S_{ij}]$) exceeds Q^{min} (assumed to be exogenous). Q^{min} is the quality level where the firm makes 0 profit on the service offered.

The algorithm we model is based on dynamic Bayesian learning models of quality commonly used in the marketing literature (Boulding et al. 1993; Ching et al. 2013; Erdem and Keane 1996; Rust et al. 1999). Our model is also a dynamic version of a common statistical discrimination model (Aigner and Cain 1977; Fryer 2007; Phelps 1972). The expression for expectation of quality, conditional on score, for consumer i from group j at time t is

$$\begin{aligned} E_t[Q_{ij} \mid S_{ij}] &= \hat{\gamma}_{jt} S_{ij} + (1 - \hat{\gamma}_{jt})\, \hat{A}_{jt} \quad \text{, where} \\ \hat{\gamma}_{jt} &= \frac{\hat{\sigma}^2_{q_{jt}}}{\hat{\sigma}^2_{q_{jt}} + \sigma^2_{\varepsilon}} \end{aligned} \tag{3}$$

Expected quality, conditional on score, is a dynamic estimate and a weighted combination of data about the individual consumer (her score S_{ij}) and data about the consumer's group (estimate of group j's mean quality, $\hat{A}_{jt}$). Because this algorithm uses group information, we henceforth refer to this algorithm as the *Group-Aware* algorithm. Given that the consumer's score is a noisy, error-prone signal of quality, data about the consumer's group is used to provide additional information to improve the accuracy of the expected quality estimate. Given that the firm is also uncertain about group j's true mean and variance of quality, the firm's Group-Aware algorithm uses historical data (scores) on prospective consumers training data to dynamically estimate the mean ($\hat{A}_{jt}$) and variance ($\hat{\sigma}^2_{q_{jt}}$) of quality at time t. We also assume that the algorithm is not forward-looking and does not have access to data about prospective consumers competitors. This assumption has been shown to hold in some typical algorithmic service environments[1]. See Web Appendix D for details on how the Bayesian learning model updates estimates from data.

The weight in the algorithm, $\hat{\gamma}_{jt}$, is known as the score validity and is the proportion of the total variance that is comprised of group j's quality. Score validity has value between 0 and 1. The quantity $\hat{\gamma}_{jt}$ must be learned and estimated from the training data. The score validity indicates the validity of consumer i's score and designates the weight placed on score S_{ij}, as opposed to information about their group j, captured in $\hat{A}_{jt}$ (the prior). If there is no measurement error (i.e., $\sigma^2_\varepsilon = 0$), then $\gamma = 1$ and implies the consumer's score is completely valid and perfectly predicts the consumer's quality. On the other hand, if score error exists, then the consumer's score is not a completely valid measure. Hence, the Group-Aware algorithm places some weight on $\hat{A}_{jt}$, which represents the historical data about the group the consumer is from. Increasing customer heterogeneity ($\hat{\sigma}^2_{q_{jt}}$) or decreasing measurement error (σ^2_ε) increases the score validity. In other words, the more valid a score is, the more the algorithm assesses the consumer on individual merits rather than the consumer's group.

Group-Aware algorithms are very commonly used (e.g. regression, logistic regression, Naive Bayes, SVM algorithms, etc.) and often include both features (variables) that are specific to the

[1]Conversations with bank lending officers indicated that this assumption was accurate for them.

individual consumer (e.g. academic achievement history, bill payment history) as well as to the sociodemographic group the consumer is from (e.g., zipcode, citizen status). This means that the Group-Aware algorithm's estimate of an individual consumer's quality will always be a function of the consumer's own history of actions as well as the consumer's group. Two consumers from different groups who have the exact same score (similar individuals) will get different expectations from the algorithm if their groups differ in mean quality ($\hat{A}_{Ht} \neq \hat{A}_{Lt}$). For example, this may have been the case with the International Baccalaureate and UK A-1 Level algorithms that were used to predict student scores. Consider two students, one from an affluent neighborhood and one from a poor neighborhood, who produce the exact same performance on the exact same academic materials and practice standardized tests in high school. If the poorer student's school has historically had lower test scores than the affluent student's school, then it is straightforward to see that the poor student would have a lower predicted test score than the affluent student. Under individual fairness standards, this is where algorithmic bias arises. We formalize our definition of algorithmic bias (D_{it}) as follows:

Algorithmic bias is the difference in expectations of two consumers who have the same score but are members of different groups. Equivalently, algorithmic bias is is the difference in expectations of consumer i if consumer i changes group membership, conditional on maintaining the same score. We define algorithmic bias (D_{it}) mathematically as follows:

$$
\begin{aligned}
D_{it} &= E_t[Q_{i,H} \mid S^*] - E_t[Q_{i,L} \mid S^*] \\
&= (\gamma_{Ht} - \gamma_{Lt})\, S^* + \left[(1-\hat{\gamma}_{Ht})\, \hat{A}_{Ht} - (1-\hat{\gamma}_{Lt})\, \hat{A}_{Lt}\right] \\
E[D_{it}] &= [1 - p\ \gamma_{Lt} - (1-p)\ \gamma_{Ht}]\,(\hat{A}_{Ht} - \hat{A}_{Lt}) \qquad (4) \\
&\text{where } S^* = S_{i,H} = S_{i,L} \\
&\text{and } p = \text{percentage of consumers that are H-group}
\end{aligned}
$$

Evidence of algorithmic bias occurs when $D_{it} \neq 0$. We find that increasing consumer heterogeneity ($\sigma^2_{q_{jt}}$), decreasing measurement error (σ^2_{ε}), or increasing score validity (γ_{jt}) reduces

$E[D_{it}]$, the magnitude of average algorithmic bias. It can be shown that the average algorithmic bias produced by a Group-Aware algorithm is always positive as long as $A_H > A_L$ and $0 < \gamma_{jt} < 1$.

If the Group-Aware algorithm selects to serve consumers with expected quality that exceeds threshold Q^{min}, then the minimum score threshold for group j, S_{jt}^{min}, is the consumer score that corresponds with Q^{min} at time t. The minimum score threshold is dynamic and estimated from the training data. The Group-Aware algorithm's minimum score threshold for group j is:

$$S_{jt}^{min} = \hat{A}_{jt} + \left(\frac{Q^{min} - \hat{A}_{jt}}{\hat{\gamma}_{jt}} \right) \tag{5}$$

When the group's estimated mean quality ($\hat{A}_{jt}$) is greater than Q^{min} (which means that, on average, the group is expected to be profitable to the firm), then the minimum score threshold is less than Q^{min}. This allows for measurement error in measuring quality. The minimum score threshold increases towards Q^{min} as score validity $\hat{\gamma}_{jt}$ increases, until it equals Q^{min} at its limit when $\hat{\gamma}_{jt} = 1$. When both the H-group and L-group are, on average, expected to be profitable to the firm, the H-group's minimum score threshold is lower than the L-group's as long as $A_H > A_L$ and $0 < \gamma_{jt} < 1$. This means that in the aggregate, L-group consumers face higher thresholds to receive service than H-group consumers.

What if the algorithm did not use group information? This type of algorithm, which we refer to as a *Group-Blind* algorithm, would effectively ignore group information and use a single minimum score threshold (S_t^{min}) to assess all consumers. Algorithms that do not use any sociodemographic data or proxies are not currently in common use. The shortage of Group-Blind algorithms may exist because they use less information (do not use group information) and are less accurate as classification algorithms. Dropping the subscript j in equation (3) and equation

(5) results in the Group-Blind algorithm's expectation of quality and minimum score threshold.

$$E_t[Q_{ij} \mid S_{ij}] = \hat{\gamma}_t S_{ij} + (1 - \hat{\gamma}_t)\,\hat{A}_t \quad \text{, where} \tag{6a}$$

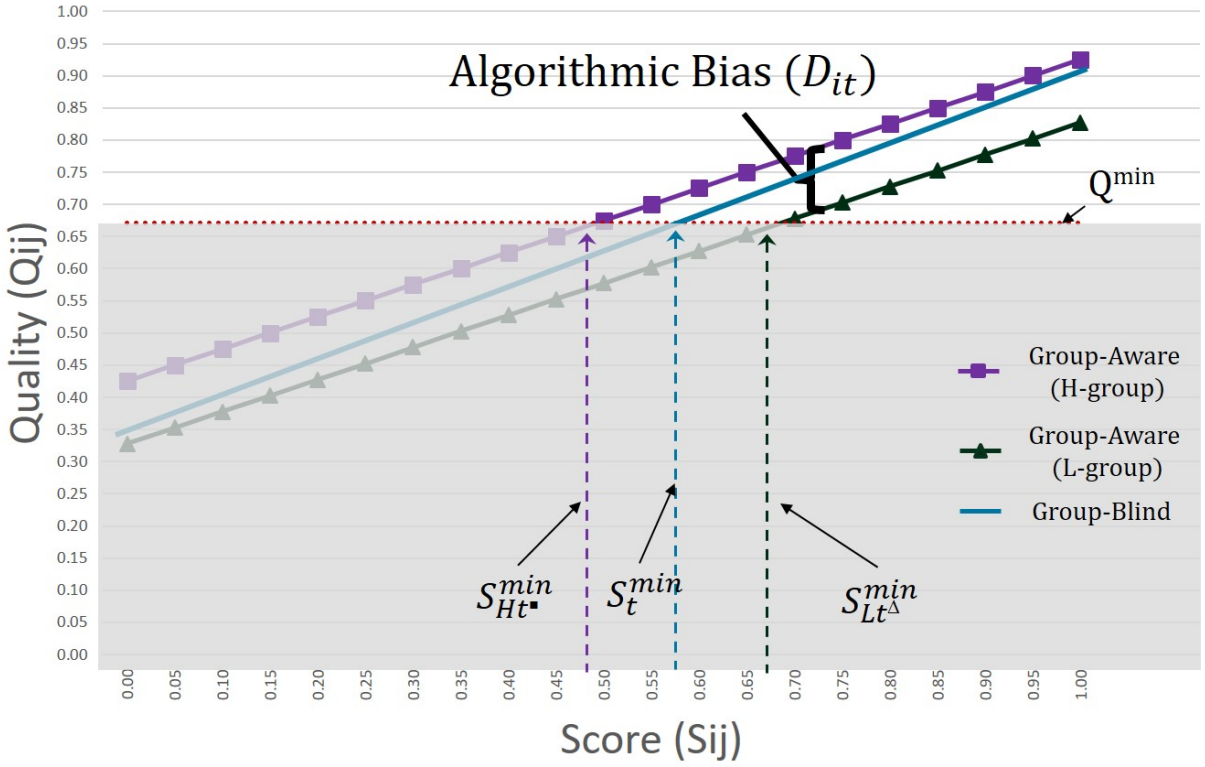
$$S_t^{min} = \hat{A}_t + \left(\frac{Q^{min} - \hat{A}_t}{\hat{\gamma}_t}\right) \tag{6b}$$

$$\hat{\gamma}_t = \frac{\hat{\sigma}_{q_t}^2}{\hat{\sigma}_{q_t}^2 + \sigma_\varepsilon^2} \tag{6c}$$

The primary difference between the Group-Aware and Group-Blind algorithms lies in the group-based information used. Instead of estimating $\hat{A}_{jt}$ and $\hat{\gamma}_{jt}$ for each group j in the training data, the Group-Blind algorithm estimates $\hat{A}_t$ (pooled mean) and $\hat{\sigma}_{q_t}^2$ (pooled variance of quality) from a mixture of two normal distributions of consumer data. Based on our definition of algorithmic bias, this algorithm is an unbiased algorithm because $D_{it} = 0$ for all values of S_{ij}.

Figure 1 displays a graphical example of the outcomes of the two types of algorithms we consider in this research. In this graph, the three diagonal lines represent the mapping of a

Figure 1: Algorithm's Expectation of Consumer Quality



consumer's score (x-axis) to the algorithm's expectation or prediction of the consumer's true quality (y-axis). The lines with square and triangle markers represent the Group-Aware algorithm's expectation of H-group and L-group consumers respectively. The line without markers represents the Group-Blind algorithm's expectation of any consumer's quality. The horizontal dotted line at the top of the shaded block represents Q^{min}, the quality level where the service makes 0 profit. Consumers with expected (or real) quality levels above this line would receive algorithmic service because they are the ones who are expected to be profitable. It follows that scores that correspond with the points where the three diagonal lines intersect with Q^{min} are the minimum score thresholds for the Group-Aware algorithm (H-group and L-group criteria) as well as for the Group-Blind algorithm. Notice that the minimum score threshold for the L-group from the Group-Aware algorithm is higher than the criterion for the H-group. This is because the average quality for the L-group is lower. Also notice that there is a gap between the H-group and L-group in terms of expected quality of a consumer for any given score in this example. This gap is our measure of algorithmic bias, as referenced in equation 4. Two people from different groups with the same score (information based on individual merit) would get a different expectation of quality because of the differences in mean quality of the groups they come from. This difference could lead to cases where those from the H-group receive service while similar members from the L-group would not.

What impact do the decisions of these algorithms have on consumer WOM? In the next section, we describe how WOM can affect the profitability of algorithms and impact consumer decisions for service over time (e.g., the cases of the Apple Card, TikTok, standardized tests, etc.).

How Does Word-of-Mouth Arise from Algorithmic Bias?

Suppose consumer *i* has two options of where to get service in the city – the firm using a Group-Aware algorithm or the firm using a Group-Blind algorithm. The consumer's objective is simply to patronize the firm that is most likely to give service and is not too far away from home.

We assume that consumer i is not necessarily aware that either service is using an algorithm, nor aware of potential bias from the algorithms. Consumer i knows that both firms screen prospective consumers before offering services. Furthermore, consumer i knows that friends may also have already sought services at either firm.

Prior research has established that WOM is highly influential on consumer choice (Brown and Reingen 1987; Trusov et al. 2009). Suppose consumer i uses WOM from the consumer's social network to inform the decision. However, unlike prior research which suggests consumers use WOM to assess the quality of services, suppose consumer i uses WOM to assess the likelihood of receiving service. Furthermore, suppose consumer i infers this likelihood by observing not only the percentage of the consumer's network that received service, but also the percentage who are friends from the consumer's own group. For example, imagine a Black content creator who wants to choose between TikTok and Fanbase to be a home platform for the creator's content. Assume the Black content creator asks their network, "Did you get your content accepted by TikTok? By Fanbase?" If the Black creator observes from responses that Tiktok accepted a higher proportion of posts overall, but FanBase accepted a higher proportion of Black creator posts, then this information could influence the Black creator's assessment of their own chances of acceptance at TikTok and Fanbase. We assume consumer i's utility for selecting firm b at time t is as follows:

$$U_{ibt} = \phi WOM_{ibt} - Dist_{ib} + \varepsilon_{ibt} \tag{7}$$

In equation (7), ϕ represents the importance the consumer places on WOM overall. The inclusion of geographical distance between the consumer and the service ($Dist_{ib}$) in the utility function is consistent with models in the consumer store choice literature (e.g., Rust and Donthu 1995). We account for additional unobservable factors that influence a consumer's utility with an extreme-valued distributed error term, ε_{ibt}.

Consumers learn about service options from their social networks. Extant research has

shown that people who share sociodemographic attributes (e.g., race/ethnicity, age, gender, income level, education level, etc.) are more likely to be socially connected (McPherson et al. 2001; Reagans 2005). Prior research has also shown that consumers are influenced by homophily and sociodemographic group membership (Lam et al. 2009; Ma et al. 2015; Uslu et al. 2013). Most relevantly, consumers give more consideration to in-group versus out-group sources of WOM in their consumption choices (Nitzan and Libai 2011; Podoshen 2006; Risselada et al. 2014). We model WOM in this regard. WOM_{ibt} represents consumer i's assessed probability of receiving service from firm b based on their social ties' answers to the question, "Have you ever received service from firm b?". WOM_{ibt} is equal to the weighted proportion of consumer i's social ties who received service from firm b:

$$\begin{aligned} WOM_{ibt} &= \frac{\sum_k w_{ik} \;\; \mathbf{1}(\text{if } b \text{ has ever offered a loan to } k \text{ as of time } t)}{\sum_k w_{ik}} \\ w_{ik} &= \frac{1 + \psi \, \mathbf{1}(i,k \in j)}{Soc_{ik}} \;\; \text{(social tie weight)} \end{aligned} \tag{8}$$

The social tie weight w_{ik} is a function of social tie strength (Soc_{ik}) and the influence of WOM sourced from an in-group social connection (ψ) between consumers i and k (Granovetter 1973). Prior research has shown that strong ties (irrespective of homophily) have a greater influence on an individual's consumption choice than weak ties (Brown and Reingen 1987). Larger Soc_{ik} indicates a weaker social tie because the connection is "further away" socially. Note that Soc_{ik} is not a measure of geographic distance between people. Although the value ϕ indicates that consumer i values WOM from in-group as well as out-group sources, the factor ψ is the additional weight that a consumer places on WOM from someone who is in the same group j as consumer i. We assume that consumers are aware of their own group membership and can observe the group membership of others.

We hypothesize that WOM will have no effect on demand for algorithmic service in scenarios where the firm is a monopoly or competing firms use the same type of algorithm. In the monopoly scenario, since there are no other service options for the consumer, WOM about whether friends were served should have no influence on whether consumer i chooses the

monopoly firm. In scenarios where all firms use the same type of algorithm, we expect that the proportion of friends served in consumer's social network (in-group as well as overall) will be, on average, the same across service options. In contrast, when Group-Aware and Group-Blind algorithms compete, we they will produces different service rates overall as well as for each group. Hence, we expect different impact on WOM which will consequently produce different impact on consumer demand. Furthermore, we propose that greater weight (importance) placed on WOM will increase demand in favor of the service that has higher overall service rates, but increased weight on in-group WOM increases service demand for the firm with higher service rates for group j.

Our consumer decision model has some important implications. First, as the overall WOM weight ϕ increases, the likelihood increases that consumer i selects the firm that has served the greater total number of social ties in consumer i's network. For this reason, we expect that increasing ϕ will increase the demand for the firm that has the greatest total demand. On the other hand, increasing the in-group WOM weight ψ increases the likelihood that consumer i chooses the firm that has served the greater number of consumer i's social ties in their group. Since Group-Aware (Group-Blind) firms are likely to have more H-group (L-group) members because it is relatively easier to get served from them, larger ψ should drive more H-group (L-group) consumers to Group-Aware (Group-Blind) firms. A second implication is that consumers can be completely unaware that algorithms make the decisions, and yet the effects of algorithmic bias on WOM can still impact the consumer's decision. If fewer members of a consumer's group are receiving service from a firm, it could decrease the consumer's likelihood of patronizing that firm.

Extant literature asserts that Group-Aware algorithms are more accurate and profit-maximizing (Phelps 1972; Fu et al. 2021; Zhang et al. 2021). Our research findings are consistent with this – as long as social effects in the form of consumer WOM are not taken into account. How would a Group-Blind algorithm perform against the Group-Aware algorithm if they competed in a marketplace where consumer WOM is taken into account? The next section addresses this question. We find conditions where the Group-Aware algorithm is

profit-maximizing in the short run, but not necessarily so in the long run.

HOW DOES ALGORITHMIC BIAS IMPACT CONSUMER DEMAND?

Our conceptual framework provides insights into how algorithms can make service decisions and how consumers could evaluate WOM to forecast their chances of acceptance for service. However, our theoretical models are limited because they model the behaviors of individual algorithms or consumers. They do not provide insight into emergent macro effects of the interactions of those individual behaviors at scale. They also do not provide insights into the dynamic effects of both the supply-side mechanism (algorithms in a competitive setting) and demand side mechanism (consumer WOM) operating together in a complex endogenous social system. Although there is a rich history of analytical models that show equilibrium outcomes of firm and consumer decisions in a market, incorporating into an analytical model the dynamic complexities of two populations of consumers that are socially connected in networks in a market of competing firms is highly difficult at best. For this reason, we turn to agent-based modeling (ABM) to examine the dynamics of algorithmic bias, algorithmic competition, and its impact on consumer WOM and demand. ABMs are well-suited to this modeling challenge because of their ability to simulate interacting autonomous individual agents (firm algorithms and consumers in our setting) in the complex setting of social networks in a competitive algorithmic market. Agent-based modeling is a method that enables the researcher to model the micro-behavior (e.g., consumer *i*'s selection of firm, firm *b*'s algorithmic evaluation of each consumer applicant) and interactions of autonomous individual agents (consumer WOM) to analyze emergent effects at scale such as demand (Goldenberg et al. 2001; Rand and Rust 2011). In the next section, we discuss our ABM to investigate the long run implications of algorithmic bias on consumer word-of-mouth.

Description of the ABM Design

The ABM models a city containing two firms and a population of socially-connected consumers from two sociodemographic groups – the H-group and L-group. Consumers are connected to other consumers in a social network. At the beginning of each ABM run, firms and consumers are randomly distributed throughout the geographic area. The two firms are banks that compete in the provision of lending services (loans, credit cards, etc.). We use lending merely as an illustrative example. Our findings would apply to other contexts that meet the criteria mentioned in section 2. Becker (1957) theorized that if non-discriminatory firms compete against discriminatory firms, market forces would drive out discrimination because non-discriminatory firms would have a labor cost advantage. In contrast, competition in our ABM allows us to investigate whether consumer WOM provides firms using unbiased algorithms (Group-Blind firms) a demand advantage.

The ABM's parameters are calibrated with empirical data. We use empirical data from the Copenhagen Networks Study (Sapiezynski, Stopczynski, Lassen, and Lehmann 2019) to form the social network of consumers. Collected in 2013, this network data is comprised of the Facebook friendship connections of 787 freshman (22% female) at the Technical University of Denmark. We designate the female nodes as members of the L-group and everyone else as H-group members. To test robustness of the model against different network structures, we also build ABM models based on synthetic networks, including a complete (fully-connected) network, a Erdös - Rényi random network (Erdös and Rényi 1959), and a Barabási-Albert preferential attachment network (Barabási and Albert 1999). In a complete network, all consumers are connected to all other consumers. In a random network, the probability that two consumers are connected is equally likely across all consumers. In a preferential attachment network, some consumers are disproportionately more likely to be connected than other consumers. All three alternative network structures are widely used in the literature (Rand and Rust 2011).

When using the using the Copenhagen Networks dataset, the parameter of population mix

(proportion of consumers that is an H-group member) is fixed and determined by the empirics (78% H-group vs. 22% L-group). In contrast, we vary the population mix parameter in the ABM models using synthetic networks. This allows us to vary and control for the impact of sociodemographic mix of the population in the outcomes of the ABM. At the beginning of each ABM run, the ABM model randomly draws the percentage of population that is H-group from a uniform distribution $U(9\%, 63\%)$. We use the uniform distribution for this parameter as well as others in the ABM because of its diffuse nature. It enables us to observe ABM outcomes over a disparate set of population mix scenarios. The distribution's lower and upper bounds are calibrated on the percentage of the population that is White in South Africa and U.S respectively according to the 2011 South African National Census (SSA 2012) and Pew Research Center Report (Taylor and Cohn 2012).

The ABM randomly assigns to each consumer a quality level and score based on the relationships displayed in equation (1). Each consumer's quality and score remain fixed during the entire run. These values are empirically derived from credit data from the 2010 Equifax Federal Bulletin Report to form the distributions of credit scores and latent quality of each consumer group in the ABM.

Each ABM dynamic model runs for 60 time periods, where each time period represents one month. We selected this time frame because organizations commonly plan over a 5 year horizon. To test whether ABM results were sensitive to time periods, we also ran models of the primary model for 100, 200, 300, 400, and 500 periods. Results were robust to increased time periods, so the following reported findings are based on 60 periods. Empirically-derived ABM input values were calibrated to match the monthly time frame for each period.

In each time period, a randomly selected group of consumers applies for service. The application rate (percentage of consumers that seek lending services at time t) is the ABM input parameter that determines the number of consumers who seek service in each time period. Application rate allows us to vary and control for overall market demand for services. At the beginning of each ABM run, the ABM model randomly draws a monthly application rate from

the uniform distribution $U(0.417\%, 3.75\%)$. The distribution's lower and upper bounds are calibrated on the monthly equivalent of annual loan and credit application rates reported by the 2013 - 2021 NY Federal Reserve service's Credit Access survey (FRBNY 2021).

Consumers first gather WOM from their social network to inform their decision of which service to select. The information they gather is in answer to the question, "Have you ever received service from firm b?" It is important to re-emphasize here that consumers are not aware that firms are using algorithms for service decisions. Nor are they aware of any biases in the service decisions. They use information learned from their social networks simply to assess their own chance of receiving service. Social network WOM is derived from each contact's history of success or failure of receiving service from the firms. We model a consumer's choice with a multinomial logit, which has the utility function described by equation (7) and equation (8). The consumer's choice set contains the two firms and an outside option (another unknown service).

After consumers select a firm, each firm's algorithm screens prospective consumers and offers lending to those with scores exceeding the minimum score threshold of the firm. Each firm may have either a Group-Aware algorithm (uses sociodemographic group information) or a Group-Blind algorithm (does not use sociodemographic group information). Firm-realized profits, based on equation (2b), are earned after service consumption when quality is revealed (e.g., consumer repayment after receiving a loan). To update their estimates about group mean quality levels and to set new minimum score thresholds in each period, algorithms learn from historical score data of consumers who have interacted with the firm. We employ a 2 (variance of H-group consumer quality: low vs. high) $\times$ 2 (variance of L-group consumer quality: low vs. high) $\times$ 2 (variance in score measurement error: low vs. high) $\times$ 720 (random-seeds for random-generators of replicates) design of experiments of algorithmic competition over 60 time periods. Hence, this is a $2^3 \times 720$ full factorial design (5,760 separate models) in each run of the agent-based model (ABM).

We use the method of Kapeller, Jäger, and Füllsack (2019) to include homophily in the design of the synthetic networks. This method allows us to preserve the theoretical structure of

the intended social network, while ensuring that group members are highly likely to be connected to each other. Table 2 provides statistics that characterize the networks we used in the ABM runs. We find that all four network structures produce qualitatively similar results in our investigation. The remainder of the article reports results based on the empirical Copenhagen network data, except where noted. More detail on the robustness analysis of the network structure is in Web Appendix B. There are two outcomes (dependent variables) of the ABM that are of interest for

Table 2: Summary Statistics of Social Network Topology

Statistic	**Complete**	**Random**	**Preferential Attachment**	**Empirical**
Nodes	787	787	787	787
Edges	309,291	124.4	788.0	105.8
Degree	786.0	12.7	2.0	14.9
Clustering Coefficient	1.000	0.016	0.000	0.304
Density	1.000	0.016	0.003	0.019
Homophily	0.585	0.636	0.892	0.642

this study. Because we are comparing competing firms, we examine the net difference between the two firms in terms of cumulative demand (consumer applications) and cumulative profits by the end of an ABM run. Two parameters for word-of-mouth (WOM) are the independent variables of interest. Described by equation (7) and equation (8), we employ the WOM parameters ϕ (overall weight placed on WOM from a consumer's social network) and ψ (weight placed on WOM from the in-group portion of a consumer's social network). At the beginning of each of the 5,760 ABM runs, we draw random values for WOM factors ϕ and ψ from the uniform distributions $U(0.01, 20)$ and $U(0.01, 3)$, respectively. The upper and lower bounds in these distributions are based on Trusov, Bucklin, and Pauwels (2009) for overall WOM influence and Brown and Reingen (1987); Podoshen (2006); Zhao and Xie (2011) for in-group WOM influence. When $\psi = 1$, consumer i equally weights in-group and out-sources of WOM. A $\psi > 1$ implies that i places greater weight on WOM from other in-group social ties.

Three of the input parameters were manipulated in the ABM: 1) high and low values for variance of quality (customer heterogeneity) in the H-group, 2) the same for the L-group, and 3)

high and low variance of score, conditional on quality (measurement error). Other input parameters used in the ABM are either randomly drawn once per ABM run or are constant values. All input parameter values used in the ABM run are empirically derived from sources that inform credible values for financial/credit card lending or sociodemographic characteristics. These input parameters values and their empirical sources are listed in table 3. Note that Q^{min} is below the mean quality levels of both groups ($Q^{min} = 620$, $A_H = 723$, and $A_L = 640$), indicating that on average, both groups are profitable. To investigate our hypothesis about the effects algorithms on

Table 3: Empirical Calibration of ABM Parameters

Parameter	Low	High	Distribution	Source
H-group mean of quality (A_H)	723	723	Constant	Equifax 2006/2010 in Federal Bulletin Report Based on the mean credit scores of U.S. White and Black consumers (723 and 640 respectively) and mean minimum score associated with loan offers (620)
H-group mean of quality (A_L)	640	640		
Marginal consumer quality (Q^{min})	620	620		
Consumer Heterogeneity (σ_q^2)	$(45.5)^2$	$(80.2)^2$	Constant	Equifax 2006/2010 in Federal Bulletin Report; Corresponds to a standard deviation of 45.5 (low condition) and 80.2 (high condition)
Measurement error (σ_ε^2)	$(45.5)^2$	$(80.2)^2$		
Application Rate/Month	0.4%	3.8%	Uniform	2013-21 Survey of Consumer Expectations by the Federal Reserve Bank of New York Range from 5% to 45% annual application rates
H-group % of Population	9%	63%	Uniform	2011 South African National Census 2011 Pew Research Center Report
WOM-overall (ϕ)	0.01	20	Uniform	Brown and Reingen (1987); Podoshen (2006); Zhao and Xie (2011); Trusov, Bucklin, and Pauwels (2009)
WOM-ingroup (ψ)	0.01	3		

WOM in monopolistic and competitive markets, we ran five separate marketplace scenarios using the Copenhagen Networks empirical data: two monopoly scenarios (a single Group-Aware or a single Group-Blind algorithm service) and three competitive duopoly scenarios (Group-Aware vs. Group-Blind algorithms, Group-Aware vs. Group-Aware algorithms, and Group-Blind vs. Group-Blind algorithms). We produce 5,760 ABM runs for each marketplace scenario. We ran the two monopoly scenarios to assess WOM effects on demand and profits without competition involved. The three duopoly scenarios help us understand whether WOM has an effect when competing firms use the same type of algorithm and when they use different types of algorithms. To reiterate our hypothesis, we expect WOM to have an effect in only the competitive scenario

where the algorithms differ. This is because Group-Aware and Group-Blind algorithms differ in service rates to sociodemographic groups. This will result in differing impact on WOM and consequently long run demand.

In addition to network structure robustness checks, we also conducted robustness checks on distributional assumptions about consumer quality. Other prior empirical research has shown that the distribution of customer revenue, a 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 quality (Q_{ij}). We also tested a uniform distribution, a diffuse distributional assumption. 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.

Each of the 5,760 ABM models in a given ABM scenario of the bank-applicant ecosystem generates one data record in our dataset. Given that we ran ten separate ABM scenarios, we generated 57,600 records. We used the NetLogo programming language (Wilensky 1999) to develop and run the agent-based models. Web Appendix A provides more details on the design and set up of the ABM.

Tests and Measures

To assess the effects of algorithmic service decisions in the presence of consumer WOM, we first calculate our dependent variables for the three competitive scenarios– the difference between the cumulative demand (applications) and profits of the two competing firms generated by the end of the 60 period (5-year). For ease of exposition, we will refer to these quantities as the net cumulative demand and net cumulative profits. In the competitive scenario where one firm uses a Group-Aware algorithm and the other uses a Group-Blind algorithm, a positive value for net cumulative demand (profits) indicates that the Group-Blind algorithmic service generated more demand (profits) than the Group-Aware algorithmic service. In the homogeneous competitive scenarios, the dependent variable is simply the difference between the first and second firm. For

the two monopolistic scenarios, our dependent variable is simply cumulative demand and profits.

We use regression to analyze the relationship between the dependent variables of net cumulative demand (profits) and the two WOM parameters – our parameters of interest. These are the weights on overall WOM (ϕ) and in-group sourced WOM (ψ). A significant (insignificant) coefficient on overall WOM in the scenarios where different (same) type of algorithms compete would support our hypothesis that WOM only matters when the competing algorithms differ. In the Group-Blind vs. Group-Aware competitive scenario, a positive (negative) and significant coefficient on overall WOM or in-group WOM would indicate that when consumers place greater weight on WOM overall or in-group WOM, demand and profits increase in favor of the Group-Blind (Group-Aware) algorithmic service.

In the regression, we also include variables to control for other factors that could influence cumulative demand and profits. We include the application rate, average distance from consumers to each of the firms, score validity for the H-group (γ_H), and score validity for the L-group (γ_L) groups. Score validity is of particular importance as a control because score validity incorporates customer heterogeneity (σ_q^2) and measurement error (σ_ε^2). Score validity influences each algorithm's service rates for each group j because score validity directly affects minimum score thresholds. Score validity also directly influences the degree of algorithmic bias produced by the Group-Aware algorithm. To summarize, we assess each of the ABM scenarios with the following OLS regression model:

$$Net.Cumulative.Demand_r = \beta_0 + \beta_1 WOM_r + \beta_2 WOM_r^{ingroup} + \beta\ controls_r + \varepsilon_r \tag{9}$$

We also use the same regression with the dependent variable of net cumulative profits. Next, we present insights from our analysis about algorithmic bias's impact on demand in the long run.

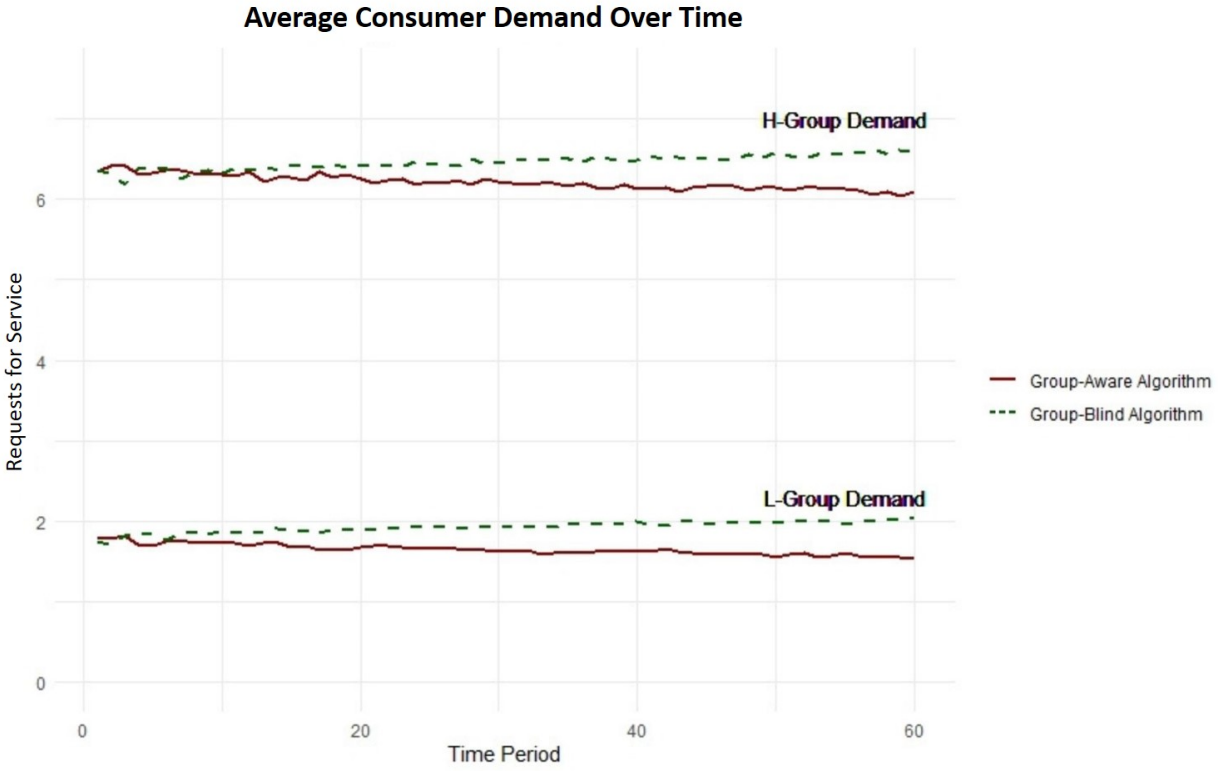
Analysis and Results

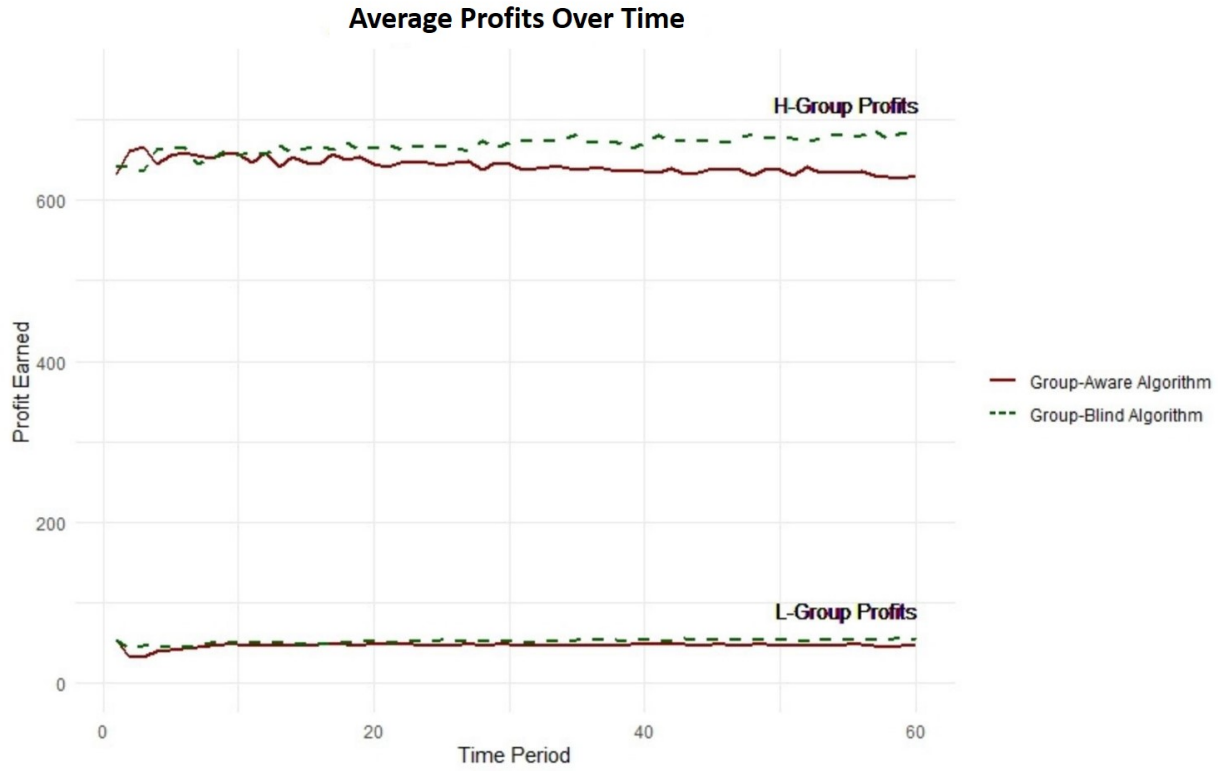
The overall finding is that algorithmic bias is profitable in the short run but unprofitable in the long run. Initially, the Group-Aware algorithm generates more profits on average, consistent with the economics literature (e.g., Phelps 1972). However, after sufficient WOM, there is a profit reversal; within months, the Group-Blind algorithm surpasses the Group-Aware algorithm. The following provides details on the statistical analyses that support this conclusion. First, average results from the ABM data indicate that Group-Aware algorithms initially dominate Group-Blind in profits. However, the average time it takes Group-Blind to surpass Group-Aware in demand and profits is 5 months. In 95% of the ABM runs where Group-Blind surpasses Group-Aware, it does so within 22 months. The two charts in Figure 2 show average consumer demand and profits over time from firms using Group-Aware and Group-Blind algorithms. Table 10 in Web Appendix C provides more detail about the dynamics of Group-Aware and Group-Blind demand and profits over time. The x-axes for both graphs are time periods. The y-axis of the top chart represents average demand per period (applications submitted). The bottom chart y-axis represents average profits per period. Averages are based on 5,760 ABM runs of competition between a firm using a Group-Aware algorithm and one using a Group-Blind algorithm.

Based on Welch two-sample t-tests, we find that firms using the Group-Blind algorithm exceed those using the Group-Aware algorithm in terms of long-term average total demand (502.99 vs. 471.43; $t(11,445) = 6.82,\ p < .001$) and total profits (43,229.51 vs. 41,364.88; $t(11,485) = 4.62,\ p < .001$). Among the H-group consumers, Group-Blind dominates Group-Aware in average long-term demand (387.62 vs. 372.71; $t(11,487) = 4.13,\ p < .001$) and profits (40,136.66 vs. 38,579.55; $t(11,485) = 4.16,\ p < .001$). The same holds for the L-group demand (115.37 vs. 98.72; $t(11,205) = 15.61,\ p < .001$) and profits (3,092.85 vs. 2,785.33; $t(11,508) = 8.92,\ p < .001$).

Recall that the Copenhagen Networks population is comprised of 78.0% H-group and 22.0% L-group. By comparison, the mix of applications at both firms reveal a shift in demand. The

Figure 2: Demand and Profits Over Time





average Group-Aware firm demand is comprised of 79.1% H-group/20.9% L-group while Group-Blind has a 77.1% H-group/22.9% L-group mix. The Group-Blind algorithm had higher loan offer rates (94.81% vs. 92.01%; $t(11,496) = 43.73\ ,\ \ p < .001$). Group-Blind also produced significantly higher loan offer rates for L-group members (86.97% vs. 68.80%; $t(7,848.7) = 77.47,\ \ p < .001$). In contrast, the Group-Blind algorithm offered H-group members loans less often than the Group-Aware, but the difference between Group-Blind and Group-Aware loan offer rates is much smaller (97.13% vs. 98.10%; $t(10,924) = -21.97,\ \ p < .001$) than for L-group consumers. H-group members had a relatively easier time getting loans from either firm whereas L-group consumers had a much better chance getting a loan from the Group-Blind firm.

Tables 4 and 5 display results from the regression analysis using equation (9). The

Table 4: Competitive Market Demand Scenarios with WOM

	Dependent variable:		
	Net Applications		
	GB vs. GA	GA vs. GA	GB vs. GB
	(1)	(2)	(3)
Intercept	-127.1^{***}	23.9	78.7^{***}
	(23.8)	(24.5)	(23.9)
WOM-overall	3.7^{***}	-0.4	0.03
	(0.4)	(0.4)	(0.4)
WOM-ingroup	-5.2^{*}	0.8	-1.0
	(2.8)	(2.8)	(2.8)
Score Validity-H-group	-83.0^{***}	1.2	-20.3
	(14.9)	(15.3)	(14.9)
Score Validity-L-group	38.8^{***}	5.5	-3.5
	(14.9)	(15.3)	(14.9)
Application Rate	$2,485.3^{***}$	276.6	670.6^{***}
	(243.0)	(248.8)	(242.3)
Distance to Second Firm	866.8^{***}	616.9^{***}	584.6^{***}
	(38.8)	(40.5)	(39.4)
Distance to First Firm	-597.6^{***}	-704.2^{***}	-796.9^{***}
	(40.1)	(41.0)	(39.9)
Observations	5,760	5,760	5,760
R^2	0.1	0.1	0.1
Adjusted R^2	0.1	0.1	0.1
Residual Std. Error (df = 5752)	177.9	182.1	177.3
F Statistic (df = 7; 5752)	132.1^{***}	80.7^{***}	95.3^{***}

Note: Net Applications = First - Second Firm Applications
GA = Group-Aware Firm; GB = Group-Blind Firm
*p<0.1; **p<0.05; ***p<0.01

dependent variables of the first column in each table are net applications and net profits: Group-Blind - Group-Aware applications (profits). The second and third columns are net

Table 5: Competitive Market Profit Scenarios with WOM

	Dependent variable:		
	Net Profits		
	GB vs. GA	GA vs. GA	GB vs. GB
	(1)	(2)	(3)
Intercept	$-9,498.64^{***}$	$1,830.49$	$7,403.53^{***}$
	$(2,079.35)$	$(2,123.78)$	$(2,073.25)$
WOM-overall	266.26^{***}	-21.77	19.42
	(35.79)	(36.47)	(35.60)
WOM-ingroup	-574.88^{**}	123.84	-40.22
	(241.87)	(245.57)	(239.73)
Score Validity-H-group	$-4,339.51^{***}$	-13.25	$-1,600.21$
	$(1,302.66)$	$(1,323.92)$	$(1,292.42)$
Score Validity-L-group	$1,929.99$	903.63	-510.04
	$(1,302.66)$	$(1,323.92)$	$(1,292.42)$
Application Rate	$152,355.29^{***}$	$24,698.72$	$60,752.92^{***}$
	$(21,218.29)$	$(21,567.06)$	$(21,053.87)$
Distance to Second Firm	$73,914.89^{***}$	$53,863.86^{***}$	$49,805.92^{***}$
	$(3,390.69)$	$(3,511.83)$	$(3,428.26)$
Distance to First Firm	$-53,288.93^{***}$	$-61,677.48^{***}$	$-70,389.23^{***}$
	$(3,497.32)$	$(3,552.60)$	$(3,468.07)$
Observations	5,760	5,760	5,760
R^2	0.13	0.09	0.10
Adjusted R^2	0.12	0.09	0.10
Residual Std. Error (df = 5752)	15,529.53	15,782.96	15,407.41
F Statistic (df = 7; 5752)	117.78^{***}	82.07^{***}	95.85^{***}

Note: Net Profits = First - Second Firm Profits
GA = Group-Aware Firm; GB = Group-Blind Firm
*p<0.1; **p<0.05; ***p<0.01

applications and profits in scenarios where the competing firms both use Group-Aware algorithms or Group-Blind algorithms respectively. Tables 4 and 5 show that WOM and score validity have no statistically significant effect when the two competing firms use the same type of algorithm. We interpret this to mean that algorithmic bias has little effect on long-term demand and profits when competing firms in a market use the same type of algorithm. In contrast, we find that WOM and score validity do matter when Group-Aware algorithms compete with Group-Blind algorithms. The negative and statistically significant intercept suggests that at the baseline, Group-Aware algorithms are more profitable. However, the greater the weight that consumers place on WOM overall, the more the demand and profits favor the Group-Blind firm. In contrast, increasing the weight on in-group WOM increases demand and profits for the Group-Aware algorithm. Consistent with what we posited, more weight placed on in-group WOM increases the likelihood that the consumer applies to the firm where the majority of their in-group goes (as opposed to the majority of consumers overall). Consequently, this drives more profits to the Group-Aware firm because it serves more profitable H-group members. We find that the in-group

WOM effect is robust for the complete and random network structures. However, we find that in-group WOM is not statistically significant for preferential attachment networks. Investigating the relationship between network structure and effects of algorithmic bias on in-group sourced WOM is outside the scope of this research. However, we offer speculation that it could be related to the degree and clustering coefficients of a social network. The preferential attachment network used in the ABM has much lower degree, density, and clustering coefficients than the other networks. It could mean that denser social connections amplifies the effects of in-group sourced WOM. Investigating these relationships could be the basis for interesting future research.

H-group score validity has a statistically significant effect on demand and profits. L-group score validity has a statistically significant effect on demand, but not profits (although, when we run the ABM for 100 time periods or more, L-group score validity is statistically significant). The intuition behind this result is that increasing score-reliability (driven by increasing within-group consumer quality heterogeneity or decreasing quality measurement error) of the H-group shifts demand and profits towards the Group-Aware firm while increasing score validity of the L-group shifts demand and profits towards the Group-Blind firm. Recall that an increase in a group's score validity increases minimum score thresholds in both the Group-Aware and Group-Blind algorithms. Given that the Group-Blind algorithm's minimum score threshold is greater than the H-group and less than the L-group minimum score thresholds of the Group-Aware algorithm, increasing the minimum score thresholds at both algorithms means that some H-group members no longer have a score high enough to receive service at the Group-Blind firm. However, their scores still exceed the Group-Aware threshold, so they patronize that firm. An analogous situation happens for L-group members, where they shift towards the Group-Blind firm. An increasing minimum score threshold will mean more consumers will be rejected for loans. Rejections influence future consumers through WOM. Higher service application rates increases demand and profits in favor of Group-Blind firms. The intuition is that higher application rates increase the number of consumers talking to others about their success or failure to receive service from the firm. This accelerates the effects of of overall WOM. These findings are robust across different

social network structures and distributional assumptions about consumer quality (see tables 7, 8, and 9 in Web Appendix B for details). Furthermore, the analysis data generated from the theoretical network structures (complete, random, and preferential attachment) allows us to vary the population mix of H-group and L-group consumers. We find that increasing the proportion of the H-group population shifts the demand and profits towards Group-Blind firms (see tables 8 and 9 in Web Appendix B). This effect is robust across network structures. The intuition here is that if a sufficient number of L-group members shift their demand towards Group-Blind firms due to algorithmic bias, their WOM influences downstream decisions of some H-group members. In a world where L-group members are connected to H-group member, algorithmic bias against the marginalized group can ultimately influence members of the non-marginalized group.

To determine whether competition is a necessary component of the outcomes, we ran monopoly marketplace scenarios of the ABM. We performed a regression analysis on 11,520 observations that are from a combined dataset of Group-Aware and Group-Blind monopoly firm ABM runs. We find that the overall weight on WOM (ϕ) has no impact on demand or profits. The weight on in-group sourced WOM (ψ) has a marginally significant positive impact on demand, but no impact on profits. From this, we learn that WOM is not an important driver in a monopoly marketplace when algorithmic bias is present. When we control for the factors in the ABM runs in a regression (see table 6 in Web Appendix B for details), we find that the Group-Aware algorithm is more profitable ($\beta = 388.02$, $SE = 52.52$, $p < .001$). We find no statistical difference in overall demand between the Group-Aware and Group-Blind monopoly algorithms ($\beta = .001$, $SE = 0.32$, *NS*). This shows that although the service demand is the same in both scenarios (as the ABM was designed to facilitate), the selected mix of consumers to provide service to differs between the two algorithm and drives profits. This is consistent with extant research that shows Group-Aware algorithms are more profitable in a monopolistic setting (Aigner and Cain 1977; Phelps 1972).

Discussion

Findings from the ABM study lead us to conclude that relative to Group-Blind algorithms, Group-Aware algorithms using sociodemographic information may be more profitable in the short run but attract less demand and profits in the long run. This is because algorithmic service decisions could activate WOM that influences future consumers on their own service choices. Group-Aware algorithms impose tougher standards for service on marginalized groups than non-marginalized groups. In contrast, Group-Blind algorithms impose the same service standards on both groups. As a result, Group-Aware algorithmic service will serve fewer marginalized consumers than a comparable Group-Blind algorithmic service. On the other hand, non-marginalized consumers have a relatively easier time obtaining service from either type of algorithm. In our ABM study, WOM is a direct function of consumers served. The more consumers served by the algorithm, the more consumers can generate WOM that influences other consumers. Furthermore, consumers from both groups can interact with and influence each other with WOM. In contrast, Group-Blind algorithmic service serves more consumers and benefits from the WOM generated over time. This is because the Group-Blind algorithm has a lower service threshold for L-group consumers. Furthermore, although Group-Blind has a higher service threshold than the Group-Aware algorithm for H-group consumers, the H-group overall has a relatively easier time than L-group consumers in receiving service from either firm Recall that service rates for H-group consumers were 98.10% from the Group-Aware algorithm vs. 97.13% from the Group-Blind. In contrast, L-group service rates were 68.80% from the Group-Aware algorithm vs. 86.97% from the Group-Blind. Our findings support the conclusion that if there is an outside option for consumers, then in the long run, myopically profitable, rationally-based algorithmic bias does not pay.

GENERAL DISCUSSION

Our research shows that algorithms using group information (Group-Aware algorithms) can have a social impact with long run implications for firm demand and profits. Prior research implies that algorithms using group information (which can produce algorithmic bias) are profit-maximizing (Aigner and Cain 1977; Arrow 1973; Phelps 1972). Our research confirms this is the case in the short run, but not necessarily so in the long run. Our models show that Group-Aware algorithms produce biased outcomes. In competition against algorithms that do not use group information (Group-Blind algorithms), word-of-mouth can shift marginalized consumers (e.g., minorities, women, less educated, etc.) to select other firms that use less discriminatory algorithms. Furthermore, if word-of-mouth from marginalized consumers can influence a sufficient number of non-marginalized consumers (e.g., majority populations, men, etc.), they too may switch their preferences to firms that use non-discriminatory algorithms.

This is especially the case in our modern era of electronic word-of-mouth and social media. WOM has exploded in the social media era, which makes it very timely to consider its influence–something we do in this study. As of January, 2021, there were 4.7 billion active social media accounts in the world. This represents an increase of more than 490 million (13.2%) over the same period in 2020 (Kemp 2021). In advanced economies such as the U.S. and Europe, seven out of ten people use social media on a regular basis (Auxier and Anderson 2021). Social media has expanded personal networks in both size and diversity of interactions. For example, in a study of Facebook users across eleven countries, 46% of users say they rarely see in-person Facebook friends with whom they regularly interact. Nevertheless, the Facebook friends were still considered, along with their in-person contacts, part of their personal network. In the same Facebook study, 66% reported regularly interacting with people who differed from them in income levels. Half reported regular interaction with people of different racial, ethnic, and religious backgrounds (Silver and Huang 2019). With such widespread use of social media where people of different sociodemographic groups interact, social media is a powerful medium where

consumers can influence each other via word-of-mouth.

Our findings apply to contexts that meet three criteria: 1) consumers can be segmented into sociodemographic groups based on an attribute observable to consumers and the algorithm, 2) firms screen prospective consumers before providing service, and 3) the algorithm uses both individual consumer and group information about consumer's quality. For example, consider how our findings apply to the scenario of an algorithm recommending whether to rent a luxury apartment to a 65-year old garbage collector versus a 65-year old business entrepreneur who are comparable in net worth. Algorithmic decisions such as these, in isolation, may seem to have little impact on firm demand and profits. But the social patterns that emerge from algorithmic bias can produce outcomes with long-term demand and profit implications.

Theoretical Contributions

This study sits at the intersection of research in algorithmic bias, marketplace discrimination, differential service treatment, and word-of-mouth. Findings from this study contribute to the algorithmic bias and algorithmic decision-making literature by demonstrating conditions where algorithms using less information (i.e., do not use sociodemographic data) can outperform algorithms that use more information (i.e., do use sociodemographic data) in terms of meeting the algorithm's objective (demand and profits). We find that if consumer word-of-mouth is sufficiently influential across sociodemographic groups in marketplaces where the two types of algorithmic services compete, algorithms that use sociodemographic group information can be more profitable in the short run but less profitable in the long run. Unlike current rhetoric which asserts that fairness and profitability in algorithmic decision-making is a zero-sum tradeoff, our findings demonstrate dynamic conditions where fair, unbiased algorithms can also be profitable. This is in contrast to prior research that suggests algorithms using sociodemographic group information are profit-maximizing (Arrow 1973; Phelps 1972; Bjerk 2008; Fryer 2007). In this way, our findings offer an alternative point of view in a growing debate in recent studies of

whether use of sociodemographic data improves algorithmic performance as well as improves outcomes for the marginalized consumers they can impact (Fu et al. 2021; Kleinberg et al. 2018; Zhang et al. 2021).

Our findings also contribute to the marketplace discrimination and differential service treatment literature by showing how non-human service providers–algorithms–can discriminate against consumers and provide differential, unfair service treatment. Furthermore, such treatment can engender a human social response via word-of-mouth. Algorithms, used as tools of innovation on providing service in marketplace contexts, are also tools of macro- and meso-level marketplace system. Because algorithms have the ability to embed, reinforce, and amplify biased decision processes across millions of consumers, this produces a different level of marketplace discrimination that is structural and systemic in nature. This study complements prior work in marketplace discrimination and differential service treatment, whose lens focuses on discrimination or treatment at primarily micro-level (individual) human interactions (Arsel et al. 2021; Bradford and Perry 2021; Ekpo et al. 2018; Johnson et al. 2019; Haenlein and Kaplan 2012; Lepthien et al. 2017).

Insights derived from our study contribute to the literature on word-of-mouth by providing theory about how consumers evaluate WOM when consumers evaluate their chances of receiving service from firms who screen their prospective consumers. This alternative motivating reason for consumer use of WOM adds to conversations in the literature about consumer use of WOM for information acquisition purposes (Berger 2014) and how consumers evaluate WOM (Frenzen and Nakamoto 1993). Furthermore, our theoretical model shows that it is possible for consumers to have a WOM response to a firm negative action (biased algorithm decision) and yet be unaware of the existence of a negative experience (algorithmic bias) or the nature of its source (algorithms). This complements research into consumer WOM responses to negative actions of firms, where the consumer is aware of the negative action and of the source (e.g., Ward and Ostrom 2006).

Implications for Consumers

Our findings suggest that consumers may adopt protective behaviors in several ways as awareness of algorithmic bias grows. For example, consumers may seek firms that claim they do not use or minimize use of group information in their algorithms. One such firm is the auto insurer Root Insurance, whose marketing tagline is ”Fair car insurance in an app”. It states on its website that “Traditional car insurance companies focus on demographics like age, ZIP code, occupation, and credit score to price your coverage. We don’t think that’s fair. When we offer a rate, your behavior behind the wheel gets more weight than any other single factor” (Root 2019). Consumers may increase use of word-of-mouth as a socially protective tool against algorithmic bias. In the Apple Card, the IB and UK Level A standardized tests, and TikTok cases, consumers used word-of-mouth as a protective tool in three ways. First, consumers learned via word-of-mouth not only about which service algorithms were potentially biased but also about less biased alternative service options. Second, word-of-mouth became a social punishment tool that led to all four organizations subsequently making public statements about implementing algorithmic bias-reducing initiatives (Adam 2020; Contreras and Martinez 2021; Simonite 2020). Third, word-of-mouth signaled to New York state regulators potential algorithmic bias issues which subsequently led to an investigation into Apple Card’s algorithm (Harris 2019). Consumers may also change what information they disclose to algorithmic service. This idea has data privacy implications; on the one hand, consumers may disclose less information out of a desire to circumvent bias. A recent example is the case of the Black homeowner who doubled the appraisal value of her home after she concealed her race and gender on her refinance application (Bahney 2021). On the other hand, consumers may increase disclosure of other types of information that signal advantaged group attributes. For example, women credit card applicants may intentionally disclose alternative sources of assets on their applications in the hopes that the algorithm treats them better. These examples raise complex issues about the intersection of algorithmic bias and consumer data disclosure.

Implications for Managers

For organizations that meet the conditions of our research, our findings suggest that they may want to consider the benefits of employing non-biased, Group-Blind algorithms. Furthermore, firms may want to consider taking downstream social impact directly into account in their algorithm's objective function. Our research provides one way to model word-of-mouth, for example. In addition, firms should take into account competitive forces if their algorithms using sociodemographic information are competing with other algorithms that are not.

If the firm must use a Group-Aware algorithm, then findings recommend taking steps to reduce the bias. Recall our findings that reducing error in measuring quality or increasing variation in consumer quality reduces the magnitude of algorithmic bias. Hence, we recommend that firms consider investing in methods that reduce measurement error or increase the variability and representativeness of the data. These steps could improve the algorithm's learning about consumer quality and reduce bias in firms' algorithms. Firms could also take steps to measure and monitor algorithmic bias. To audit algorithms for bias requires collecting sociodemographic information (but not using it to train the algorithms). Our study provides a definition of algorithmic bias that could provide guidance on bias measurement based on the principle of individual fairness (Dwork et al. 2012). These suggestions are attenuated by the fact that investment can engender a cost that may alter the profit outcomes of our results, so these steps would have to be considered carefully.

Implications for Policymakers

Our research has implications for policymakers with regard to detecting, monitoring, and measuring algorithmic bias. In 2018, California, New York, and U.S. Federal governments actively debated whether to allow firms to use of sociodemographic group categories in algorithms 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 and 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).

A particularly challenging tension arises with the question of whether use of sociodemographic data harms or protects marginalized consumers. On the one hand, our research suggests that eliminating sociodemographic variables (and their proxies) from training data has the potential to reduce bias in algorithmic service decisions made about consumers. This also has the potential to improve firm outcomes by improving demand for services in the long run. On the other hand, policymakers may want to consider allowing firms using algorithms to collect sociodemographic information (without use as training data), because having such data could facilitate detection and correction of bias from their algorithms. Otherwise, as our bias definition suggests, detecting and measuring algorithmic bias becomes difficult. Another consideration is the use of word-of-mouth as a potential bellwether for detecting algorithmic bias. This was the case for the New York regulators who launched their investigation of the Apple Card. Algorithm regulation is under active development as of this writing. For example, in April, 2021 the European Union introduced a new artificial intelligence legal framework which builds upon Article 22 of the General Data Protection Regulation (GDPR). The regulations prohibit the use of algorithms which base outcomes on special categories such as race/ethnic origin, political opinions, or health status Vollmer (2018). Firms violating the new regulations could pay a hefty fine of up to 6% of their global sales European Commission (2021).

Limitations and Opportunities for Future Research

There are limitations to our study that open up avenues for future research. For example, our models assume consumers are members of only one sociodemographic. Thus, we do not address intersectionality (Crenshaw 2017), the concept of consumers belonging to multiple sociodemographic groups with complex interactions of advantage or disadvantage (e.g., a wealthy Asian female entrepreneur with a community college education). Furthermore, our research does not account for scenarios in which sociodemographic group attributes are observable to the algorithm but not to consumers (or vice versa). Another limitation is that we do not account for training data encoded with historic, structural, or systemic biases. Investigating word-of-mouth impact of an algorithm designed to be non-discriminatory but trained on biased data would be an interesting avenue for future research. A great deal of work is still needed, but we believe that our findings serve as a starting point in exploring these and many more questions about the social effects of algorithmic bias.

Conclusion

The goal of this research was to understand under what conditions biased algorithms could have potential effects on consumer WOM and subsequent demand for services over time. Our models demonstrate that although biased algorithms are more accurate than human decision-makers and more profitable in the short run, in the long run they generate less demand and are less profitable than unbiased algorithms. This is because consumer word-of-mouth drives consumers to select unbiased algorithmic services over time. This research emphasizes the long-term benefits of employing unbiased algorithms. By doing so, a firm could reduce algorithmic bias to improve societal well-being as well as its profits.

References

Adam, Karla (2020). The U.K. used an algorithm to estimate exam results. The calculations favored elites. *The Washington Post*.

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

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

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

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

Arsel, Zeynep, David Crockett, and Maura L Scott (2021). Diversity, Equity, and Inclusion (DEI) in the Journal of Consumer Research: A Curation and Research Agenda. *Journal of Consumer Research*. ucab057.

Asher-Schapiro, Avi (2020). Global exam grading algorithm under fire for suspected bias. *Reuters*.

Auxier, Brooke and Monica Anderson (2021). Social Media Use in 2021. Technical report, Pew Research Center.

Bahney, Anna (2021). When a Black homeowner concealed her race, her home's appraisal value doubled. *CNN*.

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

Barocas, Solon and Andrew D. Selbst (2016). Big Data's Disparate Impact. *California Law Review 104*, 63. Publisher: clr.

Bass, Frank M. (1969). A new product growth for model consumer durables. *Management Science 15*(5), 215–227.

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

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

Berger, Jonah (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology 24*(4), 586–607.

Berger, Jonah and Katherine L Milkman (2012). What Makes Online Content Viral? *Journal of Marketing Research*, 14.

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.

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

Bone, Sterling A., Glenn L. Christensen, and Jerome D. Williams (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.

Boulding, William, Ajay Kalra, Richard Staelin, and Valarie A Zeithaml (1993). A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions. *Journal of Marketing Research*, 21.

Bradford, Tonya Williams and Vanessa Gail Perry (2021). Marketing while Black: commentary on the Galak and Kahn 2019 Academic Marketing Climate Survey. *Marketing Letters 32*(3), 299–306.

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

Bruce, Norris I., Keisha M. Cutright, Renée Richardson Gosline, Jacquelyn S. Thomas, and Tiffany Barnett White (2020). How Business Schools Can Help Corporate America Fight Racism. *Harvard Business Review*. Section: Business education.

Buolamwini, Joy and Timnit Gebru (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, Volume 81 of *Proceedings of Machine Learning Research*, pp. 77–91. PMLR.

CDIpress (2019). Commissioner issues regulations prohibiting gender discrimination in automobile insurance rates. [Accessed May 10, 2019].

Chevalier, Judith A and Dina Mayzlin (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 10.

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

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science 29*(5), 944–957.

Contreras, Brian and Marisa Martinez (2021). Fed up with TikTok, Black creators are moving on. *Los Angeles Times*. Section: Technology.

Corbett-Davies, Sam and Sharad Goel (2018). The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *arXiv*.

Crenshaw, Kimberlé W (2017). *On Intersectionality: Essential Writings*. The New Press.

Crockett, David (2021). Racial Oppression and Racial Projects in Consumer Markets: A Racial Formation Theory Approach. *Journal of Consumer Research*, ucab050.

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

Dwork, Cynthia, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel (2012). Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference on - ITCS '12*, Cambridge, Massachusetts, pp. 214–226. ACM Press.

Ekpo, Akon E, Benét DeBerry-Spence, Geraldine Rosa Henderson, and Joseph Cherian (2018). Narratives of technology consumption in the face of marketplace discrimination. *Marketing Letters 29*(4), 451–463.

Elliott, Marc N., Peter A. Morrison, Allen Fremont, Daniel F. McCaffrey, Philip Pantoja, and Nicole Lurie (2009). Using the Census Bureau's surname list to improve estimates of race/ethnicity and associated disparities. *Health Services and Outcomes Research Methodology 9*(2), 69.

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

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

European Commission (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) | Shaping Europe's digital future. Technical Report COM(2021) 206, European Commission.

Evett, Sophia R., Anne-Marie G. Hakstian, Jerome D. Williams, and Geraldine R. Henderson (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.

Fader, Peter S., Bruce G.S. Hardie, and Ka Lok Lee (2005). RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. *Journal of Marketing Research 42*(4), 415–430.

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

FRBNY (2021). SCE Credit Access Survey - FEDERAL RESERVE BANK of NEW YORK.

Frenzen, Jonathan and Kent Nakamoto (1993). Structure, Cooperation, and the Flow of Market Information. *Journal of Consumer Research 20*(3), 360.

Friedman, Batya and Helen Nissenbaum (1996). Bias in computer systems. *ACM Trans. Information Systems 14*(3), 330–347.

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

Fu, Runshan, Yan Huang, and Param Vir Singh (2021). Crowds, Lending, Machine, and Bias. *Information Systems Research 32*(1), 72–92.

Gill, Jeff (2007). *Bayesian Methods: A Social and Behavioral Sciences Approach, Second Edition*. CRC Press. Google-Books-ID: Iq_epk4mtM4C.

Godes, David and Dina Mayzlin (2009). Firm-Created Word-of-Mouth Communication: Evidence from a Field Test. *Marketing Science 28*(4), 721–739.

Goldenberg, Jacob, Barak Libai, and Eitan Muller (2001). 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.

Gopinath, Shyam, Jacquelyn S. Thomas, and Lakshman Krishnamurthi (2014). Investigating the Relationship Between the Content of Online Word of Mouth, Advertising, and Brand Performance. *Marketing Science 33*(2), 241–258.

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

Haenlein, Michael and Andreas M Kaplan (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.

Haggerty, Neil (2018). Trump makes repeal of CFPB auto lending rule official.

Harris, Anne-Marie G., Geraldine R. Henderson, and Jerome D. Williams (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. *Newsweek*.

Heinemeier Hansson, David [@dhh] and Steve [@stevewoz] Wozniak (2019). The @AppleCard is such a f**** sexist program. My wife .... [Accessed May 19, 2021].

Henderson, Geraldine Rosa, Anne-Marie Hakstian, and Jerome D Williams (2016). *Consumer equality: Race and the American marketplace: Race and the American marketplace*. ABC-CLIO.

Herr, Paul M., Frank R. Kardes, and John Kim (1991). Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research 17*(4), 454.

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

Hogan, John E., Katherine N. Lemon, and Barak Libai (2003). What Is the True Value of a Lost Customer? *Journal of Service Research 5*(3), 196–208.

Homburg, Christian, Mathias Droll, and Dirk Totzek (2008). Customer Prioritization: Does it Pay off, and how Should it be Implemented? *Journal of Marketing 72*(5), 110–130.

Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2011). Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science 30*(2), 195–212.

Johnson, Guillaume D, Kevin D Thomas, Anthony Kwame Harrison, and Sonya A Grier (2019). *Race in the marketplace: Crossing critical boundaries*. Springer.

Kapeller, Marie Lisa, Georg Jäger, and Manfred Füllsack (2019). Homophily in networked agent-based models: a method to generate homophilic attribute distributions to improve upon random distribution approaches. *Computational Social Networks 6*(1), 9.

Kemp, Simon (2021). Digital Trends 2021. Technical report, Hootsuite.

Khalil, Ashraf, Soha Glal Ahmed, Asad Masood Khattak, and Nabeel Al-Qirim (2020). Investigating Bias in Facial Analysis Systems: A Systematic Review. *IEEE Access 8*, 130751–130761.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan (2018). Algorithmic Fairness. *AEA Papers and Proceedings 108*, 22–27.

Kordzadeh, Nima and Maryam Ghasemaghaei (2021). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 1–22.

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

Lambrecht, Anja and Catherine Tucker (2019). Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science 65*(7), 2966–2981.

Lepthien, Anke, Dominik Papies, Michel Clement, and Valentyna Melnyk (2017). The ugly side of customer management – Consumer reactions to firm-initiated contract terminations. *International Journal of Research in Marketing 34*(4), 829–850.

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. [Accessed May 10, 2019].

Ma, Liye, Ramayya Krishnan, and Alan L. Montgomery (2015). Latent Homophily or Social Influence? An Empirical Analysis of Purchase Within a Social Network. *Management Science 61*(2), 454–473.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology 2*, 415–444.

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

Nitzan, Irit and Barak Libai (2011). Social effects on customer retention. *Journal of Marketing 75*(6), 24–38.

Noble, Safiya Umoja (2018). *Algorithms of oppression: How search engines reinforce racism*. nyu Press.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan (2019). Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science 366*(6464), 447–453.

O'Neil, Cathy (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

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

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.

Poole, Sonja Martin, Sonya A. Grier, Kevin D. Thomas, Francesca Sobande, Akon E. Ekpo, Lez Trujillo Torres, Lynn A. Addington, Melinda Weekes-Laidlow, and Geraldine Rosa Henderson (2021). Operationalizing Critical Race Theory in the Marketplace. *Journal of Public Policy & Marketing 40*(2), 126–142.

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

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

Reagans, Ray (2005). Preferences, Identity, and Competition: Predicting Tie Strength from Demographic Data. *Management Science 51*(9), 1374–1383.

Risselada, Hans, Peter C. Verhoef, and Tammo H.A. Bijmolt (2014). Dynamic Effects of Social Influence and Direct Marketing on the Adoption of High-Technology Products. *Journal of Marketing 78*(2), 52–68.

Root, Insurance Co (2019). FAQ and customer support | Answers to Root Car Insurance questions.

Russell, Stuart and Peter Norvig (2020). *Artificial Intelligence: A Modern Approach* (4th edition ed.). Hoboken: Pearson.

Rust, Roland, Valarie Zeithaml, and Katherine Lemon (2000). *Driving Customer Equity : How Customer Lifetime Value is Reshaping Corporate Strategy*. New York: Free Press.

Rust, Roland T and Naveen Donthu (1995). Capturing Geographically Localized Misspecification Error in Retail Store Choice Models. *Journal of Marketing Research 32*(1), 103–110.

Rust, Roland T., J. Jeffrey Inman, Jianmin Jia, and Anthony Zahorik (1999). What You *Don't* Know About Customer-Perceived Quality: The Role of Customer Expectation Distributions. *Marketing Science 18*(1), 77–92.

Sapiezynski, Piotr, Arkadiusz Stopczynski, David Dreyer Lassen, and Sune Lehmann (2019). Interaction data from the Copenhagen Networks Study. *Scientific Data 6*(1), 315. Number: 1 Publisher: Nature Publishing Group.

Schmittlein, David C. and Robert A. Peterson (1994). Customer Base Analysis: An Industrial Purchase Process Application. *Marketing Science 13*(1), 41–67.

Silver, Laura and Christine Huang (2019). In Emerging Economies, Smartphone and Social Media Users Have Broader Social Networks. Technical report, Pew Research Center.

Simonite, Tom (2020). Meet the Secret Algorithm That's Keeping Students Out of College. *Wired*. Section: tags.

Srinivasan, Raji and Gülen Sarial-Abi (2021). When Algorithms Fail: Consumers' Responses to Brand Harm Crises Caused by Algorithm Errors. *Journal of Marketing 85*(5), 74–91.

SSA (2012). 2011 Census | Statistics South Africa.

Sweeney, Latanya (2013). Discrimination in Online Ad Delivery. *Communications of the ACM 56*(5), 44–54.

Taylor, Paul and D'Vera Cohn (2012). A Milestone En Route to a Majority Minority Nation. *Pew Research Center's Social & Demographic Trends Project*.

Thomaz, Felipe, Natalia Efremova, Francesca Mazzi, Gregory Clark, Ewan Macdonald, Rhonda Hadi, Joseph Bell, and Andrew T. Stephen (2021). Ethics for AI in Business. *SSRN Electronic Journal*.

TikTok (2021). Thanks a billion! [press release]. *Newsroom | TikTok*.

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

Uslu, Aypar, Beril Durmuş, and Sina Taşdemir (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. *New York Times (Online)*.

Vollmer, Nicholas (2018). Article 22 EU General Data Protection Regulation (EU-GDPR). Library Catalog: www.privacy-regulation.eu Publisher: SecureDataService.

Ward, James C. and Amy L. Ostrom (2006). Complaining to the Masses: The Role of Protest Framing in Customer-Created Complaint Web Sites. *Journal of Consumer Research 33*(2), 220–230.

Watts, Duncan J. and Peter Sheridan Dodds (2007). Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research 34*(4), 441–458.

Wieringa, Maranke (2020). What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona Spain, pp. 1–18. ACM.

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

Zhang, Shunyuan, Nitin Mehta, Param Vir Singh, and Kannan Srinivasan (2021). Frontiers: Can an Artificial Intelligence Algorithm Mitigate Racial Economic Inequality? An Analysis in the Context of Airbnb. *Marketing Science 40*(5), 813–820.

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

Web Appendix A: ABM Rules of Engagement

The agent-based model, which was developed and implemented with the NetLogo programming language (Wilensky 1999), simulates a city with a population of consumers comprised of two groups: the red H-group people and the green L-group people. The city has two competing banks. One bank uses a Group-Blind algorithm to make decisions about which applicants get loan offers. The other bank uses a Group-Blind algorithm. Please refer to subsection "The Algorithm's Decision" with the "CONCEPTUALIZING ALGORITHMIC DISCRIMINATION" section of the paper for more information on how populations groups and algorithms types are defined in this research.

Consumers who want to apply for a loan make choose a bank that maximizes their utility. Their choice, modeled by a multinomial logit, has a utility function that is described by Equation (7) and Equation (8). Their choice set contains the Group-Aware bank, the Group-Blind bank, or an outside option (another unknown bank). They assess the probability of acceptance through via word-of-mouth (WOM). The consumer's personal assessment of the probability she will be offered a loan depends on the information she gathers via WOM from her social network. She differentially weights WOM information based on the source (strength of social ties, in-group vs. out-group sources).

Timeline of events that initiates the ABM simulation (this happens once at the start of the simulation):

1. A Group-Blind bank and a Group-Aware bank are randomly placed in geographic locations in the ABM city

2. A large population of people is randomly "born" and distributed throughout geography of city.

3. Each person is endowed with randomly provided characteristics (e.g., latent quality,, credit score, group membership: H-group or L-group)

4. People are given a color according to group membership. Red people are H-group members and green people are L-group members.

5. For the empirical social network, those that are H-group and L-group members is pre-determined by the data. For synthetic networks (complete, random, and preferential attachment), the proportion of people who are H-group is an input in the model and is randomly drawn from a uniform distribution whose lower and upper bounds are based on empirical data.

6. A randomly selected application rate (percentage of people who decide to apply for a bank loan each period) is drawn from a uniform distribution whose lower and upper bounds are based on empirical data. This percentage will remain fixed through the rest of the simulation.

7. An ABM factorial design value is drawn from the ABM's permutation through design levels for high or low quality variance (one for each group) and high or low measurement error. These values will remain fixed through the rest of the simulation.

Timeline of events within each period of the ABM simulation

1. A randomly selected number of people (based on the application rate) decide to apply for a bank loan.

2. Each applicant selects one of 2 possible banks to apply for the loan or some other unknown bank (outside option)

   (a) Applicant gathers WOM information about each bank from her social network.

   (b) Applicant computes her own likelihood of receiving a loan offer from each bank based on gathered WOM. Strong ties and in-group WOM get greater weight than weak ties and out-group WOM. See Equation (8 for details.

   (c) Applicant computes her own utility for each bank in choice set. See Equation (7 for details.

(d) Applicant selects bank or chooses outside option based on a multinomial logit choice model.

3. The banks review each loan applicant's score. If a Group-Aware algorithm is reviewing, it also observes group-membership information. The algorithms form an expectation of applicant quality (interpreted as ability to repay loan) based on its own methods.

   (a) Group-Blind algorithm forms an expectation of applicant quality based on applicant score and a prior based on historical scores of all past applicants.

   (b) Group-Aware algorithm forms an expectation of applicant quality based on applicant score and a prior based on historical scores of all past applicants from the applicant's group.

4. Group-Aware algorithms offer loans to applicants whose scores exceed the minimum score threshold for the applicant's group at time t (S_{jt}^{min}). All other applicants are rejected.

5. Group-Blind algorithms offer loans to applicants whose scores exceed a single minimum score threshold for all applicant at time t (S_{t}^{min}). All other applicants are rejected.

6. Banks update the prior with estimates of parameters of the normal distribution of quality ($\hat{A}_{jt}$ and $\hat{\sigma}_{q_{jt}}^2$) with applicant's information. See Web Appendix D for details.

7. Banks update minimum score thresholds. See Equation (5) for details.

8. 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) For any bank, the applicant will update probability of loan acceptance based on gathered WOM information in the next period.

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

Web Appendix B: Additional Analysis of Agent-Based Model

Analysis of Monopoly Marketplace

The following table displays results from ABM simulations of a monopolistic marketplace with only one bank using one type of algorithm. We compare and contrast each type of algorithm (Group-Aware vs Group-Blind) by focusing on the variable of interest, $Group-Aware$. The results indicate that the Group-Aware algorithm is more profitable in a monopolistic market (see column 2).

Sensitivity Analysis of Distributional Assumptions

To examine whether the agent-based model is robust to different specifications of distributional assumptions about consumer quality (Q_{ij}), we conducted an analysis of a set of ABM simulations using the Copenhagen Networks empirical dataset. 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 7 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 ABM is robust to other distributional assumptions. Furthermore, all three models show statistically significant effects on the WOM-overall, WOM-ingroup (except for Log Normal), and score validity variables.

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 the ABM with alternative social network structures. Although our primary model's social network structure is based on the Copenhagen Networks

Table 6: Algorithmic Bias and WOM: Comparisons of Monopolies

	Dependent variable:	
	Applications	Profits
	(1)	(2)
Intercept	-26.3^{***}	$-6,741.91^{***}$
	(1.3)	(205.53)
WOM-overall	0.04	8.77^{*}
	(0.03)	(4.59)
WOM-ingroup	0.3^{*}	43.67
	(0.2)	(31.01)
Score Validity-H-group	-0.000	$2,468.28^{***}$
	(1.0)	(167.18)
Score Validity-L-group	0.002	$5,952.63^{***}$
	(1.0)	(167.18)
Application Rate	$47,199.4^{***}$	$4,113,960.30^{***}$
	(16.8)	(2,723.12)
Distance to Bank	-10.6^{***}	$-1,616.35^{***}$
	(2.7)	(430.09)
Group-Aware Bank	0.001	388.02^{***}
	(0.3)	(52.52)
Observations	11,520	11,520
R^2	0.995	0.995
Adjusted R^2	0.995	0.995
Residual Std. Error (df = 11512)	17.4	2,818.6
F Statistic (df = 7; 11512)	$1,136,418.1^{***}$	$330,478.3^{***}$

Note: Applications and Profits are cumulative. *p<0.1; **p<0.05; ***p<0.01

Table 7: Comparison of Distributional Assumptions for Consumer Quality

	Dependent variable:		
	Net Applications		
	Normal	Log Normal	Uniform
	(1)	(2)	(3)
Intercept	-127.1^{***}	-107.4^{***}	-11.7
	(23.8)	(24.9)	(23.3)
WOM-overall	3.7^{***}	3.1^{***}	2.4^{***}
	(0.4)	(0.4)	(0.4)
WOM-ingroup	-5.2^{*}	-3.5	-15.3^{***}
	(2.8)	(2.9)	(2.7)
Score Validity-H-group	-83.0^{***}	-32.8^{**}	-84.7^{***}
	(14.9)	(15.6)	(14.6)
Score Validity-L-group	38.8^{***}	-4.8	44.8^{***}
	(14.9)	(15.6)	(14.6)
Application Rate	$2,485.3^{***}$	941.7^{***}	$2,440.1^{***}$
	(243.0)	(254.3)	(237.6)
Distance to Bank-Group Aware	866.8^{***}	863.2^{***}	814.5^{***}
	(38.8)	(40.6)	(38.0)
Distance to Bank-Group Blind	-597.6^{***}	-587.6^{***}	-789.6^{***}
	(40.1)	(41.9)	(39.1)
Observations	5,760	5,760	5,760
R^2	0.1	0.1	0.2
Adjusted R^2	0.1	0.1	0.2
Residual Std. Error (df = 5752)	177.9	186.1	173.9
F Statistic (df = 7; 5752)	132.1^{***}	102.6^{***}	155.1^{***}

Note: Net Applications = Group-Blind - Group-Aware Applications *p<0.1; **p<0.05; ***p<0.01

Facebook empirical dataset (Sapiezynski et al. 2019), other network structures may produce results that differ in outcome. For this reason, we test robustness of the model 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 social network analysis and marketing literature (Rand and Rust 2011). Tables 8 and 9 present results from this analysis using OLS regression. As indicated in the tables, our model of algorithmic discrimination's impact on demand and profits is generally robust to network structure specification. All four network specifications are qualitatively consistent with each other. Furthermore, three out of the four models show statistically significant effects with positive signs on the overall WOM parameter ϕ and negative signs on the in-group WOM parameter ψ. Preferential Attachment is not statistically significant, but we find significance and consistence with the other models when the model is run on a longer time span (100 steps). This analysis suggests that in many social network structures, the weight of overall WOM in the utility function of the consumer leads to greater long-run Group-Blind algorithm profits than Group-Aware. Greater strength of weight on in-group WOM can attenuate the effect.

Table 8: Algorithmic Discrimination Impact on Demand: Network Structure Comparison

	Dependent variable: Net Applications			
	Empirical	Complete	Random	Pref. Attach
	(1)	(2)	(3)	(4)
Intercept	-127.1^{***}	949.6^{**}	8.5	-10.4
	(23.8)	(394.2)	(25.8)	(11.4)
WOM-overall	3.7^{***}	6.7^{***}	2.1^{***}	0.4^{*}
	(0.4)	(1.6)	(0.4)	(0.2)
WOM-ingroup	-5.2^{*}	-30.7^{***}	-5.0^{*}	1.3
	(2.8)	(10.3)	(2.9)	(1.3)
Score Validity-H-group	-83.0^{***}	-387.2^{***}	-177.4^{***}	-75.7^{***}
	(14.9)	(56.4)	(16.0)	(7.0)
Score Validity-L-group	38.8^{***}	234.8^{***}	149.7^{***}	51.4^{***}
	(14.9)	(56.4)	(16.0)	(7.0)
H-Group % of Population		141.2^{**}	23.5	21.3^{***}
		(56.7)	(16.1)	(7.1)
Application Rate	$2,485.3^{***}$	$4,864.1^{***}$	$2,353.2^{***}$	905.5^{***}
	(243.0)	(930.7)	(263.2)	(116.1)
Distance to Bank-Group Aware	866.8^{***}	1,047.7	815.1^{***}	409.5^{***}
	(38.8)	(1,068.8)	(43.0)	(18.5)
Distance to Bank-Group Blind	-597.6^{***}	$-4,585.9^{***}$	-890.2^{***}	-400.9^{***}
	(40.1)	(1,034.3)	(40.6)	(18.2)
Observations	5,760	5,760	5,760	5,760
R^2	0.1	0.02	0.2	0.2
Adjusted R^2	0.1	0.02	0.2	0.2
Residual Std. Error	177.9 (df = 5752)	672.1 (df = 5751)	190.3 (df = 5751)	83.9 (df = 5751)
F Statistic	132.1^{***} (df = 7; 5752)	15.5^{***} (df = 8; 5751)	139.8^{***} (df = 8; 5751)	148.8^{***} (df = 8; 5751)

Note: Net Applications = Group-Blind - Group-Aware Applications *p<0.1; **p<0.05; ***p<0.01

Table 9: Algorithmic Discrimination Impact on Profits: Network Structure Comparison

	Dependent variable: Net Profits			
	Empirical	Complete	Random	Pref. Attach
	(1)	(2)	(3)	(4)
Intercept	$-9,498.64^{***}$	$61,433.91^{***}$	561.88	-215.56
	(2,079.35)	(22,209.02)	(1,474.54)	(712.60)
WOM-overall	266.26^{***}	414.53^{***}	78.33^{***}	13.25
	(35.79)	(88.39)	(25.37)	(12.22)
WOM-ingroup	-574.88^{**}	$-2,118.57^{***}$	-356.93^{**}	77.14
	(241.87)	(581.39)	(167.08)	(80.51)
Score Validity-H-group	$-4,339.51^{***}$	$-20,371.51^{***}$	$-8,034.41^{***}$	$-1,229.26^{***}$
	(1,302.66)	(3,175.91)	(913.66)	(440.45)
Score Validity-L-group	1,929.99	$14,024.67^{***}$	$7,473.81^{***}$	$1,005.07^{**}$
	(1,302.66)	(3,175.91)	(913.66)	(440.34)
H-Group % of Population		$12,567.52^{***}$	1,423.97	87.15
		(3,195.59)	(919.06)	(442.73)
Application Rate	$152,355.30^{***}$	$275,390.70^{***}$	$79,513.79^{***}$	3,672.91
	(21,218.29)	(52,433.15)	(15,061.43)	(7,256.61)
Distance to Bank-Group Aware	$73,914.89^{***}$	33,867.53	$45,598.74^{***}$	$22,289.99^{***}$
	(3,390.69)	(60,210.82)	(2,458.09)	(1,154.70)
Distance to Bank-Group Blind	$-53,288.93^{***}$	$-273,667.80^{***}$	$-48,744.43^{***}$	$-22,031.21^{***}$
	(3,497.32)	(58,268.62)	(2,326.07)	(1,135.33)
Observations	5,760	5,760	5,760	5,760
R^2	0.13	0.02	0.14	0.12
Adjusted R^2	0.12	0.02	0.14	0.12
Residual Std. Error	15,529.53 (df = 5752)	37,861.14 (df = 5751)	10,892.05 (df = 5751)	5,248.49 (df = 5751)
F Statistic	117.78^{***} (df = 7; 5752)	17.00^{***} (df = 8; 5751)	118.21^{***} (df = 8; 5751)	95.94^{***} (df = 8; 5751)

Note: Net Profits = Group-Blind - Group-Aware Profits

*p<0.1; **p<0.05; ***p<0.01

Web Appendix C: Sensitivity Analysis Regarding Time Horizon

Table 10: Time Period When Group-Blind Exceeds Group-Aware

		Percentage of	Time Periods			
	Description	**Simulations**	**Mean**	**Std. Dev.**	**Min**	**Max**
Demand	Group-Blind demand surpasses Group-Aware demand	55.95%	4.92	8.60	1	60
	Group-Blind H-Group demand surpasses Group-Aware H-Group demand	52.74%	4.62	8.09	1	60
	Group-Blind L-Group demand surpasses Group-Aware L-Group demand	61.61%	5.91	9.48	1	60
Profits	Group-Blind profit surpasses Group-Aware profit	54.36%	3.86	7.10	1	60
	Group-Blind H-Group profit surpasses Group-Aware H-Group profit	53.70%	4.04	7.34	1	60
	Group-Blind L-Group profit surpasses Group-Aware L-Group profit	55.68%	4.63	8.13	1	60
Note:	Based on 5,760 ABM simulations. ABM social network is based on the Copenhagen Networks Facebook dataset.					

Web Appendix D: Derivations

The following are assumptions and derivations of equations discussed in the "CONCEPTUALIZING ALGORITHMIC DISCRIMINATION" section of the paper. The relationships between A_j, Q_{ij}, and S_{ij} are as follows:

$$Q_{ij} = A_j + \upsilon_{ij}, \quad \upsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{q_j}) \tag{10a}$$

$$S_{ij} \mid Q_{ij} = Q_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\varepsilon_j}), \text{ where } \upsilon_{ij} \perp\!\!\!\perp \varepsilon_{ij} \tag{10b}$$

$$S_{ij} \sim \mathcal{N}(A_j, \sigma^2_{q_j} + \sigma^2_{\varepsilon_j}) \tag{10c}$$

$$\text{where } A_H > A_L > 0 \tag{10d}$$

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:

$$P(A_j \mid \sigma^2_{q_j}, S_j) \sim \mathcal{N}\left(\frac{n_0 A_{j_0} + n_j \bar{S}_j}{n_0 + n_j}, \frac{\sigma^2_{q_j}}{n_0 + n_j}\right)$$

$$P(\sigma^2_{q_j} \mid S_j) \sim \mathcal{IG}\left(\frac{n_0 + n_j}{2}, \frac{n_0 \sigma^2_{q_{j0}} + n_j \bar{\sigma}^2_{q_j} + \frac{n_0 n_j}{n_0 + n_j}(A_{j_0} - \bar{S}_j)^2}{2}\right)$$

$$\text{where } \bar{S}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} S_{ij} \text{ and } \bar{\sigma}^2_{q_j} = \frac{1}{n_j}\sum_{i=1}^{n_j}(S_{ij} - \bar{S}_j)^2$$

$$\{A_{j_0}, \sigma^2_{q_{j0}}, n_0\} = \{\text{priors on } A_j, \text{ and } \sigma^2_{q_j}, \text{ and } n_j \text{ (number of data points in group } j)\text{, respectively}\} \tag{11}$$

The Group-Aware algorithm's estimates of the mean and variance of quality for group j,

learned from score training data, are derived from the expectations of the mean and variance the distributions in Equation (11):

$$\begin{aligned}\hat{A}_{jt} &= \frac{(N_{jt}-n_{jt})\hat{A}_{j_{t-1}}+n_{jt}\bar{S}_{jt}}{N_{jt}} \\ &= \hat{\sigma}^2_{q_{jt}} = \frac{(N_{jt}-n_{jt})\hat{\sigma}^2_{q_{jt-1}}+n_{jt}\bar{\sigma}^2_{q_{jt}}+\frac{(N_{jt}-n_{jt})\,n_{jt}}{N_{jt}}(\hat{A}_{j_{t-1}}-\bar{S}_{jt})^2}{N_{jt}} \\ \text{where } N_{jt} &= \sum_{t=1}^{t} n_{jt}\end{aligned} \tag{12}$$

The score validity has the following important properties:

$$0<\gamma_j<1, \quad \frac{\partial\gamma_j}{\partial\hat{\sigma}^2_{q_j}} = \frac{\sigma^2_{\varepsilon_j}}{(\hat{\sigma}^2_{q_j}+\sigma^2_{\varepsilon_j})^2} > 0, \text{ and } \frac{\partial\gamma_j}{\partial\sigma^2_{\varepsilon_j}} = \frac{-\hat{\sigma}^2_{q_j}}{(\hat{\sigma}^2_{q_j}+\sigma^2_{\varepsilon_j})^2} < 0 \tag{13}$$

Let p and $(1-p)$ 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 Group-Blind algorithm's estimates of mean quality, variance in quality, and score validity are as follows:

$$\hat{A}_t = p\hat{A}_{Ht} + (1-p)\hat{A}_{Lt} \tag{14a}$$

$$\hat{\sigma}^2_{q_t} = p\;\hat{\sigma}^2_{q_{Ht}} + (1-p)\;\hat{\sigma}^2_{q_{Lt}} + p(1-p)(\hat{A}_{Ht}-\hat{A}_{Lt})^2 \tag{14b}$$

$$\hat{\gamma}_t = \frac{\hat{\sigma}^2_{q_t}}{\hat{\sigma}^2_{q_t}+\sigma^2_{\varepsilon}} \tag{14c}$$

$$S_t^{min} = Q^{min} + \left(Q^{min}-\hat{A}_t\right)\left(\frac{1-\hat{\gamma}_t}{\hat{\gamma}_t}\right) \tag{14d}$$