# How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity

Jinyan Zang

Highlights

- This study examines racial and ethnic biases in the targeting of ads on Facebook's advertising platform in January 2020 and January 2021, before and after a major July 2020 boycott of Facebook by advertisers over issues of misinformation and civil rights.
- In 2021, Facebook's new "African-American Culture" ad targeting option contained 75% fewer White users than the old "African American (US)" option removed the previous year.
- Facebook's tools to help advertisers find users similar to their existing customers generated biased target lists that included either more African-Americans or more Whites, depending on which racial group was dominant in an advertiser's submitted customer list. This was true for both the Lookalike Audience tool and the Special Ad Audience tool that Facebook designed to explicitly not use sensitive demographic attributes when finding similar users.
- The degree of bias toward African-Americans or Whites in the composition of a Lookalike or Special Ad audience was greater when customer lists of individuals with racially stereotypical names or ZIP codes were used as the basis for each tool.
- Similarly, lookalike audiences can also be biased toward Asians. An audience generated by using a customer list of Asians with stereotypical names and ZIP codes as the source list was 100% Asian. Lookalike audiences based on source lists of Hispanics also overrepresented Hispanics.

Sample share of African-American voters in Lookalike Audiences based on lists of NC voters with different traits
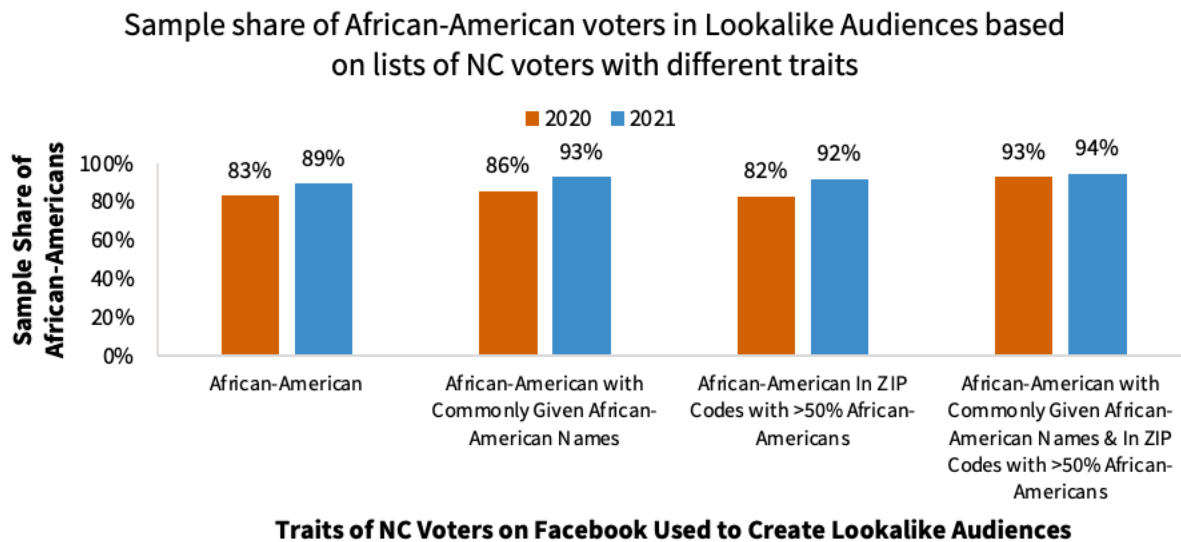
As more stereotypically African-American attributes were added to a customer list used to create a Lookalike audience of North Carolina voters, the sample shares of African-American voters in the corresponding Lookalike audiences offered to Facebook advertisers became more biased, increasing to 93% in 2020 and 94% in 2021.

Abstract

Over the last 5 years, Facebook has faced repeated criticism and lawsuits over the potential for discrimination on its ad platform. In July 2020, advocacy groups focused on issues of misinformation and civil rights organized a high-profile boycott of Facebook's advertising platform that successfully pressured more than a thousand major corporations to stop advertising on Facebook for the month. Facebook responded by releasing a civil rights audit and announcing the removal of its much-criticized multicultural affinity groups, such as "African American (US)", "Asian American (US)", and "Hispanic (US – All)", as ad targeting options. This study examines whether Facebook has gone far enough to prevent discrimination on its advertising platform. I collected data on Facebook's ad platform in January 2020 and in January 2021. I compared the racial and ethnic breakdown of the old multicultural affinity groups with the similar-sounding cultural interest groups "African-American Culture", "Asian American Culture", and "Hispanic American Culture," which were still usable by advertisers in 2021. I also used a set theory approach to study racial and ethnic biases in Facebook's Lookalike Audience and Special Ad Audience algorithms in both time periods.

Results summary: I found that in 2021 Facebook's "African-American Culture" ad targeting option produced audiences that contained 75% fewer White users than the old "African American (US)" option removed in the previous year. Facebook's tools to help advertisers find users similar to their existing customers exhibited bias toward including more African-Americans or Whites depending on which racial group was dominant in an advertiser's customer list. This was true for the Lookalike Audience tool as well as the Special

Ad Audience tool that Facebook created to explicitly not use sensitive demographic attributes when finding similar users. The degree of bias toward including more African-Americans or Whites in a Lookalike or Special Ad audience was larger when the advertiser submitted customer lists of individuals with racially stereotypical names or ZIP codes as the basis for each audience. There was also significant bias toward Asians in Lookalike audiences based on Asians, reaching up to 100% Asian in one case. Finally, Lookalike audiences based on Hispanics overrepresented Hispanicss. This study shows that in 2021, advertisers were able to use Facebook's ad platform to discriminate by race and ethnicity by using cultural interest groups as targeting options or by using the Lookalike and Special Ad Audience tools. It also provides evidence that "fairness through unawareness", or trying to prevent discrimination by eliminating the use of protected class variables or close proxies in a model, does not reduce the potential for algorithmic bias.

Introduction

Facebook is the second largest digital advertising platform, behind Google, in the US today [1]. It offers advertisers multiple tools to target their ads on the platform.

- An advertiser can use Facebook's own "Detailed Targeting" options, which are categories of users that share the same demographic attributes, interests, or behavior based on Facebook's analysis of its user data.

- Facebook also allows advertisers to create a "Custom Audience," a list of customers or individuals that the advertiser already has data on, to use as a basis for Facebook audience targeting.

- Facebook also offers to create a "Lookalike Audience", based on an advertiser's existing Custom audience, by finding the users whom Facebook identifies as most similar to the ones currently in the Custom audience.

- Finally, for ads related to housing, employment, and credit, Facebook can create a "Special Ad Audience" which is like a Lookalike audience, except that Facebook does not use sensitive attributes such as "age, gender or ZIP code" in considering which users are similar enough to include [2].

Over the last five years, Facebook has faced repeated criticism, lawsuits, and controversies over the potential for discrimination on its ad platform. Journalists have demonstrated how easy it is to exclude users whom Facebook classifies as being in racial or ethnic affinity groups from target groups for housing or employment ads [3, 4]. Researchers have demonstrated racial and ethnic biases in Facebook's Lookalike Audience and Special Ad Audience algorithms [5]. Facebook has been sued by the National Fair Housing Alliance [6], the ACLU [7], the Communications Workers of America [8], the U.S. Department of Housing and Urban

Development [9], and others [10] for violating civil rights laws such as the Fair Housing Act and the Civil Rights Act of 1964 by enabling discrimination on its advertising platform. There is also ongoing controversy over how Facebook's platform can be used by political actors, both foreign and domestic, to spread misinformation and target racial and ethnic minorities in the 2016 and 2020 election cycles [11, 12, 13, 14, 15].

In July 2020, a high-profile boycott of Facebook's advertising platform over issues of misinformation and civil rights was organized by advocacy groups including the NAACP, the Anti-Defamation League, Color of Change, and others. The activists called on major corporations to stop advertising on Facebook for the month [16]. More than 1,000 large companies, including Microsoft, Starbucks, and Target, participated in the boycott [17].

On July 8, 2020, Facebook released its own civil rights audit, conducted by Laura Murphy, former Director of the ACLU Legislative Office, and attorneys at the law firm Relman Colfax [18]. The audit criticized Facebook for having "placed greater emphasis on free expression" than on the "value of non-discrimination" [18]. Facebook Chief Operating Officer Sheryl Sandberg stated that the audit "is the beginning of the journey, not the end" [19].

As a concrete step on that journey, on August 11, 2020, Facebook announced that it would retire its controversial "multicultural affinity" groups, which allowed advertisers to target users whom Facebook categorized as "African American (US)", "Asian American (US)" or "Hispanic (US – All)" [20].

Has Facebook gone far enough to prevent discrimination on its advertising platform?

Advertising is commonly targeted to enable sellers to reach the desired buyers for their product; for example, textbooks are marketed to students. This study seeks to reveal to what degree Facebook's ad platform supports a particular kind of targeting, discrimination by race or ethnicity, through its tools. Such discrimination may or may not be the intended goal of a Facebook advertiser. If Facebook helps advertisers discriminate by protected classes like race and ethnicity, then that would violate existing U.S. civil rights laws including the Fair Housing Act, the Civil Rights Act of 1964, and the Equal Credit Opportunity Act [49].

I studied this question by testing Facebook's advertising platform in two waves, first in January 2020 and again in January 2021. While in 2021 Facebook no longer offered multicultural affinity groups as targeting options for advertisers, I tested the similar-sounding cultural interest groups that Facebook still offered as targeting options, such as "African-American Culture", "Asian American Culture", and "Hispanic American Culture". I also tested whether Facebook's other advertising tools, such as Lookalike Audiences and Special Ad Audiences, can discriminate by race and ethnicity by creating biased target audiences that over-represent a specific demographic group. Thus, this study examined the following questions about Facebook's advertising platform:

- Are the cultural interest groups as racially and ethnically homogeneous as the old multicultural affinity groups?

- Do Lookalike and Special Ad audiences reflect the racial and ethnic biases of the lists of individuals used to generate them?

- Is the degree of racial and ethnic bias in Lookalike and Special Ad audiences affected by well-established racial proxies from the offline world, such as the name or ZIP code of individuals used to create the audience?

- How did the type and degree of bias observed in Facebook's advertising tools change between 2020 and 2021?

Background

The Rise of Digital Advertising and Microtargeting

In 2019, digital advertising spending ($129 billion) in the US eclipsed traditional advertising ($109 billion) for the first time Facebook accounted for 22% of all digital ad spending, second only to Google's 37% market share [1]. Regulatory agencies, advocacy groups, the media, and researchers are increasingly focused on the need to prevent advertising discrimination on the country's second most popular digital advertising platform.

Facebook's advertising platform provides multiple targeting tools to help an advertiser "micro target" only the users they want. These tools are the focus of this study.

- "Detailed Targeting" options allow an advertiser to target a prepackaged group of Facebook users who share common attributes based on Facebook's data analysis of the ads they click, the pages they engage with, the activities they conduct on its websites, and other data. "Detailed Targeting" options are organized into three categories: demographics, interests, and behaviors (Figure 1).

Figure 1. Adding Detailed Targeting Option to a Facebook Advertising Campaign.

- "Custom Audiences" allow an advertiser to either upload a contact list of customers or other individuals for Facebook to target or create an audience by integrating Facebook's trackers on their websites or apps. When an advertiser uploads a customer list, Facebook then matches data fields such as email, phone number, first name, last name, city, state, country, ZIP code, date of birth, gender, and more with the data it has on its users. Choosing a Custom audience means targeting ads only to Facebook users that match an advertiser's customer list (Figure 2). Facebook can also create Custom audiences based on which Facebook users have interacted with an advertiser's content on the Facebook platform by an act such as commenting on a video, liking an Instagram post, or attending a Facebook event (Figure 3).

Figure 2. Identifiers Used to Match an Advertiser's Customer List with Users on Facebook.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/
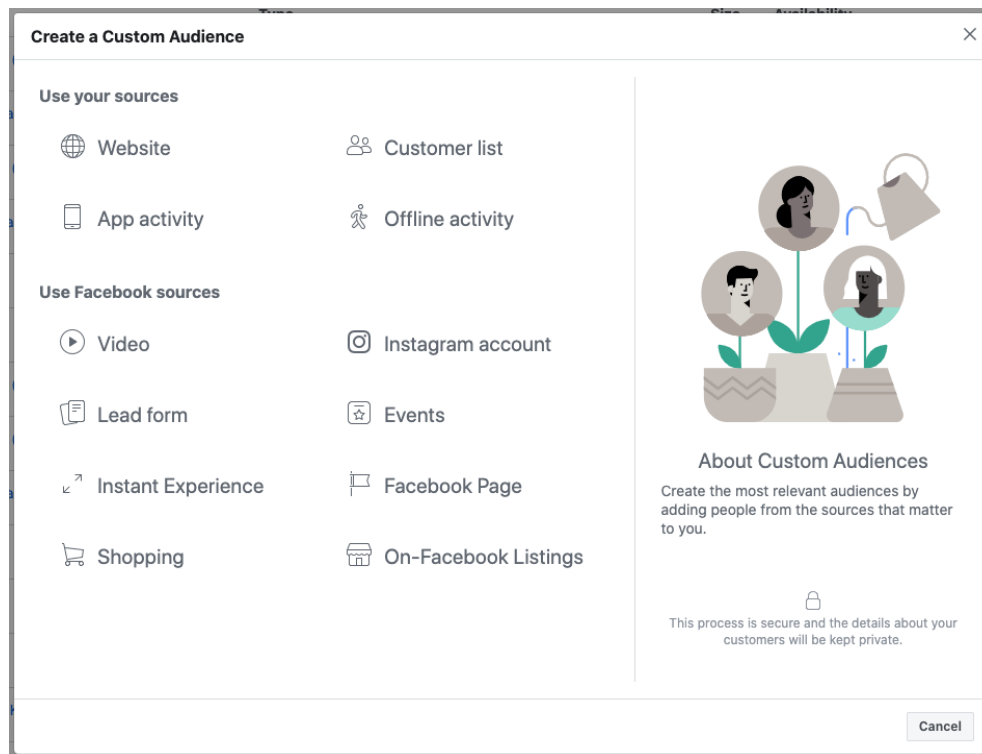
Figure 3. Sources for Custom Audience on Facebook.

- "Lookalike Audiences" allow an advertiser to reach people who resemble members of a designated source audience created through the Custom Audiences tool. According to Facebook, "You choose a source audience, and we identify the common qualities of the people in it. Then we find people like them, using your selected location and desired audience size." The advertiser can choose one or multiple countries as the "Audience Location", and Facebook will find the 1-10% of the population of Facebook users in those countries that are most similar to the advertiser's source audience (Figure 4).

Figure 4. Creating a Lookalike Audience.

- "Special Ad Audiences" allow an advertiser to create a Lookalike Audience that finds people similar to a source audience "in online behavior without considering things like age, gender or ZIP code"[2] specifically for ads in the categories of housing, employment, and credit, which are regulated by anti-discrimination laws (Figure 5). The settings for creating a Special Ad Audience are similar to the Lookalike Audience settings. The advertiser chooses a source audience, then selects one or multiple countries as the "Audience Location", and finally asks Facebook to find the 1-10% of the population of Facebook users in those countries who are most similar to the advertiser's source audience (Figure 6).
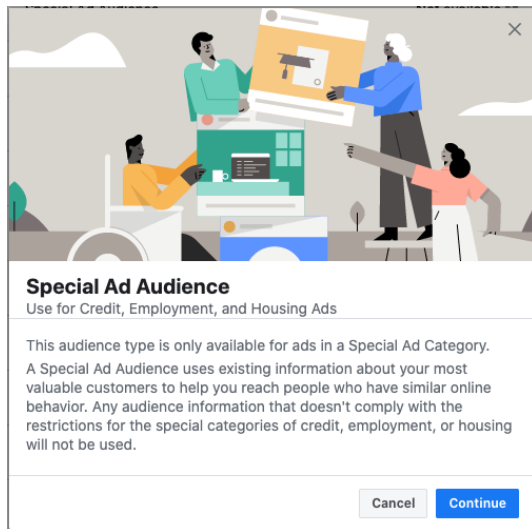
Figure 5. Special Ad Audience Tool.

Figure 6. Creating a Special Ad Audience.

An advertiser can use the tools offered on Facebook's ad platform to define a desired intersection of user attributes by using Facebook's Audience selection tool (Figure 7). First, the advertiser can choose one or multiple Custom, Lookalike, or Special Ad audiences to "include" as potential targets. Advertisers can also choose one or multiple Custom or Lookalike audiences to exclude as potential targets, though they are not able to choose to exclude a Special Ad audience. Next, in terms of geographical targeting options, the advertiser

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

can choose to limit the audience to individuals currently living in, recently located in, or traveling to a particular country, state, city, ZIP code, media market, or Congressional District. Another option is to drop a pin on the map and target all individuals within f 1 to 50 miles of the pin's location. The advertiser also has options to target by age, gender, and language. The advertiser can choose to add one or more of the predefined "Detailed Targeting" options based on demographics, interests, and behaviors created by Facebook to include or exclude as criteria for their target audience. There are also additional options for an advertiser to include or exclude users who have connections with the advertiser's Facebook pages, apps, or events.

Figure 7. Facebook's Audience selection tool.

Digital advertising platforms like Facebook give advertisers abilities to target individuals, offering options that traditional advertising does not provide or cannot offer in one place.

First, with traditional print or broadcast advertising, the advertiser will reach anyone who reads the newspaper or watches the TV show that displays the ad. On Facebook, the advertiser can target specific individuals they want to reach through Facebook's Custom Audiences or match a very specific intersection of targeting criteria, combining demographics, interests, behaviors, and more (Figure 7).

Second, with physical mail or email advertising, the advertiser can usually only reach those whose physical or email address they already have. On Facebook, however, the advertiser can potentially target millions of new individuals similar to existing customers without needing to know contact information for this new audience. Facebook offers this capability through its Lookalike Audiences or Special Ad Audiences tools.

Facebook and Race

The increased targeting capabilities of Facebook's advertising platforms also present new risks for advertising to have a discriminatory result.

Given this potential for discrimination, it is important to note that Facebook does not ask users to indicate their race. However, until August 11, 2020, the Facebook Ads platform did infer a user's "Multicultural Affinity" based on their behavior on Facebook and how similar it is to others of the same affinity group. Based on this inference Facebook offered advertisers the ability to target by Multicultural Affinity classifications for "African American", "Asian American", and "Hispanic", with the following Hispanic sub-categories: "Hispanic – Bilingual", "Hispanic – Spanish Dominant", and "Hispanic – English Dominant".

According to Facebook:

> The word "affinity" can generally be defined as a relationship like a marriage, as a natural liking, and as a similarity of characteristics. We are using the term "Multicultural Affinity" to describe the quality of people who are interested in and likely to respond well to multicultural content. What we are referring to in these affinity groups is not their genetic makeup, but their affinity to the cultures they are interested in…. The Facebook multicultural targeting solution is based on affinity, not ethnicity. This provides advertisers with an opportunity to serve highly relevant ad content to affinity-based audiences [21].

While Facebook argued that these multicultural affinity classifications did not facilitate racial and ethnic discrimination by advertisers, it faced repeated criticism from journalists, researchers, civil rights groups, and law enforcement agencies over this issue.

Media Attention

In October 2016, ProPublica journalists Julia Angwin and Terry Parris Jr. found that "Facebook lets advertisers exclude users by race" [3]. Advertisers could target or exclude users based on Facebook's "Ethnic Affinity" categories, which included African American, Asian American, and Hispanic. In order to demonstrate the possibility of discrimination, ProPublica ran its own housing-related ads on Facebook and used these categories to exclude users of a given ethnic affinity (Figure 8). Facebook responded that ethnic affinity categories existed as part of their "multicultural advertising" effort and that ethnic affinity was not the same as race but rather a membership category that Facebook created based on the pages and posts a user liked or engaged with on its website [3].
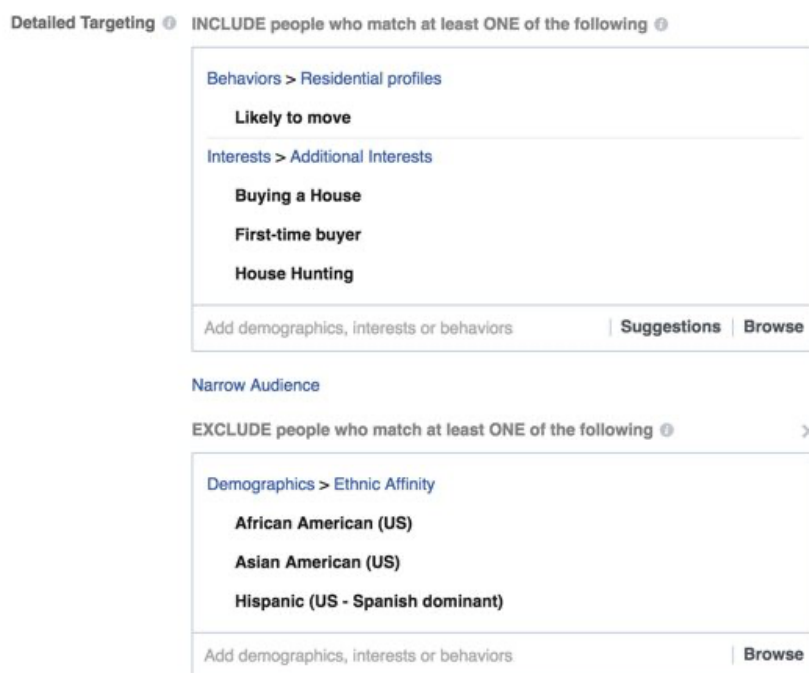
Figure 8. ProPublica's example in October 2016 of using the "Ethnic Affinity" categories within Facebook's Detailed Targeting options to potentially exclude minority users from seeing a housing-related ad [3].

One year after their initial reporting, ProPublica journalists reported in a November 2017 story, "Facebook (Still) Letting Housing Advertisers Exclude Users by Race" [4], that discrimination in housing advertising was still allowed on the platform. ProPublica was able to purchase housing-related ads that excluded users on the basis of multicultural affinity groups, as well as mothers of high school kids, people interested in wheelchair ramps, Jews, expats from Argentina, and Spanish speakers. These categories appear related to "protected classes" in anti-discrimination law. Facebook approved every ad proposed by ProPublica

within three minutes. Facebook acknowledged that "This was a failure in our enforcement and we're disappointed that we fell short of our commitments" [4]. ProPublica noted that Facebook had renamed "Ethnic Affinity" as "Multicultural Affinity" and moved it from the Demographics targeting options to the Behaviors list. Since Facebook also lets advertisers select which ZIP codes to target, ProPublica could also target ads at majority non-Hispanic White ZIP codes in Brooklyn. This demonstrated a practice ProPublica described as similar to "redlining" [4], a set of historical discriminatory practices used by landlords, brokers, and lenders to prevent African-Americans from moving to predominantly White neighborhoods. Redlining is now prohibited by the Fair Housing Act.

Academic Research

Academic researchers have also found discriminatory potential in Facebook's other targeting tools, including Lookalike Audiences and Special Ad Audiences. A December 2019 study found that Lookalike and Special Ad audiences could be significantly biased by the demographics of the source audience [5]. For example, when the source audience was all women, the Lookalike audience was 96.1% women, and the Special Ad audience was 91.2% women. Similar relationships were also observed for racial groups. For example, a source audience that was 100% Black created a Lookalike audience which had a 61% overlap with a list of 900,000 Black voters and only a 16% overlap with a second list of 900,000 White voters in the same state. The corresponding Special Ad audience had a 62% overlap with the Black voter list and a 12% overlap with the White voter list. On the other hand, a Lookalike Audience based on a source audience that was 100% White had a much smaller overlap of 17% with the Black voter list and a 42% overlap with the White voter list. Similarly, the corresponding Special Ad Audience had a 10% overlap with the Black voter list and a higher 36% overlap with the White voter list.

Legal Actions Taken Against Facebook and Its Advertisers

Facebook's policies and practices have been challenged in numerous lawsuits alleging illegaladvertising discrimination. On November 3, 2016, a class action lawsuit, *Mobley v. Facebook*, filed in U.S. District Court, argued that Facebook violated federal anti-discrimination laws for housing (Fair Housing Act) and employment (Civil Rights Act) by citing the ProPublica reporting [22, 23]. On March 27, 2018, a coalition of housing advocacy groups led by the National Fair Housing Alliance (NFHA) filed a lawsuit against Facebook for violating the Fair Housing Act by allowing advertisers to discriminate against legally protected groups such as mothers, the disabled, and Spanish-language speakers [24]. The NFHA conducted its own investigation of Facebook's advertising platform and found that it could exclude individuals that Facebook classified as "Disabled American Veterans", "moms of preschool kids", and individuals interested in "English as a second language" [6]. On September 18, 2018, the ACLU, the Communications Workers of America (CWA), and the employment law firm Outten & Golden LLP filed charges with the Equal Employment Opportunity Commission (EEOC) against Facebook and ten major corporations that targeted

ads for jobs to younger male Facebook users only, excluding all women and older users [7]. In a separate 2018 lawsuit filed by the Communications Workers of America vs. T-Mobile, Amazon, and 1,000 other large employers using Facebook ads, the plaintiffs allege that not only did the employers use Facebook's prepackaged targeting options in a discriminatory manner, but they also used Facebook's Lookalike Audience tool to target candidates demographically similar to their existing workforce in ways that marginalized older workers [8, 25].

Regulatory agencies and other law enforcement offices have also investigated and filed suit against Facebook for advertising discrimination. In 2016, Washington State's Office of the Attorney General started a 20-month investigation of Facebook's advertising platform. The investigators purchased ads that excluded protected categories of people from being targeted and found real-world examples of ads that did just that [10, 26, 27]. Washington Attorney General Bob Ferguson said, "Facebook's advertising platform allowed unlawful discrimination on the basis of race, sexual orientation, disability and religion…That's wrong, illegal, and unfair" [26].

On March 28, 2019, the U.S. Department of Housing and Urban Development (HUD) sued Facebook, alleging it "unlawfully discriminates based on race, color, national origin, religion, familial status, sex, and disability by restricting who can view housing-related ads on Facebook's platforms and across the internet" and "mines extensive data about its users and then uses those data to determine which of its users view housing-related ads based, in part, on these protected characteristics" [9].

Political Controversy and Criticism from Civil Rights Advocacy Groups

Finally, while it is not unlawful in the United States to target political or news ads in a racially biased manner, the ability to use Facebook's advertising platform to spread misinformation or inflame racial tensions generated high-profile political controversy in the 2016 and 2020 General Elections.

Researchers studying the 3,519 ads that Facebook shared with Congress as part of investigations into Russian interference with the 2016 presidential election found that many of the ads focused on Black identity issues such as police shootings, Black Lives Matter, and discrimination [28]. The most popular Facebook targeting options used by Russian actors included targeting users interested in Martin Luther King, African-American Civil Rights Movement, African-American history, Black Power, and related categories [15]. In fact, 17 ads used Facebook's "African-American (US)" multicultural affinity group [15]. Another study found that across all the Russian-linked ads disclosed by Facebook, 52% had more than double the proportion of African-Americans in their target audience compared to Facebook's US baseline [29]. Besides the Russians, other political actors may have also used Facebook's advertising tools to target minority voters in controversial ways in 2016. For example, in October 2016, Trump Campaign Manager Brad Parscale announced that the Trump campaign

would target an ad featuring opponent Hillary Clinton disparaging young African-Americans as "superpredators", using Facebook's advertising platform to reach "only the people we want to see it", in order to suppress support for Clinton among African-American voters [30].

In 2020, similar controversies arose about misinformation by campaigns targeting minority voters on Facebook. For example, racial appeals and a focus on American social unrest during the summer of 2020 were part of misinformation campaigns by groups linked to Russia, Iran, and China on Facebook and Twitter [11, 12]. Domestic political actors also used these platforms to promote misleading or false ads [11, 13, 14]. In August 2020, *The Washington Post* reported that FreedomWorks, a conservative political advocacy group established by David and Charles Koch, spent $1,500 on paid Facebook posts using LeBron James' picture and falsely implying that James had criticized mail-in voting [14]. These ads targeted voters in swing states with high concentrations of minority voters [14]. Other political groups were able to exploit Facebook's fact-checking system to re-post and distribute nearly identical copies of ads that Facebook's fact-checking partners helped remove [13]. A significant amount of misinformation also targeted Hispanics in 2020, particularly Hispanic voters in Florida [31, 32, 33, 34]. According to Jaime Longoria, an investigative researcher with First Draft News, one tactic was to draw on "anti-blackness" bias to promote the idea that Black individuals were "harassing" Latinos under the guise of activism. One viral video that was shared 180,000 times on Facebook showed two Black women harassing a Latino family celebrating at a party. The video falsely labeled the women as members of Black Lives Matter [31]. Other misleading ads labeled presidential candidate Joe Biden a "communista" or stated that vice presidential candidate Kamala Harris supported abortion up to the minutes before birth [31, 32]. A common problem was that Facebook's fact-checking was not as robust or strongly enforced for Spanish-language ads and posts [33].

In response to the widespread misinformation on Facebook, often targeting minority users, on June 17, 2020, civil rights groups including the NAACP, the Anti-Defamation League, and Color of Change launched the "Stop Hate for Profit" campaign. The campaign pressured major corporate advertisers to stop advertising on Facebook for the month of July 2020 [35]. More than 1,000 major advertisers joined the boycott, including Microsoft, Starbucks, Unilever, and Target [17]. The organizers of the boycott recommended 10 actions for Facebook to take, including hiring a C-suite executive to review the company's products for discrimination, hate, and bias; participating in a regular third-party audit on identity-based misinformation and hate; stopping the amplification of content with ties to hate, misinformation, or conspiracies; ending the exemption of politicians from fact-checking; and other changes [35]. The organizers of the boycott argued that the prevalence of hate speech, white supremacy activism, and misinformation on Facebook reflects the systemic racism in America today and that Facebook's technologies and the company's laid-back approach to moderation further reinforce racism [36].

Racial Discrimination by Humans

Outside of Facebook, racial discrimination and racial bias can occur in many different situations in everyday life, from applying for a job or a loan to trying to get a doctor's appointment. In these cases, individuals or a group of individuals working within a larger organization make the biased decisions. Just as Facebook does not collect race directly as a variable from its users, discrimination by others in society may not rely on an explicit racial variable but rather on proxies for race.

One common proxy is an individual's name. Different racial groups tend to have distinct first and/or last names [37, 38]. Field experiments have shown that individuals with racially affiliated names can experience discrimination [39, 40]. For example, in 2003, Bertrand and Mullainathan found that resumes with White-sounding first names received 50% more callbacks for interviews than resumes with Black-sounding first names [39]. In 2016, Kang, DeCelles, et al. found a similar interview callback gap of 40% for James (a White-sounding first name) versus Lamar (a Black-sounding first name) when both resumes appeared "White", i.e. the resumes did not describe participation in ethnically affiliated groups, such as an African-American fraternity. They found no statistical difference in the callback rates when the resumes appeared "Black". When resumes appeared "Asian", Luke Zhang (White-sounding first name, Asian-sounding last name) received a statistically significant 83% more interview callbacks than Lei Zhang (Asian-sounding first and last name) [40]. A meta-analysis of these field experiments in 2017 found no change in African-American callback rates since 1989 but found a decline in discrimination against Latinos [41]. Other studies have looked at gaps in response rates for a housing inquiry, a mortgage application, a request for a doctor's appointment, a request for help from a public official, and more [42, 43, 44].

An individual's home address can also serve as a proxy in other forms of racial discrimination. Historical factors have resulted in high levels of racial segregation in many US cities [45]. Before the passage of the Fair Housing Act in 1968, lenders and the real estate industry commonly practiced "redlining", where residents of certain minority neighborhoods were denied access or offered inadequate access to affordable mortgages, and minority renters and borrowers were often prevented from moving to White-dominant neighborhoods [45]. While the Fair Housing Act made redlining and housing discrimination illegal, disparities in access to housing and borrowing still exist between racial groups [45]. For example, a 2015 study in Baltimore found that race is the most statistically significant factor in predicting who gets a mortgage, and that White borrowers were far more likely to get loans relative to the population baseline than African-American borrowers [46]. Similar patterns exist for 61 metro areas around the country [47]. Hanson and Hawley found in an audit study across the 10 largest US cities that African-Americans inquiring about rental vacancies faced lower response rates from landlords than Whites in racially mixed neighborhoods, areas with rent above the median rent for the city, and neighborhoods close to the city center or first-ring suburbs [44]. According to research by Raj Chetty, Nathaniel Hendren, et al., these disparities contribute to "substantially lower rates of upward mobility and higher rates of downward mobility" for African-Americans relative to Whites, "leading to large income disparities that persist across

generations" [48]. Chetty et al. found that fewer than 5% of Black children grew up in rich areas with a poverty rate below 10% while more than 63% of White children did [48]. Among researchers studying algorithmic bias, these findings have contributed to concerns over the inclusion of geographic variables such as ZIP codes as data inputs. Such variables may serve as proxies for race and contribute to algorithmic bias [49, 50].

The research summarized above documents ways that discrimination often occurs "offline". This study examines how it can occur online on sites like Facebook. I test whether the algorithms used by Facebook to create Lookalike and Special Ad audiences generate higher degrees of racial and ethnic bias when the advertising tools are given source audiences composed of individuals with racially and ethnically stereotypical names and ZIP codes.

Discrimination by Algorithms and Anti-Discrimination Laws

The Civil Rights Movement and its legislative successes resulted in three major federal laws that forbid racial and other types of discrimination in housing (Fair Housing Act), employment (Civil Rights Act of 1964), and credit (Equal Credit Opportunity Act), including in advertisements for those sectors [49]. These laws provide a legal recourse for victims of discrimination and a potential threat of legal repercussions to discriminatory actors. These are the laws that the ACLU, HUD, and others argue that Facebook and advertisers on Facebook's platform violated. However, as discrimination has moved from decisions made by humans to those by machines and algorithms, detecting discrimination and enforcing anti-discrimination laws has become more complicated.

Latanya Sweeney's 2013 study on Google's ad platform was an early, high-profile example of a digital audit study to detect discrimination. Sweeney found that statistically significantly more ads using Black-identifying first names contained the word "arrest" in the ad's text than ads using White-identifying first names, with an adverse impact ratio of 77% for Reuters.com, which had ads served by Google, and an adverse impact ratio of 40% for Google.com search results [51]. Other instances of algorithmic bias have been found in facial recognition systems [52], online shopping [53], search engines [54, 55, 56, 57], job sites and hiring software [58, 59, 60], translation services [61], healthcare [62], and other systems.

Since much of the work on algorithmic discrimination has focused on the products of private sector companies, the legal concepts of "disparate treatment" and "disparate impact" established by Title VII of the 1964 Civil Rights Act are the primary applicable jurisprudence [63].

Disparate treatment focuses on intentional discrimination, which occurs when individuals are treated differently because of their protected class attribute, such as race, color, national origin, religion, sex, disability, or familial status [63]. There can be overt evidence of disparate treatment, as when a lender has a policy of a higher credit limit for older borrowers [64]. Or there can be comparative evidence of disparate treatment, as when two borrowers who are

otherwise similar are treated differently by a lender on the basis of a protected class attribute [64].

Disparate impact occurs when there is a disproportionate burden in outcomes for a specific group. Showing intentionality is not required if a significant disparity can be demonstrated [63, 64]. The law uses a burden-shifting framework to decide whether a company is liable under disparate impact. First, a plaintiff must show evidence of disproportionate outcomes across demographic groups. The Equal Employment Opportunity Commission's (EEOC) 80-20 Rule is often cited by algorithmic bias literature, although it primarily applies to labor issues [65, 66]. In response, a defendant corporation can try to show a "business necessity" to justify its decision-making process [63, 64]. For example, if a job requires lifting heavy supplies, a business may argue that business necessity requires it to consider a candidate's ability to lift weights in its hiring process, even if that leads to hiring more male candidates. Finally, the plaintiff must show that there is a less discriminatory alternative that could meet the business necessity [63].

These concepts may be applied to the Facebook ad targeting tools. If an advertiser uses Facebook's targeting options in an intentionally discriminatory way, such as creating age limits or gender criteria, or choosing multicultural affinity targeting options like "African-American (US)," that may be considered disparate treatment. If an advertiser has a predominantly White customer base, they might then use either Facebook's Lookalike Audience or Special Ad Audience tools to create a new target audience of other users whom Facebook identifies as being similar to their existing customers. If the resulting target audience is racially biased by being predominantly White, under the disparate impact framework, it doesn't matter whether the advertiser or Facebook intentionally meant to discriminate. Plaintiffs challenging this choice would need to show that Black and other minority users were impacted disproportionately by not being targeted by the advertiser relative to White peers. The advertiser could argue that it was a business necessity to use Facebook's tools to find new ad audiences similar to their existing customers. Facebook might argue that business necessity required creating the best possible Lookalike or Special Ad audience to serve their advertising clients. Finally, the plaintiffs would need to show that there are less discriminatory alternatives that could still achieve the same business necessities without predominantly targeting more White users.

This study examines to what degree Facebook's racially affiliated targeting options lead to disparate treatment by race and ethnicity. This study also examines to what degree Facebook's Lookalike Audience and Special Ad Audience tools lead to disparate impact by race and ethnicity depending on the demographics of the source audience.

Facebook's Response to Charges of Discrimination

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

After being repeatedly accused of discrimination by journalists, advocacy groups, and law enforcement agencies, over time Facebook has become more responsive in making changes to its advertising platform

In response to the original ProPublica report from 2016 [3], Facebook announced efforts in February 2017 to icreate enforcement tools to disapprove ads that use multicultural affinity groups – the new name "Ethnic Affinity" – as targeting options when offering housing, employment, or credit opportunities. Facebook also required advertisers to self-certify that they are not discriminating [67].

On July 24, 2018, Facebook signed a legally binding agreement with the state of Washington to make changes to its advertising tools as a result of the 20-month investigation conducted by Washington Attorney General Bob Ferguson's office [10]. Facebook agreed to pay $90,000 in costs and fees to the Attorney General's Office and removed targeting options that may allow advertisers to exclude users on the basis of race, religion, sexual orientation, veteran status, and other protected classes from seeing ads for housing, employment, credit, and insurance ads [10]. On August 21, 2018, Facebook announced it would eliminate 5,000 targeting options related to ethnicity or religion, such as "Native American culture", "Passover", "Evangelicalism", and "Buddhism", from being used by advertisers [68, 69].

On March 19, 2019, Facebook settled with the ACLU, the Communications Workers of America, the National Fair Housing Alliance, and others on their multiple lawsuits against Facebook over advertising discrimination [70, 71]. Facebook agreed to create a separate portal for ads for housing, employment, and credit [72]. On this portal, Facebook agreed to offer "a much more limited set of targeting options so that advertisers cannot target ads based on Facebook users' age, gender, race, or categories that are associated with membership in protected groups, or based on zip code or a geographic area that is less than a 15-mile radius, and cannot consider users' age, gender, or zip code when creating 'Lookalike' audiences for advertisers" [72].

On August 26, 2019, Facebook started implementing this agreement by requiring advertisers purchasing employment, housing, and credit ads to use "Special Ad Categories" tools instead of the regular Facebook advertising tools [2]. The "Special Ad Categories" tools restrict options that "allow targeting by age, gender, ZIP code, multicultural affinity, or any detailed options describing or appearing to relate to protected characteristics." Advertisers must create a Special Ad Audience that finds users who are similar to their existing customers "in online behavior without considering things like age, gender or ZIP code" [2].

Before settling these lawsuits, Facebook repeatedly argued that it is not liable for discrimination on its platform that might result from the actions of its advertisers, because it is protected by Section 230 of the Communications Decency Act [26, 68]. According to Facebook, since it is a platform rather than a publisher, Section 230 means that it should not be liable for the content that it hosts [26]. In 2018, the U.S. Department of Justice filed

Statements of Interest in two discrimination-related lawsuits against Facebook, arguing that Section 230 liability protections do not apply to Facebook's advertising platform [73, 74].

Finally, in response to the July 2020 Stop Hate for Profit boycott organized by the NAACP, the Anti-Defamation League, Color of Change, and others, Facebook on July 8, 2020 released a civil rights audit conducted by Laura Murphy, former Director of the ACLU Legislative Office, and attorneys at the law firm Relman Colfax [18]. The audit focused primarily on issues of misinformation and hate speech on Facebook and how Facebook has "placed greater emphasis on free expression"rather than balancing that objective with the "value of non-discrimination", values that the report noted do not need to be mutually exclusive [18]. For example, the audit criticized Facebook for deciding that President Trump's post stating "when the looting starts, the shooting starts" – in regards to the May 2020 protests following George Floyd's death – did not violate its content policies about the incitement of violence and for leaving the post up without any warning labels [18]. Regarding advertising discrimination, the audit acknowledged that HUD had filed charges against Facebook for violating fair housing laws with its ad targeting options [18]. The audit also acknowledged that research from Northeastern University and Upturn [5] had shown it is possible for the Special Ad Audiences tool to produce a biased result even though it does not use protected class information [18]. Regarding this audit and Facebook's civil rights policies, Facebook's Chief Operating Officer, Sheryl Sandberg, said that "it is the beginning of the journey, not the end" [19]. On August 11, 2020, Facebook announced that it would finally retire its controversial multicultural affinity advertising targeting options for racial and ethnic groups [20].

The boycott by most major advertisers ended by August 2020, and Facebook does not appear to have suffered significant financial damage as a result of the boycott. The boycott coincided with the COVID-19 pandemic, which saw many small and medium businesses increase their web presence and online sales [75]. Besides the civil rights audit, Facebook also responded to some of the demands of the boycott's organizers, banning Holocaust denial posts and blackface posts [75] and agreeing to hire a civil rights vice president, though not a C-suite executive as demanded by the boycott's organizers [76].

At the end of 2020, Facebook also began to update some "race-blind" policies that had a disproportionate impact on marginalized groups. In December 2020, Facebook started re-engineering its automated moderation systems, which previously did not distinguish groups that historically were targets of hate speech from groups that were not [77]. Thus, comments such as "White people are stupid" were treated the same way as anti-Semitic or racist slurs [77]. Black users complained that the old system removed posts such as "Thank a Black woman for saving our country." Civil rights experts argued that "you can't have the conversation if it is being filtered out, bizarrely, by overly blunt hate speech algorithms"[77]. Facebook is de-prioritizing its moderation of negative comments about "Whites", "men", and

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

"Americans" as less likely to be harmful, while acknowledging that underrepresented groups need more protection [77].

This study examines changes in the discriminatory potential of Facebook's advertising platform between January 2020 and January 2021. It was a tumultuous year, which saw organized action by advocacy groups and by Facebook's users, and largest advertisers to apply pressure for greater civil rights protections on the platform. Facebook has also begun to publicly take some steps toward reform, such as removing multicultural affinity targeting options and re-engineering its content moderation system.

Has Facebook gone far enough to prevent discrimination on its advertising platform?

Methods

Overview

This study is the first to examine the racial and ethnic breakdowns of Facebook's multicultural affinity groups and the similar-sounding cultural interest groups. It is also the first to examine the racial and ethnic biases of Facebook's Lookalike Audience and Special Ad Audience tools across multiple time periods and attempt to determine whether the degree of bias is affected by populating a source audience with individuals with racially stereotypical names and ZIP codes.

I conducted the study over two waves, with Wave 1 occurring in January 2020 and Wave 2 in January 2021. During each wave, I conducted two types of tests of Facebook's ad platform to answer my research questions about potential racial and ethnic biases.

- Test 1 – Studying the Racial and Ethnic Breakdowns of Facebook Targeting Options

    - Detailed Targeting options

        - 2020 – Multicultural Affinity groups

            - African American (US)

            - Asian American (US)

            - Hispanic (US – All)

        - 2021 – Cultural Interest groups

            - African-American Culture

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

- - Asian American Culture

    - Hispanic American Culture

  - Research question: Are the cultural interest groups as racially and ethnically homogeneous as the old multicultural affinity groups?

- Test 2 – Studying the Bias in Facebook's Lookalike Audience and Special Ad Audience Tools Using A Set Theory Approach

  - Lookalike Audiences (2020 and 2021)

  - Special Ad Audiences (2020 and 2021)

  - Research questions:

    - Do Lookalike and Special Ad audiences reflect the racial and ethnic biases present in the lists of individuals used to generate them?

    - Is the degree of racial and ethnic bias in Lookalike and Special Ad audiences affected by well-established racial proxies from the offline world, such as the name or ZIP code of individuals used to create the audience?

    - What are the differences in the type and degree of bias observed in Facebook's advertising tools in 2021 versus 2020?

In each wave, I used that month's North Carolina voter list of ~8 million voters to generate source audiences to test Facebook's advertising tools. The North Carolina voter data is useful for this study, because it includes important personal information that Facebook can use to match a voter to a user profile. It also contains self-reported race and ethnicity data that I could use to create subsets of voters for each test.

This study tested how Facebook's ad platform relates to different samples of African-American, White, Asian, and Hispanic voters, which are racial and ethnic groups that exist in the voter data and align with Facebook targeting options. For each test, I first created lists of voters with known race and ethnic data. I then used those lists as source audiences for different Facebook ad tools to see if the estimated reach according to Facebook changed in a biased way based on the demographics of the list that was used and the Facebook tool being tested.

Preparing the North Carolina Voter Data for Testing on Facebook

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

The North Carolina voter dataset is publicly available and contains the following variables that overlap with what Facebook requests in order to match an uploaded list of voters with their corresponding Facebook profiles for ad targeting through the Custom Audience tool:

- First name

- Last name

- City

- State

- ZIP code

- Country

- Gender

- Age

- Year of birth

- Phone number

I filtered the voter data for "Active" and "Verified" voters to ensure that I used voters with the most up-to-date city and ZIP code fields to match against Facebook profiles.

Voter registration forms in North Carolina ask voters to indicate their race and ethnicity, which are captured in the state voter data (Figure 9). I used the African American/Black, Asian, and White race categories and the Hispanic and Not Hispanic/Latino ethnicity categories to create subsets of voters to test the degree of bias in Facebook's advertising platform.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

NORTH CAROLINA VOTER REGISTRATION APPLICATION (fields in red text are required)   2020.02   **06w**

1 Indicate whether you are qualified to vote or preregister to vote based on U.S. citizenship and age.

Are you a citizen of the United States of America?
IF YOU CHECKED "NO" IN RESPONSE TO THIS CITIZENSHIP QUESTION, DO NOT SUBMIT THIS FORM. YOU ARE NOT QUALIFIED TO VOTE   ☐ Yes ☐ No

Will you be at least 18 years of age on or before election day?   ☐ Yes ☐ No

Are you at least 16 years of age and understand that you must be 18 years of age on or before election day to vote?
IF YOU CHECKED "NO" IN RESPONSE TO BOTH OF THESE AGE QUESTIONS, DO NOT SUBMIT THIS FORM. YOU ARE NOT QUALIFIED TO REGISTER OR PREREGISTER TO VOTE.   ☐ Yes ☐ No

2 Provide your full legal name.

Last Name   Suffix

First Name

Middle Name

3 Provide your date of birth and identification information.

Date of Birth (MM/DD/YYYY)   State or Country of Birth

/   /

NC Driver License or NC DMV ID Number   Last 4 Digits of Social Security Number

☐ Check if you do not have a driver license or Social Security number.   State Voter Registration Number (Optional: To locate, check "Voter Lookup" at www.NCSBE.gov.)

4 Provide your residential address - where you physically live.
Do not enter a P.O. Box or a mail drop location.

Address Number   Street Name and Type

Address Line 2 (e.g., apartment, lot or unit number)

City   State   Zip Code

County   Have you lived at this address for 30 or more days? ☐ Yes ☐ No   If "No", date moved?

5 Provide a mailing address.

Do you receive mail at your residential   ☐ Yes ☐ No
If "No", you are required to provide a mailing address.

Mailing Address Line 1

Mailing Address Line 2

Mailing Address Line 3

City   State   Zip Code

No Physical Address? If you do not have an address, use the space to the right to illustrate where you normally live or sleep. Write in the names of the nearest crossroads (or streets). Draw an **X** on the map to show where you live or usually sleep.

IMPORTANT: You should also provide a valid mailing address above to permit the board of elections to send you a voter card.

NORTH ⬆

6 Provide your demographic information (optional).

Gender   ☐ Male   ☐ Female

Ethnicity   ☐ Not Hispanic/Latino   ☐ Hispanic/Latino

Race   ☐ African American/Black   ☐ American Indian/Alaska Native   ☐ Asian   ☐ Multiracial   ☐ Native Hawaiian/Pacific Islander   ☐ White   ☐ Other

7 Provide your choice for political party affiliation.

☐ Constitution Party   ☐ Libertarian Party   ☐ Other _____
☐ Democratic Party   ☐ Republican Party
☐ Green Party   ☐ Unaffiliated

If you select a party that is not recognized in North Carolina, you will be registered as Unaffiliated.

8 Complete if you are currently registered to vote in another NC county or in another state.
(This information will be used to cancel your previous voter registration in the other county or state.)

First Name Used in Last Registration   Middle Name Used in Last Registration   Last Name Used in Last Registration   Suffix

Address Where You Were Last Registered   City/State/Zip Code of Last Registration   County of Last Registration

9 Provide your contact information (optional).
(This information is helpful if we need to contact you concerning your voter registration. Your contact information may be disclosed as a public record.)

Area Code   Phone Number   Email Address   Would you like to be contacted to be a poll worker? ☐ Yes ☐ No

10 Sign below to attest to your qualifications to vote.
FRAUDULENTLY OR FALSELY COMPLETING THIS FORM IS A CLASS I FELONY UNDER CHAPTER 163 OF THE NC GENERAL STATUTES.

I attest, under penalty of perjury, that in addition to having read and understood the contents of this form, that: (1) I am a United States citizen, as indicated above; (2) I am at least 18 years of age, or will be by the date of the general election; or I am at least 16 years old and understand that I must be at least 18 years old on the day of the general election to vote; I shall have been a resident of North Carolina, this county, and precinct for 30 days before the date of the election in which I intend to vote; (3) I will not vote in any other county or state after submission of this form and if I am registered elsewhere, I am canceling that registration at this time; and (4) I am not currently serving a felony sentence, including any probation, post-release supervision, or parole OR I am serving an extended term of probation, post-release supervision, or parole, I have outstanding fines, fees, or restitution, and I do not know of another reason that my probation, post-release supervision, or parole was extended.

X _____   _____
Signature Required   Date

Figure 9. North Carolina Voter Registration Form from 2021. This form asks the voter to provide their demographic information in terms of ethnicity and race in the box highlighted in red.

Since Test 2 involved seeing whether the degree of bias in Lookalike and Special Ad audiences changed with the use of source lists of individuals with commonly given names of one race or

ethnicity for the source audience, I started with the ethnicolr library in Python to predict a voter's race or ethnicity based on their first and last names [78]. I then created subsets of voters of each demographic group who have commonly given names of that race or ethnicity according to ethnicolr predictions.

Research has revealed trends in the popular names that are given to babies of different races and ethnicities [37, 38, 51]. In fact, these trends are often used in political science and social science research to predict the race and ethnicity of individuals in a dataset that doesn't contain explicit race and ethnicity fields [38]. A common approach is to use the U.S. Census Bureau's Frequently Occurring Surnames dataset, which contains all last names occurring more than 100 times in the Decennial Census and provides a racial and ethnic breakdown in the usage of each last name [79]. However, using only last names may incorrectly predict the race of some individuals, especially African-Americans, who have last names that are also frequently used by White families. In addition, African-Americans often have more racially distinct first names. Thus, Sood and Laohaprapanon created the ethnicolr library, which uses a Long Short Term Memory neural network trained on Florida voter registration data to predict an individual's race or ethnicity using both their first and last name [78]. They found that their model using both names, which had a precision of 83% and a recall of 84%, performed better than a model using only the last name in sample testing. Since its release in 2018, the ethnicolr library has been used in multiple research studies in medicine, political science, economics, education, and other fields [80, 81, 82, 83, 84]. Figure 10 shows the results from the ethnicolr library for four names. The model outputs probabilities that a given first and last name belongs to one of four classes: White, African-American, Asian, or Hispanic. It then predicts a label based on which class had the highest probability. The first two rows compare "Jinyan Zang" to "Jinyan Zane", where changing the last name resulted in a shift in the predicted label from Asian to White. The last two rows compare "Latanya Sweeney" to "Tanya Sweeney." The change in the first name resulted in the predicted label changing from African-American to White.

First name	Last name	Predicted Label	Probabilities			
			White	African-American	Asian	Hispanic
Jinyan	Zang	Asian	19%	2%	77%	2%
Jinyan	Zane	White	49%	31%	14%	5%
Latanya	Sweeney	African-American	10%	88%	1%	1%
Tanya	Sweeney	White	89%	9%	1%	2%

Figure 10. Example Results from the ethnicolr Library in Python Using Four Names. Changing "Jinyan Zang" to "Jinyan Zane" resulted in the predicted label switching from Asian toward White. Changing "Latanya Sweeney" to "Tanya Sweeney" resulted in the predicted label switching from African-American to White.

Since Test 2 also involved seeing whether the degree of bias in Lookalike and Special Ad audiences changed when source lists used voters living in racial or ethnic enclaves, I started with the U.S. Census Bureau's American Community Survey to find the racial and ethnic breakdown of every ZIP code in North Carolina [85]. According to the Census, North Carolina's population is 71% White, 22% African-American, 3% Asian, and i10% Hispanic [86]. I classified African-Americans living in ZIP codes with >50% African-Americans, Whites living in ZIP codes with >90% Whites, Asians living in ZIP codes with >20% Asians, and Hispanics living in ZIP codes with >20% Hispanics as having a racially or ethnically stereotypical ZIP code.

Combining ethnicolr predictions based on names with Census data identifying racial enclave ZIP codes, I created four versions of 10K voter samples for each race/ethnicity to study the demographic biases of Lookalike and Special Ad audiences in Test 2 (Figure 11).

Race/Ethnicity	10K Voter Samples Used to Create Lookalike and Special Ad Audiences
African-American	African-American
	African-American with Commonly Given African-American Names
	African-American In ZIP Codes with >50% African-Americans
	African-American with Commonly Given African-American Names & In ZIP Codes with >50% African-Americans
White	White
	White with Commonly Given White Names
	White In ZIP Codes with >90% Whites
	White with Commonly Given White Names & In ZIP Codes with >90% White
Asian	Asian
	Asian with Commonly Given Asian Names
	Asian In ZIP Codes with >20% Asians
	Asian with Commonly Given Asian Names & In ZIP Codes with >20% Asians
Hispanic	Hispanic
	Hispanic with Commonly Given Hispanic Names
	Hispanic In ZIP Codes with >20% Hispanics
	Hispanic with Commonly Given Hispanic Names & In ZIP Codes with >20% Hispanics

Figure 11. 10K Voter Samples with Different Traits Used to Create Lookalike and Special Ad Audiences for Each Race/Ethnicity for Test 2 Analysis.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Test 1 – Studying the Racial and Ethnic Breakdowns of Targeting Options By Facebook
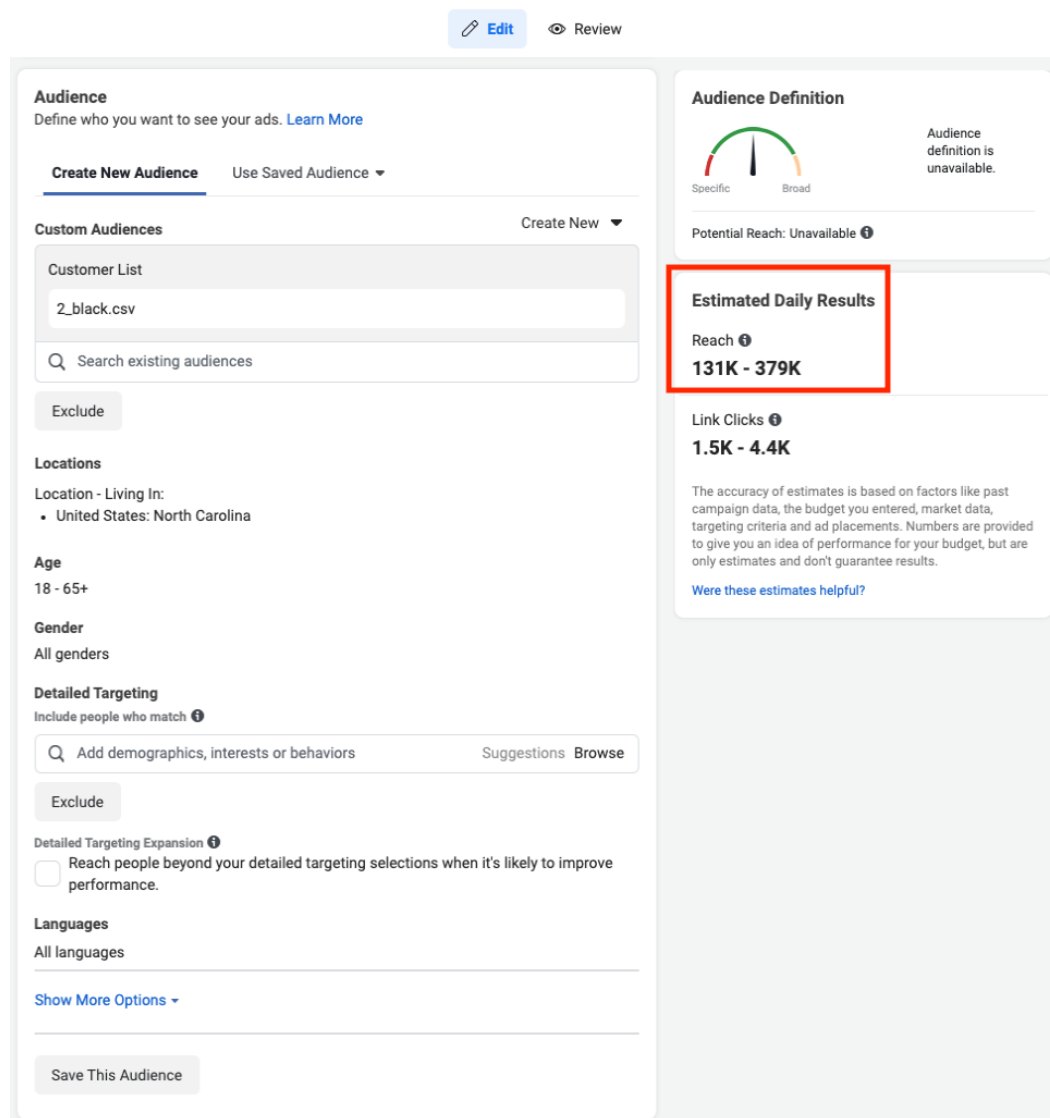
I started by recording the target size for the relevant Facebook targeting options in 2020 and 2021 using its ad planning tool, shown in Figure 12. For 2020, the options were "African American (US)", "Asian American (US)", and "Hispanic (US – All)". In 2021, the options were "African-American Culture", "Asian American Culture", and "Hispanic American Culture". I then compared the target size of each Facebook option to the corresponding estimate from the US Census for the 18+ population of that demographic group [87].
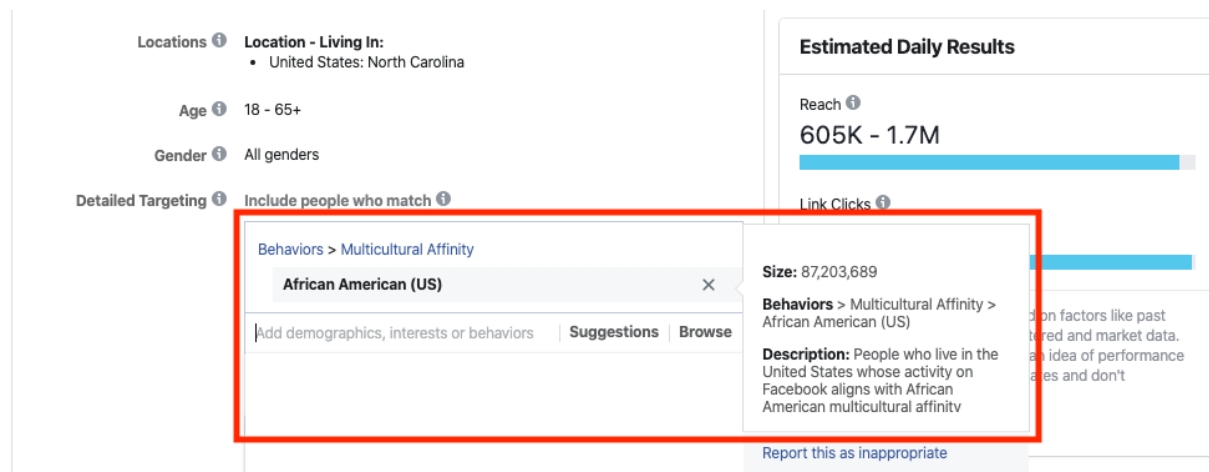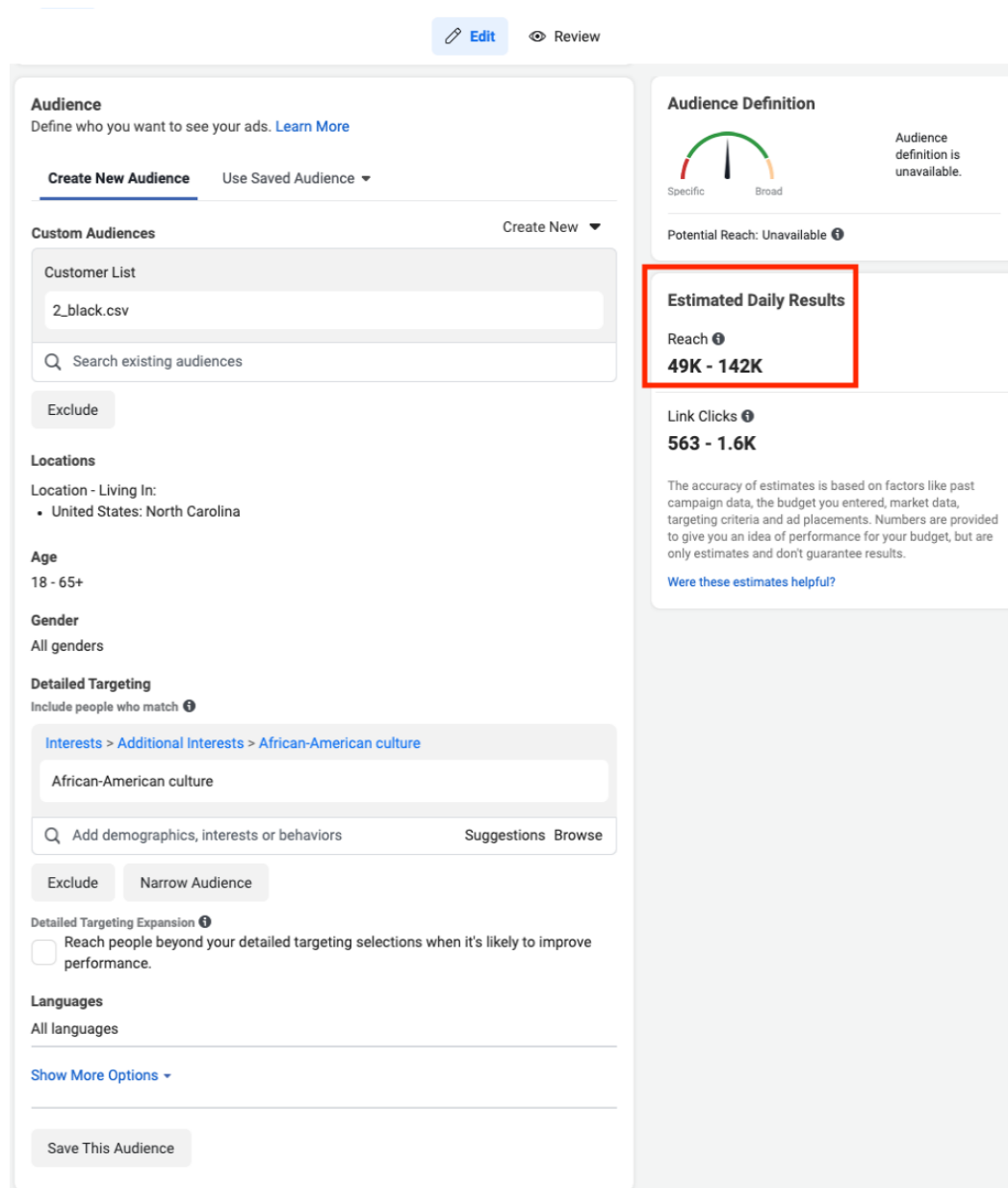

Figure 12. Target Size of the "African American (US)" Option in 2020. The size in January 2020 was 87,203,689 Facebook users.

For each wave, I created different Custom audiences on Facebook by segmenting the North Carolina voter list by different racial groups – African-American, Asian, and White – and by different ethnic groups – Hispanic and Non-Hispanic. I then randomly sampled 10,000 voters from each of the segments to use as the source audience to create a Lookalike audience of the 1% most similar users in the United States on Facebook.

I iterated and set each Custom or Lookalike audience as the target on Facebook's ad planning tool (Figure 13). I then added the targeting option being tested under the "Detailed Targeting" setting of Facebook's ad planning tool and recorded the updated daily reach estimate. In the example shown in Figure 14, an ad that targeted the African-American voters Custom audience and also matched the "African-American Culture" interest option would only reach 142,000 North Carolina users daily. This is 37% of the reach of an ad that targeted the same Custom audience without the additional "African-American Culture" criterion (Figure 13).

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Figure 13. Example of the Estimated Daily Reach of the African-American Voters Custom Audience on Facebook's Ad Planning Tool in 2021. The maximum estimated daily reach was 379K, as highlighted in the red box.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Figure 14. Example of the Estimated Daily Reach of Targeting an African-American Voters Custom Audience Who Match the "African-American Culture" Interest Option on Facebook's Ad Planning Tool in 2021. The maximum estimated daily reach was 142K, or 37% of the reach in Figure 13.

When the estimated reach is below 1 million, Facebook's ad planning tool in general rounds to the nearest thousand by using "K", which may have introduced some small rounding errors into the results shown in this study. I maximized the ad budget to $1 million in most cases to ensure the maximum reach estimate is used.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Test 2 – Studying the Bias in Facebook's Lookalike Audience and Special Ad Audience Tools Using A Set Theory Approach

While Facebook doesn't permit direct demographic queries for a Lookalike or Special Ad audience, I was able to leverage the daily reach estimates of Facebook's ad planning tool to indirectly observe the racial and ethnic breakdown of a given Lookalike or Special Ad audience amongst North Carolina voters by using the set theory approach described in Sapiezynski et al. [5].

For example, to study the African-American versus White bias of a Lookalike or Special Ad audience, I first created a 2 million voter sample made up of 1 million randomly sampled African-American voters and 1 million randomly sampled White voters. From the remaining voters in North Carolina not in the sample, I then created lists of 10,000 randomly sampled voters with particular racial- or ethnic-related traits to create corresponding Lookalike and Special Ad audiences of the 1% most similar Facebook users in the United States (Figure 11). Thus, to see whether a Lookalike audience based on African-Americans was biased toward including more African-American than White voters, I first measured the estimated reach of the 1 million African-American voters Custom audience (Figure 15). I then measured the estimated reach of a 1 million African-American audience that excluded the Lookalike Audience based on African-Americans (Figure 16). Finally, I repeated the process for the 1 million White voters Custom audience.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Figure 15. Example of the Estimated Daily Reach of Targeting a 1 Million African-American Voters Custom Audience in 2021. The maximum estimated reach was 299K, as highlighted in the red box.
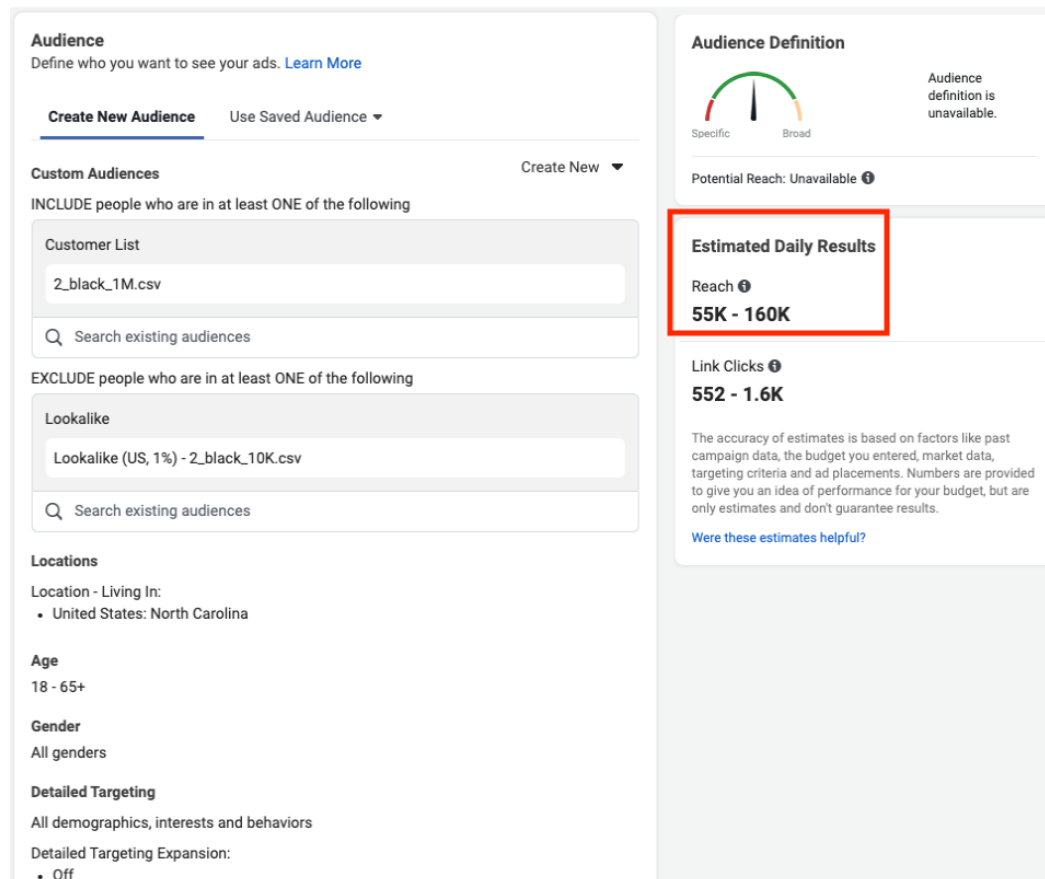
Figure 16. Example of the Estimated Daily Reach of Targeting a 1 Million African-American Voters Custom Audience While Excluding a Lookalike Audience Based on African-American Voters in 2021. The maximum estimated reach was 160K, which is 139K fewer users than Figure 15.

In this case, the estimated reach of the 1 million African-American voters Custom audience decreased by 139,000 after excluding the Lookalike audience based on African-Americans, but the estimated reach of the 1 million White voters Custom audience decreased by only 17,000 under the same circumstances. Thus, within the intersection of the Lookalike audience and the 2 million voter sample, 89% of the overlap were African-American voters, while 11% were White voters, as shown in Figure 17. This means that African-Americans were overrepresented, being far above the 50% baseline in the 2 million voter sample. In this paper, I refer to statistics like the 89% above as the sample share of African-American voters in the Lookalike audience, and I refer to statistics like the 11% above as the sample share of White voters in the

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Lookalike audience. I also describe a demographic group with a size exceeding the baseline level as "overrepresented" and a group with a size below the baseline as "underrepresented". If the set theory approach found that a Lookalike or Special Ad audience overrepresented a demographic group, I describe that audience in this paper as being "biased" toward that race or ethnicity.
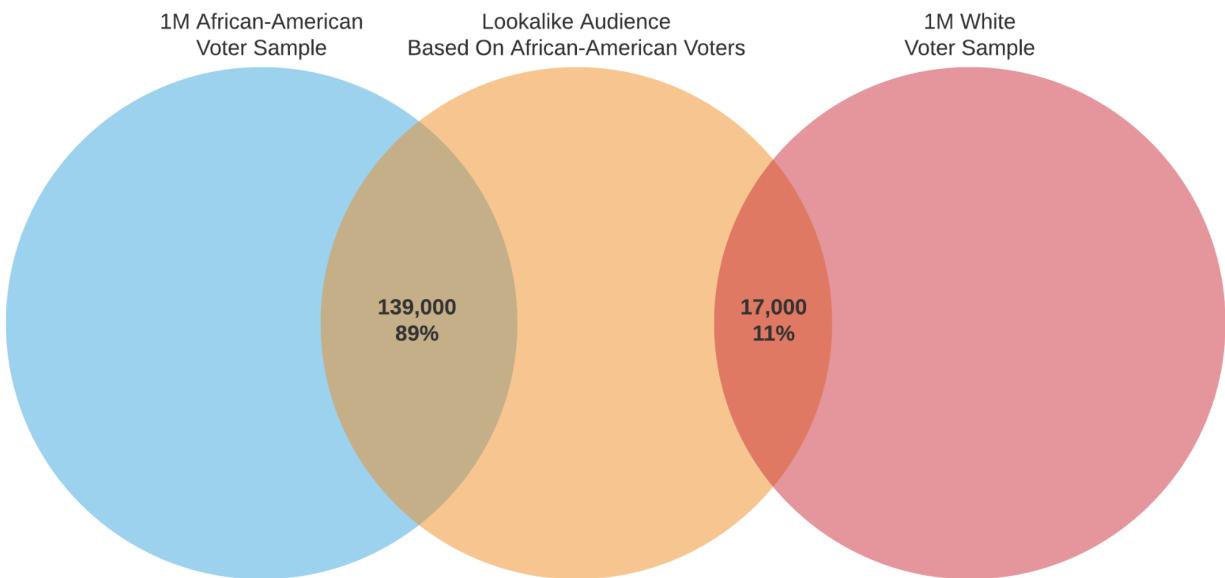
Figure 17. Intersection of the Lookalike Audience Based on African-American Voters with the 1 Million African-American Voter Sample and the 1 Million White Voter Sample in 2021. Because 89%, or 139,000 users in the audience, were African-American voters and just 11% were White voters, the Lookalike audience overrepresented African-American voters and appears to be biased toward African-Americans.

There are far fewer Asian voters in North Carolina thanAfrican-American and White voters. As Figure 18 illustrates, I used a 150,000-voter sample as the comparison set, built from samples of 50,000 Asian, African-American, and White voters each, to study the degree of bias in a Lookalike or Special Ad audience toward Asians using the set theory approach. Similarly, Figure 19 illustrates the use of a 200,000-voter sample as the comparison set, half Hispanic and half Non-Hispanic, to study the degree of bias in a Lookalike or Special Ad audience toward Hispanics. Finally, since Facebook does not allow Special Ad audiences to be in the "Exclusion" position, I simply flipped the settings described earlier by first measuring the reach of the Special Ad audience on its own and then measuring the reach of the Special Ad audience while excluding a given voter sample, in order to implement the set theory approach for measuring biases in Special Ad audiences.
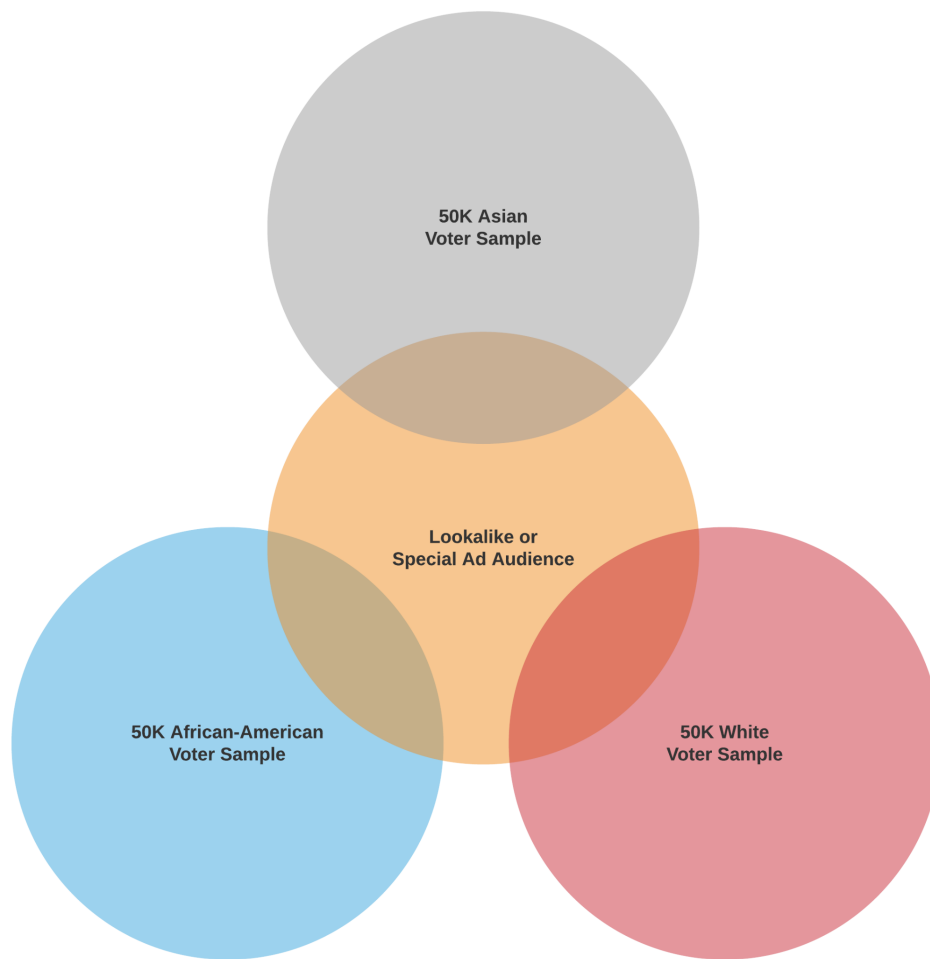
Figure 18. Set Theory Approach to Study the Bias Toward Asians in Lookalike or Special Ad Audiences Using a 150,000 Sample of Asian, African-American, and White Voters.
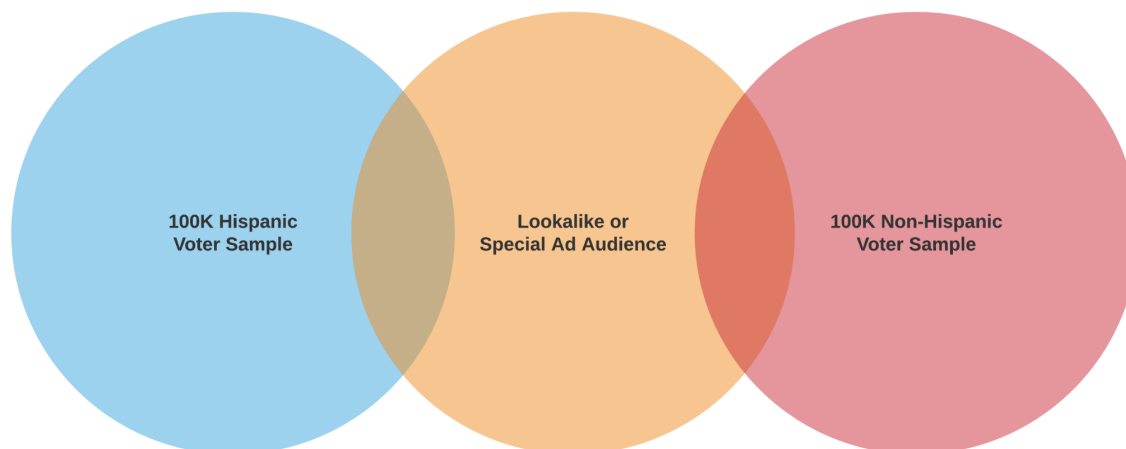
Figure 19. Set Theory Approach to Study the Bias toward Hispanics in Lookalike or Special Ad Audiences Using a 200,000 Sample of Hispanic and Non-Hispanic Voters.

Results

Test 1 – Studying the Racial and Ethnic Breakdowns of Targeting Options By Facebook

The number of Facebook users in a target group can bear little relation to US Census estimates of the population of the related ethnic or racial group. The Asian and Hispanic "culture" target groups examined in our study in 2021 were much closer to the Census estimates for Asians and Hispanics than the discontinued "affinity" groups. In 2021, Facebook's "African-American Culture" ad targeting option contained 75% fewer White users than the old "African American (US)" option removed in the previous year. By contrast,the number of non-Asians and Non-Hispanics increased significantly for the Asian- and Hispanic-related targeting options.

In 2020, Facebook's multicultural affinity group targeting option for "African-American (US)" included 2.49 times more people than the Census estimate of the African American adult population in the US, while "Asian American (US)" was 0.28 times the Census estimate for that ethnic group, and "Hispanic (US – All)" was 0.50 times the Census estimate for Hispanics (Figure 20). In 2021, the number of Facebook users interested in "African-American Culture" was 2.26 times the Census estimate of African-Americans, while the target size of "Asian American Culture" was 0.96 times the Census estimate of Asian-Americans and that of "Hispanic American Culture" was 1.31 times the Census estimate of Hispanic residents (Figure 20). Thus, Facebook's Asian American Culture and Hispanic American Culture option provided a target size closer to related Census population estimates in 2021, while the target size of the African-American Culture option stayed at twice the population estimate, declining only slightly from the size of the "African American (US)" option in 2020.

	Facebook Advertising Targeting Option	Facebook Target Size	Census (Age 18+ Population)	Facebook-Census Ratio
2020	African American (US)	87,203,689	35,079,870	2.49x
	Asian American (US)	4,972,438	17,502,608	0.28x
	Hispanic (US – All)	21,542,628	43,089,980	0.50x
2021	African-American Culture	79,388,010	35,079,870	2.26x
	Asian American Culture	16,807,470	17,502,608	0.96x
	Hispanic American Culture	56,515,880	43,089,980	1.31x

Figure 20. Ratio of Facebook Advertising Target Size to Census Population Estimate. The Census population estimates are the US Census' 2019 population estimates [87] for each race or ethnicity for individuals 18 and older, since that was the age limit for the

Facebook Detailed Targeting. Facebook's ad targeting sizes became much closer to the Census estimates by 2021 for Asians and Hispanics.

In 2020, the share of NC voters on Facebook who could be reached by targeting "African American (US)" was 43% of African-American voters, 23% of Asian voters, and 39% of White voters (Figure 21). However, since there are far more White voters than African-American voters in North Carolina, this means that the "African-American (US)" targeting option could reach approximately 150,000 African-American voters and 428,000 White voters (Figure 21). In 2021, the conditional probability of being interested in "African-American Culture" was significantly higher for African-American voters at 37%, compared to 8% for Asian and White voters (Figure 21). This resulted in 142,000 African-American voters, 109,000 White voters, and only 2,000 Asian voters being interested in "African-American Culture" (Figure 21). Thus, 75% fewer Whites were included in "African-American Culture" in 2021 than in "African American (US)" in 2020.
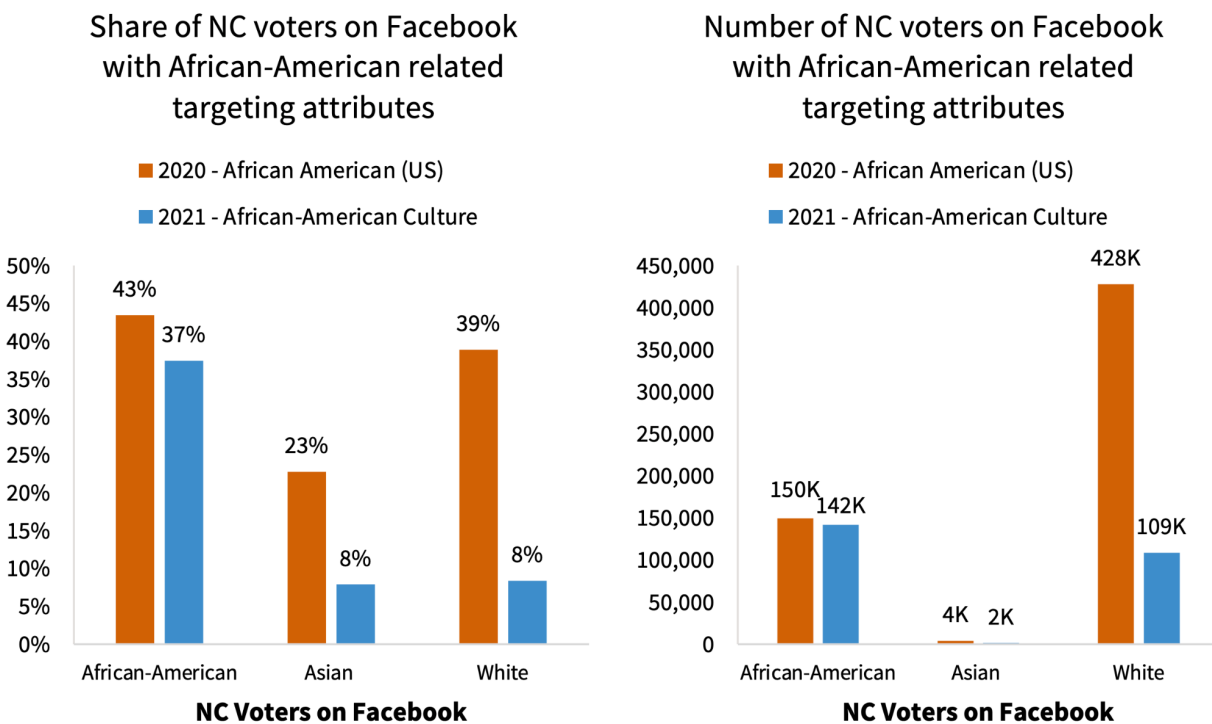
Figure 21. Share and Number of NC Voters on Facebook with African-American Related Targeting Attributes. 75% fewer Whites were included in the African-American Culture interest group in 2021 than in "African American (US)" in 2020.
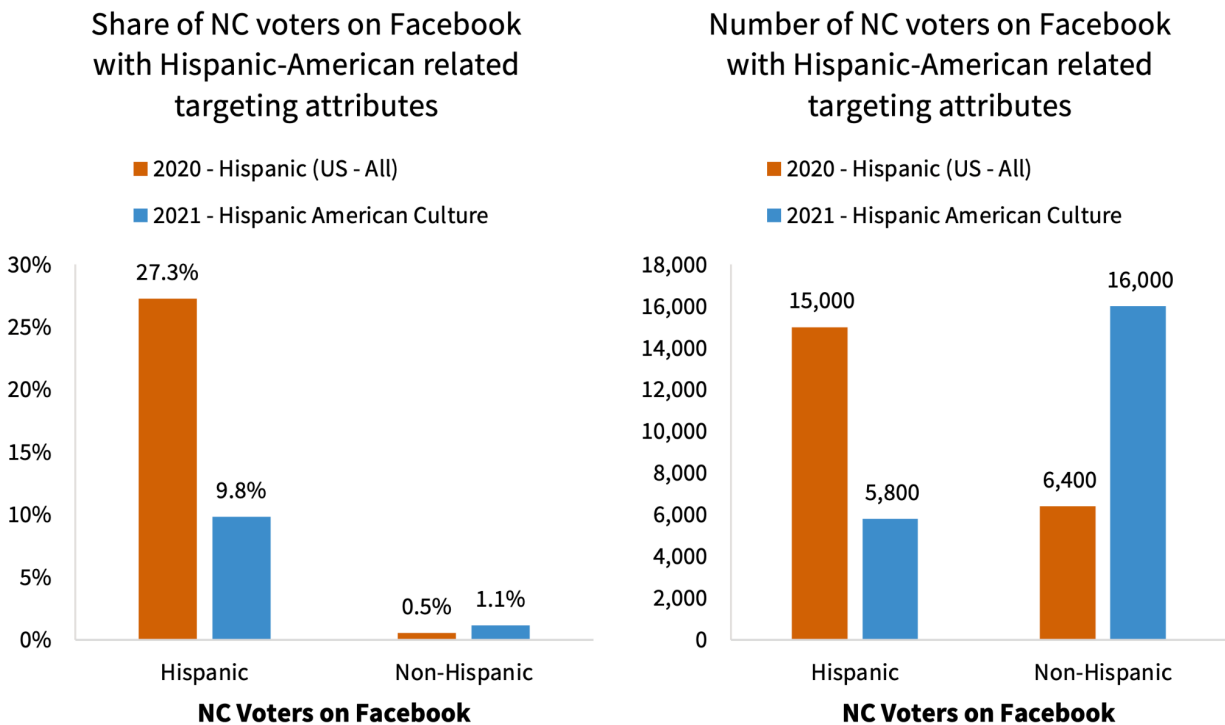
In 2020, the share of Asian voters on Facebook reached by the "Asian American (US)" option was 8.9%, which dwarfed the 0% of African-American voters and the 0.1% of White voters (Figure 22). However, in terms of absolute numbers, this means that approximately 1,600

Asian and White voters could be reached with the same multicultural affinity option (Figure 22). In 2021, the "Asian American Culture" target option could reach more non-Asians with a 7.4% share of Asian voters, but this option also reached a 2.5% share of African-American voters and 1.4% share of White voters (Figure 22). This means that targeting "Asian American Culture" would reach only 1,400 Asian voters, far fewer than the 9,400 African-American voters and 18,000 White voters who share this interest (Figure 22).
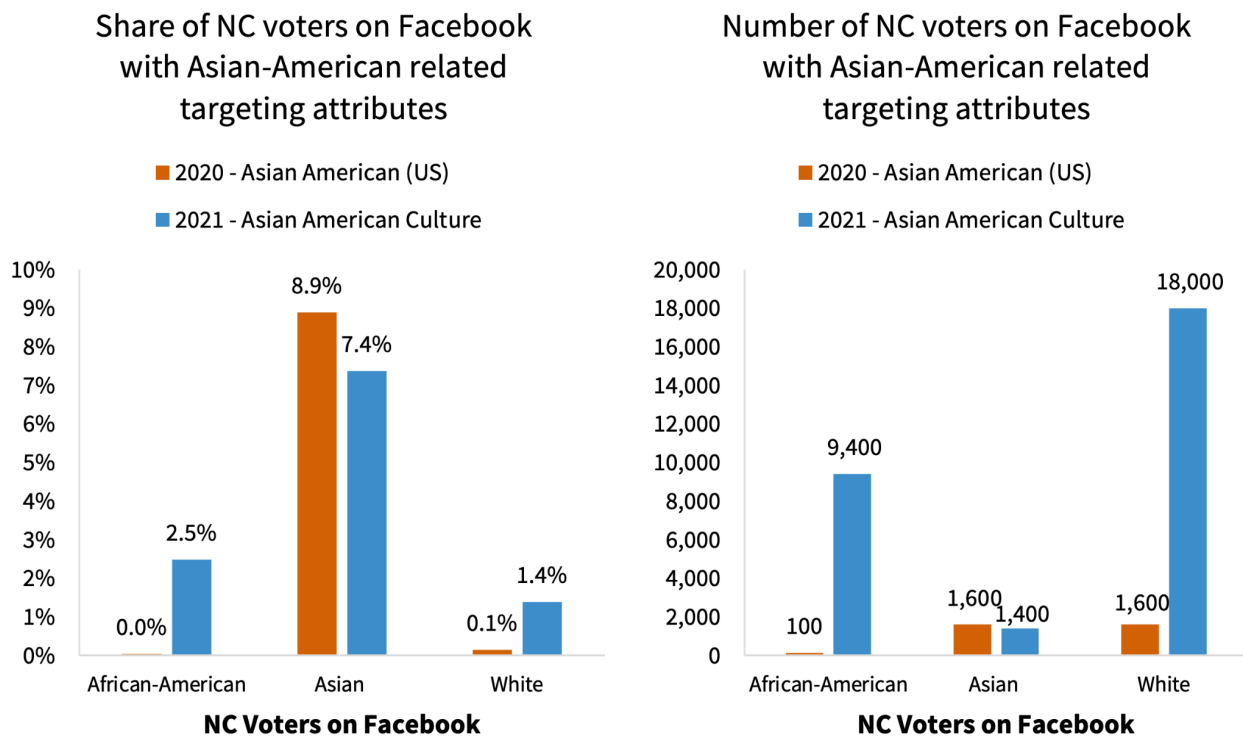
Figure 22. Share and Number of NC Voters on Facebook with Asian-American Related Targeting Attributes. In 2021, the "Asian American Culture" target option reached far more non-Asians than Asians, with an estimated reach of 1,400 Asian voters but 9,400 African-American voters and 18,000 White voters.

In 2020, Hispanic voters were more likely than Non-Hispanic voters to be reached by Facebook's "Hispanic (US -All)" targeting option in both relative and absolute terms. Of the Hispanic voters on Facebook in our sample, 27.3% had the "Hispanic (US – All)" attribute, compared to only 0.5% of Non-Hispanic voters, which translates into approximately 15,000 Hispanic voters and 6,400 Non-Hispanic voters (Figure 23). In 2021, only 9.8% of Hispanic voters were classified as interested in "Hispanic American Culture", and 1.1% of Non-Hispanic voters had the same interest (Figure 23). This results in only 5,800 Hispanic voters but 16,000 Non-Hispanic voters being reached by the targeting option, since there are far more Non-Hispanic than Hispanic voters on Facebook (Figure 23).

Share of NC voters on Facebook with Hispanic-American related targeting attributes

- 2020 - Hispanic (US - All)
- 2021 - Hispanic American Culture

Number of NC voters on Facebook with Hispanic-American related targeting attributes

- 2020 - Hispanic (US - All)
- 2021 - Hispanic American Culture

Figure 23. Share and Number of NC Voters on Facebook with Hispanic-American Related Targeting Attributes. In 2020, more than twice as many Hispanic voters as Non-Hispanic voters are targeted by "Hispanic (US – All)", but that ratio flips in 2021, with more than twice as many Non-Hispanics as Hispanics reached by the "Hispanic American Culture" targeting option.

I also found that in 2021, Lookalike audiences based on African-American, Asian, White, or Hispanic voters tended to have shares interested in "African-American Culture" or "Hispanic American Culture" that were similar to those ethnic group's shares of the corresponding voter lists. However, all Lookalike audiences, regardless of which voter list was used to create them, had about 2-3% of their users interested in "Asian American Culture", including the Lookalike audience based on Asian voters. For more details about these results, see Appendix A.

Test 2 - Studying the Bias in Facebook's Lookalike Audience and Special Ad Audience Tools Using A Set Theory Approach

In both 2020 and 2021, Facebook's Lookalike and Special Ad audiences overrepresented the sample share of either African-American or White voters, depending on which race was dominant in the customer list used as the source audience. This bias increased when using customer lists with stereotypically African-American or White names or ZIP codes. Similar biases were observed for Lookalike audiences based on Asians or Hispanics. In one case, a customer list of Asians with stereotypical Asian names and ZIP codes generated a Lookalike

audience having a 100% sample share of Asian voters. In a shift from 2020, I found that Special Ad audiences based on Asians did not significantly overrepresent Asian voters in 2021.

When studying the breakdown of the 2 million NC voter sample intersecting with Lookalike audiences, 83% of the overlap between the sample and a Lookalike audience based on African-Americans in 2020 were African-American voters. That overlap increased to 89% in 2021 (Figure 24). As the customer list used to create the Lookalike audience took on more stereotypically African-American traits by being limited to voters with commonly given African-American names, voters who live in a ZIP code with >50% African-Americans, or both, the sample shares of African-American voters in the Lookalike audiences increased to 93% in 2020 and to 94% in 2021 (Figure 24).
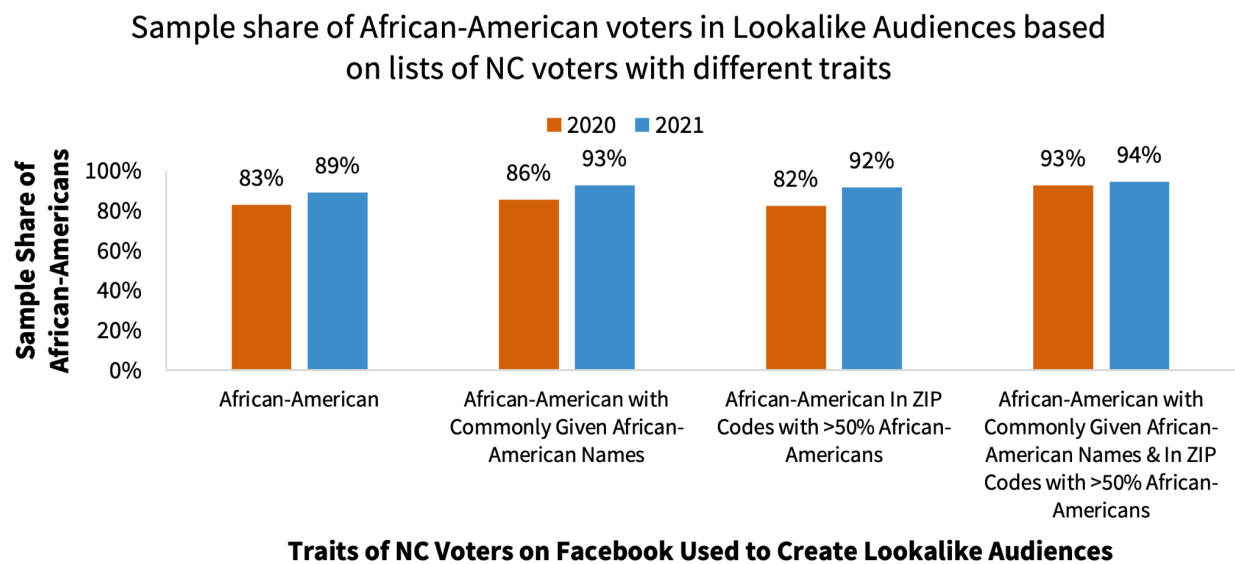
Sample share of African-American voters in Lookalike Audiences based on lists of NC voters with different traits

Figure 24. Sample Share of African-American Voters in Lookalike Audiences Based on Lists of NC Voters with Different Traits. As the customer list used to create the Lookalike audience contained more stereotypically African-American traits in terms of names and ZIP codes, the sample shares of African-American voters in the corresponding Lookalike audiences increased, reaching 93% in 2020 and 94% in 2021.

Similarly, I also found that Lookalike audiences based on Whites contained a large majority of White voters in their overlap with the 2 million NC voter sample in both waves. In 2020, 73% of the voters shared between the 2 million voter sample and the Lookalike audience based on White voters were White. In 2021 the rate was similar at 71% (Figure 25). As the White voters used to create the Lookalike audience took on additional White-affiliated traits by having commonly given White names, living in a >90% White ZIP code, or both, the sample shares of Whites peaked at 87% in 2020 and 84% in 2021 for the resulting Lookalike audiences (Figure 25).
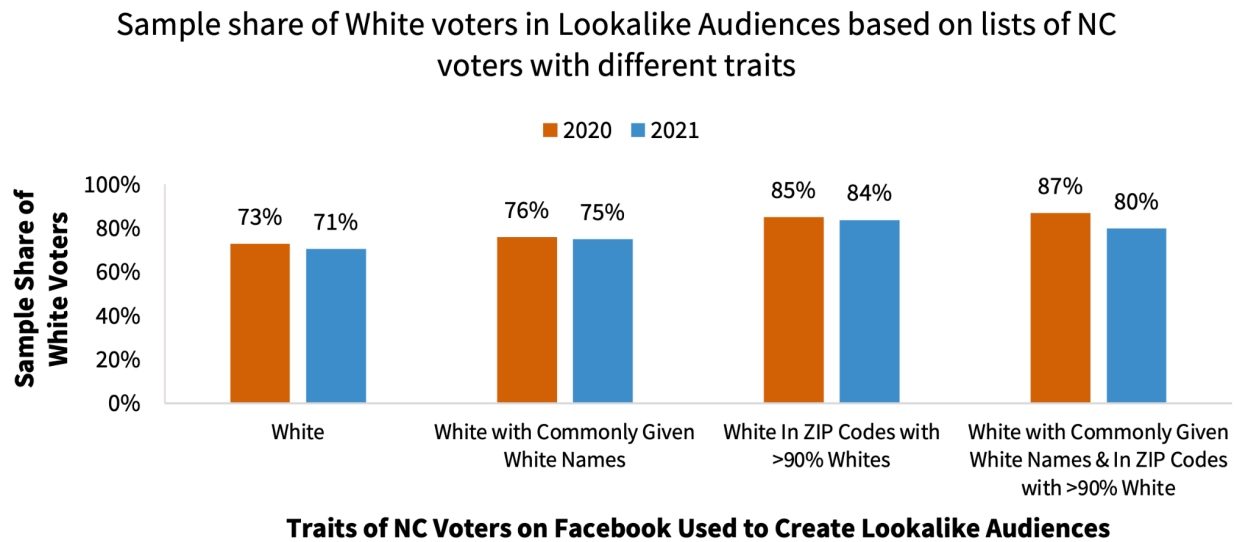
Sample share of White voters in Lookalike Audiences based on lists of NC voters with different traits

Figure 25. Sample Share of White Voters in Lookalike Audiences Based on Lists of NC Voters with Different Traits. Lookalike audiences based on Whites overrepresented White voters in both waves, with a sample share of 73% White voters in 2020 and 71% in 2021.

In 2020, Lookalike audiences based on Asian voters appeared to have a moderate tendency to favor Asians. This became a significantly stronger trend in 2021. In 2020, in a comparison of the 150,000-voter sample with Lookalike audiences based on Asian voters, 51% of the voters who overlapped were Asian. This overlap increased to 83% for the Lookalike audience based on Asian voters who had commonly given Asian names and lived in Asian enclave ZIP codes (Figure 26). In 2021, the overlap rates started at 68% for the Lookalike audience based on Asians and increased to 100% for the Lookalike audience based on Asians with stereotypically Asian traits of name and ZIP code (Figure 26). In all cases, the sample shares of Asian voters were far above the 33% baseline in the 150,000-voter sample. It is possible that the actual sample share of Asian voters in the most extreme case was slightly below 100%. A few Black and White voters could have been part of the intersection between the Lookalike audience and the voter sample, but those voters might not have been captured because Facebook rounds the advertising reach estimate on its ad planning tool to the nearest 1,000.
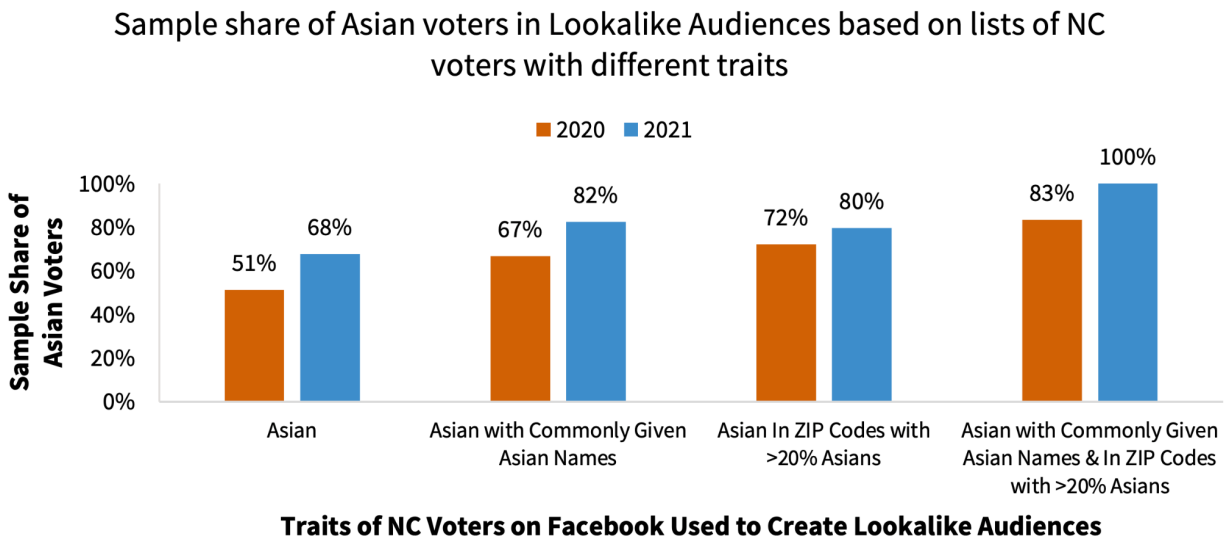
Sample share of Asian voters in Lookalike Audiences based on lists of NC voters with different traits

■ 2020 ■ 2021

Sample Share of Asian Voters

- Asian: 2020 51%, 2021 68%
- Asian with Commonly Given Asian Names: 2020 67%, 2021 82%
- Asian In ZIP Codes with >20% Asians: 2020 72%, 2021 80%
- Asian with Commonly Given Asian Names & In ZIP Codes with >20% Asians: 2020 83%, 2021 100%

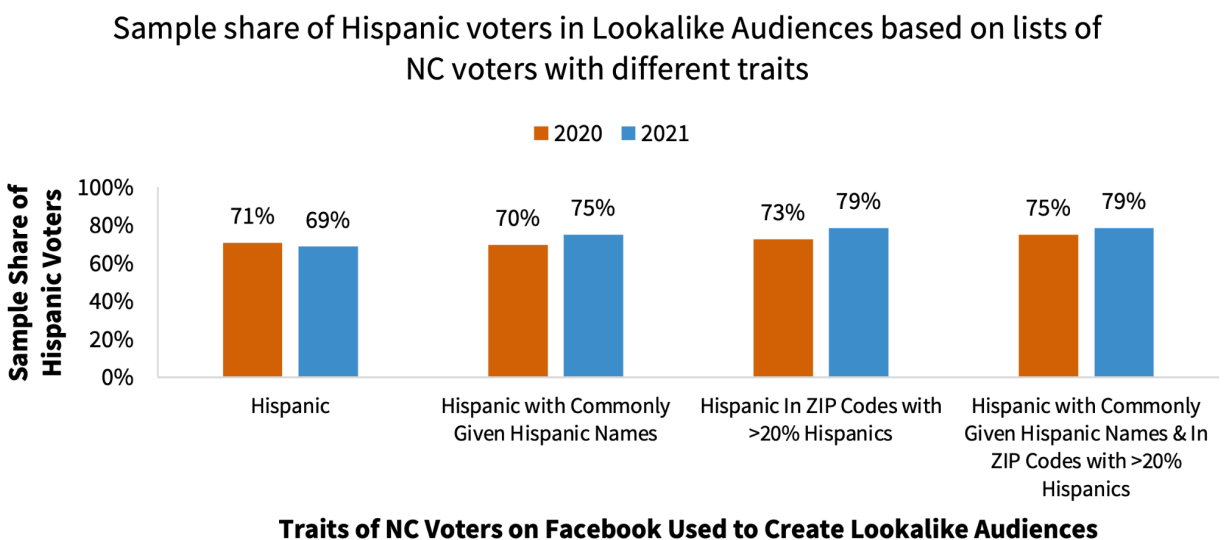**Traits of NC Voters on Facebook Used to Create Lookalike Audiences**

Figure 26. Sample Share of Asian Voters in Lookalike Audiences Based on Lists of NC Voters with Different Traits. In 2021, Lookalike audiences based on Asians were significantly biased toward including Asian voters, reaching a 100% sample share of Asian voters in one case.

In both 2020 and 2021, Lookalike audiences based on Hispanics were biased toward including more Hispanic voters. In 2020, Hispanic voters accounted for 71% of the overlap between the 200,000-voter sample and the Lookalike audience based on Hispanics. A similar 69% share was seen in 2021 (Figure 27). As the customer list used to create the Lookalike audience appeared more stereotypically Hispanic by name or ZIP code, the sample share of Hispanic voters increased slightly to 75% in 2020, and 79% in 2021 (Figure 27).

Sample share of Hispanic voters in Lookalike Audiences based on lists of NC voters with different traits

■ 2020 ■ 2021

Sample Share of Hispanic Voters

- Hispanic: 2020 71%, 2021 69%
- Hispanic with Commonly Given Hispanic Names: 2020 70%, 2021 75%
- Hispanic In ZIP Codes with >20% Hispanics: 2020 73%, 2021 79%
- Hispanic with Commonly Given Hispanic Names & In ZIP Codes with >20% Hispanics: 2020 75%, 2021 79%

Traits of NC Voters on Facebook Used to Create Lookalike Audiences

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/
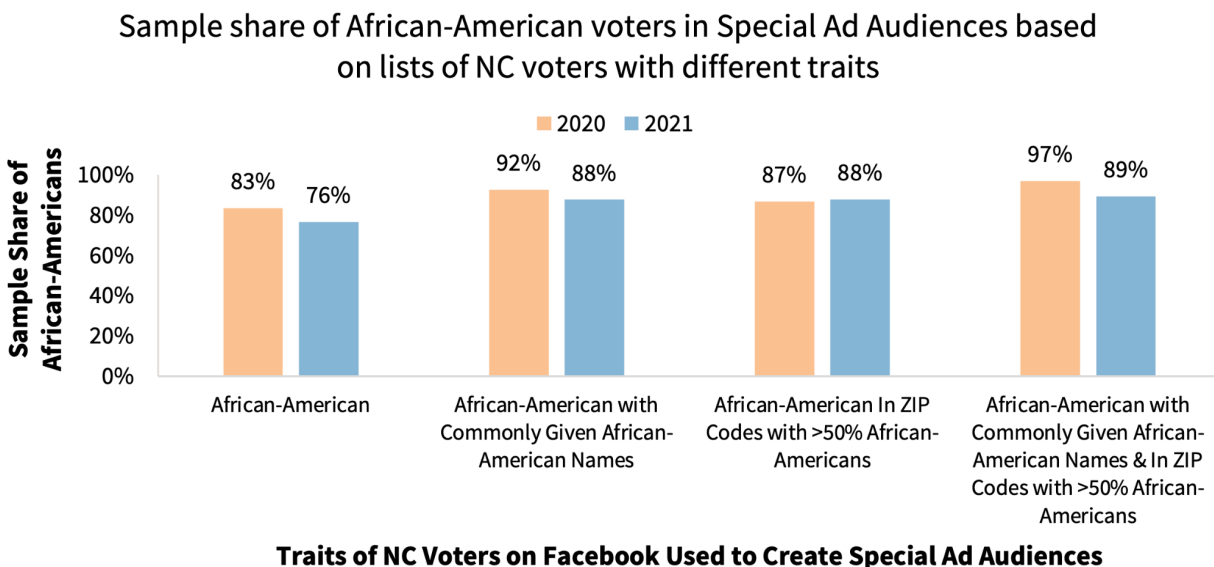
Figure 27. Sample Share of Hispanic Voters in Lookalike Audiences Based on Lists of NC Voters with Different Traits. In both 2020 and 2021, Lookalike audiences based on Hispanics included more Hispanic voters than Non-Hispanic voters, with a sample share of 71% Hispanic voters in 2020 and 69% in 2021.

Even though Facebook created Special Ad audiences as an anti-discrimination tool for housing, employment, and credit-related ads, I found that Special Ad audiences could still be biased toward including more African-American or White voters, depending on the demographics of the source audience.

In 2020 and 2021, the Special Ad audiences based on African-Americans demonstrated significant biases toward including more African-American voters. In 2020, the sample shares of African-American voters in the relevant Special Ad audiences started at 83% and went up to 97% for the Special Ad audience based on African-Americans with stereotypically African-American names and ZIP codes (Figure 28). These sample shares are similar to those observed for the Lookalike audiences based on African-Americans shown in Figure 24. In 2021, slightly fewer African-Americans were in the intersection of the 2 million voter sample with the relevant Special Ad audiences, starting at a 76% sample share and increasing to 89% for the Special Ad audience based on a list of African-American voters with commonly given African-American names living in ZIP codes with >50% African-Americans (Figure 28).

Sample share of African-American voters in Special Ad Audiences based on lists of NC voters with different traits
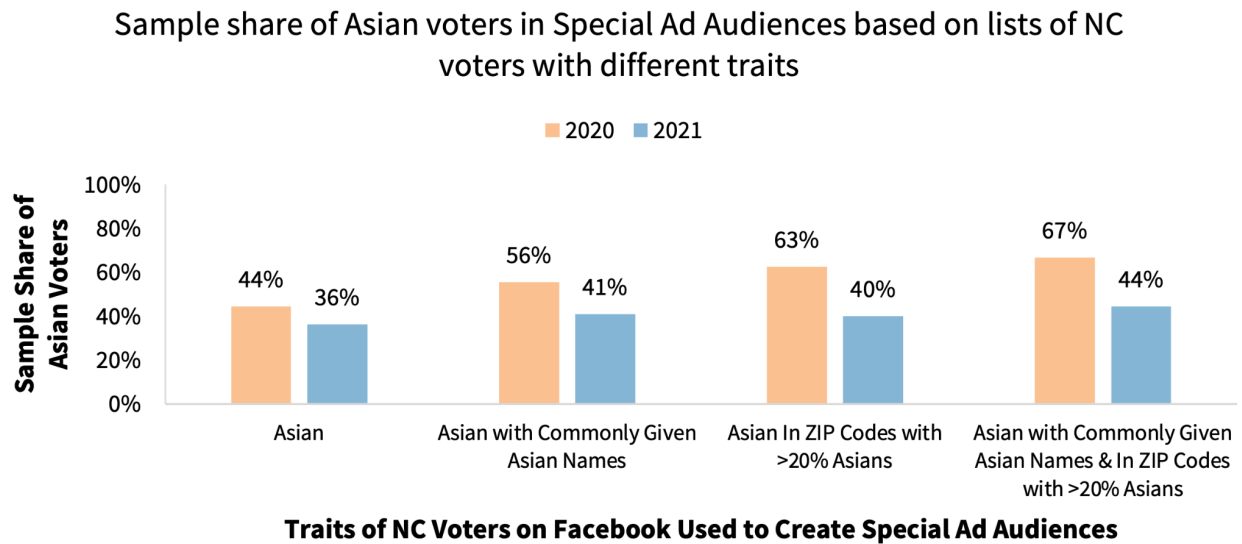

Figure 28. Sample Share of African-American Voters in Special Ad Audiences Based on Lists of NC Voters with Different Traits. In both waves, the Special Ad audiences based on African-Americans demonstrated significant biases toward including more African-American voters, with a sample share of 83% African-American voters in 2020 and 76% in 2021.

41

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Special Ad audiences based on Whites also exhibited strong biases toward including more White voters, with sample shares similar to the Lookalike audiences shown in Figure 25. In 2020, White voters accounted for 83% of the overlap between the 2 million voter sample and the Special Ad audience based on White. In 2021, the sample share was 81% (Figure 29). In 2020, the Special Ad audience based on Whites with commonly given White names had the largest bias, with a 90% sample share of White voters (Figure 29). In 2021, the Special Ad audience based on Whites living in >90% White ZIP codes had the largest bias, with a sample share of 91% of White voters (Figure 29).
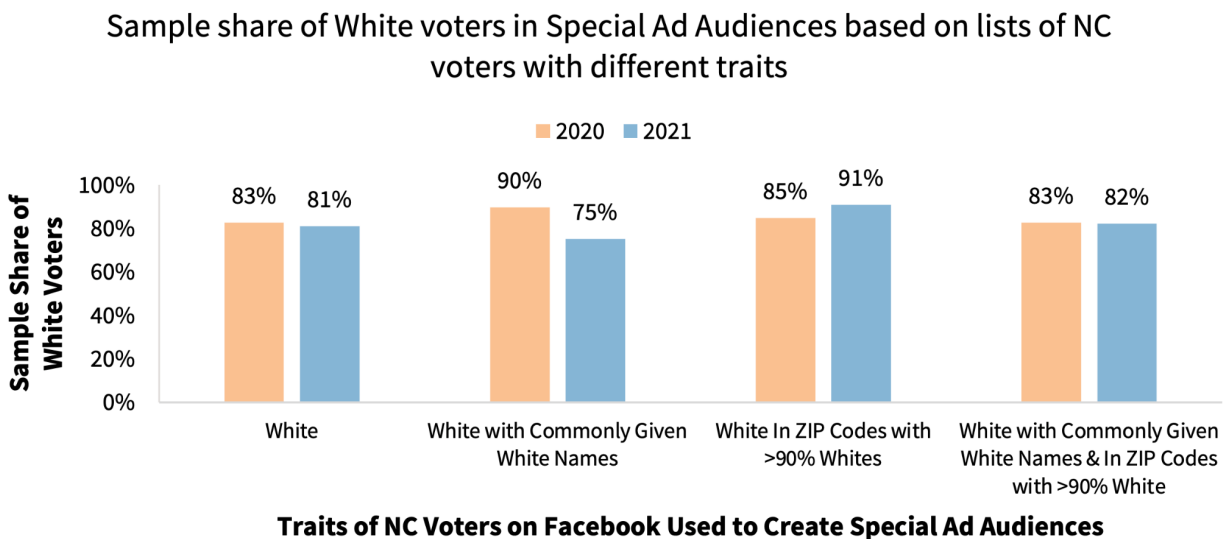
Sample share of White voters in Special Ad Audiences based on lists of NC voters with different traits
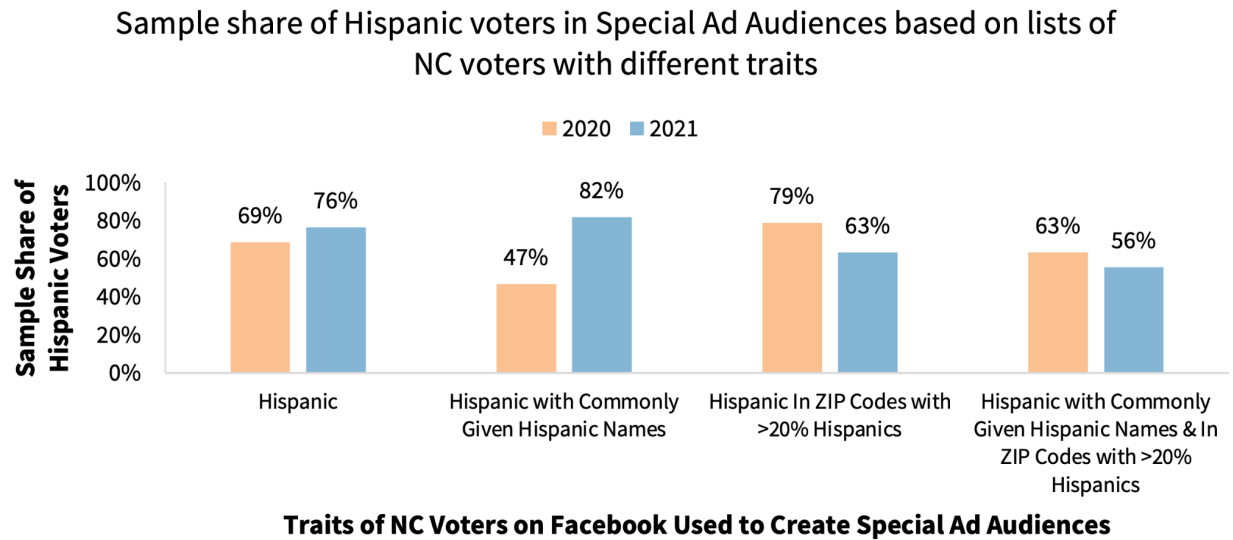

Figure 29. Sample Share of White Voters in Special Ad Audiences Based on Lists of NC Voters with Different Traits. In 2020, White voters were overrepresented, constituting 83% of the overlap between the 2 million voter sample and the Special Ad audience based on Whites. The same was true in 2021, with the sample share of White voters being 81%.

Interestingly, while in 2020 Special Ad audiences based on Asian voters demonstrated a bias toward Asians, that bias was significantly reduced in 2021. For example, in 2020, the sample share of Asian voters in Special Ad audiences based on different types of Asian voters started at 44% and increased to 67% when Asians with stereotypically Asian names and ZIP codes were used to create the Special Ad audience (Figure 30). In 2021, those sample shares go from 36% to 44% for the corresponding Special Ad audiences (Figure 30). Considering that the 150,000-voter sample used for each test has a baseline of 33% or 50,000 Asian voters, this means that Asian voters were only slightly overrepresented in the intersection of the voter sample and the relevant Special Ad audiences in 2021. This contrasts significantly with the very strong bias toward Asians, up to 100% sample share in one case, observed in Lookalike audiences based on similar customer lists of Asians shown in Figure 26.

Sample share of Asian voters in Special Ad Audiences based on lists of NC voters with different traits

■ 2020 ■ 2021

Figure 30. Sample Share of Asian Voters in Special Ad Audiences Based on Lists of NC Voters with Different Traits. Interestingly, while in 2020 Special Ad audiences based on Asian voters demonstrated a bias toward Asians, that bias was significantly reduced in 2021, with sample shares of Asian voters within 8 percentage points of the expected baseline of 33% for all four tests of Special Ad audiences.

I found that Special Ad audiences based on Hispanic source audiences with varying attributes were not consistently biased toward including more Hispanic voters across the two waves. In 2020, 3 of the 4 Special Ad audiences based on Hispanics had a sample share of Hispanic voters more than 10 percentage points above the 50% baseline in the 200,000-voter sample (Figure 31). In 2021, bias was also evident in 3 of the 4 Special Ad audiences, though a slightly different combination of audiences (Figure 31). For example, the Special Ad audience based on Hispanics with commonly given Hispanic names had a relatively low 47% sample share of Hispanic voters in 2020, while a Special Ad audience based on similar individuals in 2021 had a much higher 82% sample share of Hispanic voters (Figure 31). These patterns contrast with the consistent overrepresentation of Hispanic voters in Lookalike audiences based on Hispanics shown in Figure 27.

Sample share of Hispanic voters in Special Ad Audiences based on lists of NC voters with different traits

Figure 31. Sample Share of Hispanic Voters in Special Ad Audiences Based on Lists of NC Voters with Different Traits. Special Ad audiences based on different types of Hispanics did not consistently exhibit bias toward including more Hispanic voters in 2020 and 2021.

Discussion

This study shows that despite the advertising boycott and Facebook's response to pressure on civil rights issues in 2020, there remained multiple ways to use the tools of Facebook's advertising platform to discriminate by race and ethnicity in 2021. These include targeting options, Lookalike Audiences, and Special Ad Audiences.

Key Findings

The results of this study support the following key findings:

- **In 2021, Facebook's "African-American Culture" ad targeting option provided audiences with 75% fewer White users than the old "African American (US)" option removed in the previous year.** On the other hand, targeting options related to Asians and Hispanics included more non-Asians and Non-Hispanics in 2021. In 2020, 43% of African-American voters and 39% of White voters can be reached by the "African American (US)" targeting option. In 2021, those rates were 37% and 8% respectively for the "African-American Culture" targeting option (Figure 21). In 2021, the number of White voters reached by the "Asian American Culture" targeting option increased by 16,400 compared to the White voters reached by the "Asian American (US)" option in 2020, while the number of Asian voters reached by the two options declined by 200 over the same period (Figure 22). Similarly, the number of Hispanics reached by the "Hispanic American Culture" targeting option in 2021 was 9,200 fewer

than the reach of the "Hispanic (US – All)" option in 2020, while the number of Non-Hispanics increased by 9,600 over the same year (Figure 23).

- **Facebook's tools to help advertisers find similar users to their existing customers allow advertisers to create audiences that include more African-Americans or Whites depending on which racial group is dominant in an advertiser's customer list. In my tests, this was true for both the Lookalike Audience tool and the Special Ad Audience tool that Facebook designed to explicitly not use sensitive demographic attributes when finding similar users.** In 2020, the sample share of African-American voters in Lookalike audiences based on a source audience of African-Americans was 83%. This increased to 89% in 2021 (Figure 24). For Special Ad audiences based on the same list of African-Americans, in 2020, the sample share of African-American voters was 83%, and in 2021, it was 76% (Figure 28). Lookalike audiences based on Whites saw a slightly smaller degree of bias with a 73% sample share of White voters in 2020 and a 71% sample share in 2021 (Figure 25). Special Ad audiences based on White voters had a 83% sample share of White voters in 2020 and 81% in 2021 (Figure 29).

- **The degree of bias toward including more African-Americans or Whites in a Lookalike or Special Ad audience was larger when customer lists of individuals with racially stereotypical names or ZIP codes were supplied as the basis for each tool.** Lookalike audiences using a list of African-Americans with commonly given African-American names or living in ZIP codes with >50% African-Americans had sample shares of African-American voters up to 93% in 2020 and up to 94% in 2021 (Figure 24). Special Ad audiences based on similar customer lists had up to 97% sample shares of African-American voters in 2020 and 89% in 2021 (Figure 28). Lookalike audiences based on Whites with stereotypically White names or ZIP codes had sample shares of White voters up to 87% in 2020 and 84% in 2021 (Figure 25).

- **Similarly, Lookalike audiences can also become biased toward Asians, reaching up to 100% Asian in one case using a customer list of Asians with stereotypical names and ZIP codes. Lookalike audiences based on Hispanics overrepresented Hispanics. In a shift from 2020, Special Ad audiences based on Asians in 2021 did not appear to overrepresent them.** Figure 26 shows that the sample share of Asian voters in a Lookalike audience based on Asians is 51% in 2020 and 68% in 2021, compared to a baseline of 33% Asian in the 150,000-voter sample. The sample share of Asian voters increased to 83% in 2020 and 100% in 2021 for the Lookalike audience based on Asians with commonly given Asian names and living in ZIP codes with >20% Asians. Special Ad audiences based on Asians in 2020 had a 44% sample share of Asian voters. The share increased to 67% for audiences based on Asians with stereotypically Asian names and ZIP codes (Figure 30). On the other hand, in 2021, all of the Special Ad audiences based on Asians had sample shares of Asian voters within 8 percentage

points of the expected baseline of 33%. Finally, in 2020, the sample shares of Hispanic voters in Lookalike audiences based on different types of Hispanics ranged from 70-75% and in 2021 ranged from 69% to79%, well above the 50% baseline share of Hispanics in the 200,000 voter sample (Figure 27).

I found that removing "African-American (US)", "Asian American (US)", and "Hispanic (US – All)" as targeting options in August 2020 did not mean that the similar-sounding cultural interest groups available in 2021 could not be used for racial and ethnic targeting in 2021. In fact, 75% fewer Whites were targeted by the "African-American Culture" option in 2021, compared to the "African-American (US)" option in 2020. I also found that Lookalike and Special Ad audiences can become biased to include either more African-Americans or more Whites, based on which race is more dominant within the customer list used as the source audience, especially if the customers also have racially stereotypical names and ZIP codes. The sample shares of African-American voters reached up to 93%-94% for some Lookalike audiences in 2020 and 2021. Lookalike audiences based on Asians or Hispanics were also biased toward overrepresenting the corresponding demographic group in both 2020 and 2021, with the degree of bias reaching up to a 100% sample share of Asian voters for a Lookalike audience based on Asians with stereotypical Asian names and ZIP codes. One shift in 2021 was that Special Ad audiences based on Asians did not appear to overrepresent them, a change from 2020.

Limitations of This Study

While the results of this study describe how Facebook's ad targeting algorithms could bias the race or ethnicity of advertising audiences in 2020 and 2021, they are not able to fully describe why these biases occur. Even if an algorithm wasn't designed to be discriminatory, bias can occur through a series of different mechanisms.

- Underlying biases in the decisions made by humans may be reflected in the training data [60].

- The training data may be unrepresentative or incomplete [52, 88, 89].

- A reinforcement learning algorithm may become biased over time due to the biased behaviors of users [57, 90].

- Trade-offs in balancing an algorithm's performance in different fairness and accuracy metrics may result in biased outcomes for different demographics groups [65, 91, 92, 93, 94, 95, 96, 97, 98].

- A complex algorithmic decision-making system like Facebook's ad platform does not have just one algorithm but rather a series of different algorithms interacting with one

another, so even if individual algorithms are not biased on their own, their interactions may result in biased outcomes [99].

In this case, it is possible that a combination of multiple causes contributed to the biased outcomes observed in this study. Researchers have found that racial and ethnic groups tend to behave differently from each other online. They visit different websites [100, 101, 102, 103], follow different social media [104, 105], and even browse the web using different devices [106, 107]. In addition, as an online social network, Facebook also has data on the friends of each user. Researchers have found that Americans tend to have very racially homogeneous friend networks [108]. For White Americans, on average 91% of the members of their social network are also White [108]. For Black Americans, on average 83% of their social network are also Black [108]. Similarly, 75% of White Americans and 65% of Black Americans report having a core social network, defined as "people with whom they discuss important matters," entirely of their own race [108]. Future work would need to be done, ideally with deeper access to Facebook's data and systems than possible with this external digital audit approach, in order to explain why Facebook's advertising algorithms are biased.

Finally, this study focused on how Facebook's algorithms treated African-Americans, Whites, and Asians as racial groups and Hispanics and Non-Hispanics as ethnic groups. This focus was chosen because these demographic groups are the focus of Facebook's own ad targeting options and also exist as categories in the North Carolina voter data used for testing. Future studies may also examine other demographic groups such as Native Americans, multiracial individuals, and others. For studies of smaller demographic groups, the way that Facebook generally rounds the reach estimate on its ad planning tool to the nearest thousand may require other approaches than the set theory approach used here, since the reach estimate may not change by more than 1,000 users when using different ad targeting settings.

Implications for Facebook

While not every case of advertising discrimination is illegal or even potentially undesirable, this study highlights how the lack of transparency by Facebook to the public and to its advertisers about how its ad platform can potentially discriminate by race and ethnicity may be exploited by discriminatory advertisers while undermining the goals of non-discriminatory ones. For example, discriminatory advertisers may already know that the "African-American Culture" targeting option contains fewer White users than the "African-American (US)" option Facebook removed in 2020. Discriminatory advertisers may be using proxy variable techniques similar to the ones tested in this study based on racially stereotypical names and ZIP codes to create biased Lookalike and Special Ad audiences. On the other hand, non-discriminatory advertisers may be unintentionally choosing similar targeting settings while unaware of how Facebook's ad platform is carrying out racially and ethnically biased targeting on their behalf.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Investigations of Facebook by regulators have uncovered evidence that the tools of Facebook's ad platform examined in this study are being used to target political as well as housing, employment, and credit ads by race and ethnicity. For example, when Congress investigated Russian interference in the 2016 elections through tactics that included the purchase of political ads on Facebook, they discovered that Russian-affiliated advertisers used multicultural affinity groups as targeting options, including the "African-American (US)" option on Facebook [15]. In a different investigation of discrimination in housing, employment, and credit ads on Facebook by the Washington State's Attorney General's Office, which occurred over 20 months from 2016 to 2018, investigators found that "some third-party advertisers have chosen to target advertisements they created and placed on the Facebook platform using ethnic affinity targeting options" [27, 117].

Except when required by these regulatory investigations, Facebook has not disclosed relevant ad targeting information in its publicly accessible Ad Library that would be needed to detect racial discrimination by an advertiser. As of January 2021, Facebook's Ad Library publishes limited data about political, housing, employment and credit-related ads [109]. For political ads, Facebook publishes metadata about how much was spent, how many viewed the ad, and who saw the ad by gender and state (Appendix C). Facebook does not release data in the Ad Library about which targeting options, such as the racially affiliated interest groups studied here, were used. It also does not release information about the racial and ethnic breakdown of who saw an ad. For housing, employment, and credit-related ads, no metadata about who was targeted or who saw the ad is released (Appendix C). This is especially problematic in light of the final version of the U.S. Department of Housing and Urban Development's (HUD) "Implementation of the Fair Housing Act's Disparate Impact Standard" published on September 24, 2020. This regulation requires a plaintiff to present evidence of a "robust causal link" in order to bring a disparate impact discrimination lawsuit in the first place [110]. This study demonstrates how Facebook's targeting options—Lookalike Audiences, and Special Ad Audiences—can be used to discriminate by race or ethnicity, but Facebook's Ad Library doesn't currently release any data that could document a "robust causal link" between how an advertiser is using Facebook's tools and the discriminatory impact on who sees their ads.

Currently, Facebook's anti-discrimination efforts are concentrated on limiting the targeting options for "Special Ads" related to housing, employment, or credit. Facebook disables the use of such sensitive targeting options for Special Ads as multicultural affinity groups in 2020 and cultural interest groups in 2021 (Appendix C). Facebook only allows an advertiser to add more attributes as Detailed Targeting options and does not allow an advertiser to exclude any attribute (Appendix C). Facebook also provides notices on its ad planning tool to encourage advertisers to not discriminate when targeting a Special Ad (Appendix C). Finally, Facebook does not allow Special Ads to use regular Lookalike audiences. Rather, they have to target Special Ad audiences, which Facebook designed to not use sensitive demographic attributes such as "age, gender or ZIP code" in considering which users to include [2]. However, this study found that in 2020 and 2021 Special Ad audiences could become racially biased at rates

similar to Lookalike audiences when an advertiser used the same demographically homogeneous customer list to create both types of audiences. Regular ads not related to housing, employment, or credit do not face any of these restrictions or notices.

In order to better tackle discrimination in the future, Facebook can leverage its data and analytical capabilities to better detect potential racial and ethnic discrimination for both "Special" and regular ads. Right now, Facebook's ad planning tool provides daily reach estimates given any combination of different ad targeting options, Custom audiences, Lookalike audiences, or Special Ad audiences, which is how I collected data about Facebook's ad platform for this study. As a potential feature, Facebook can enrich its estimated reach report by displaying who will see an ad on the basis of race, ethnicity, gender, age, geography, and other categories. If a particular Custom, Lookalike, or Special Ad audience is racially or ethnically biased, Facebook can flag those audiences when they are first created in order to notify the advertiser and potentially limit their usage.

In order to carry out these digital audits for potential racial and ethnic biases in an advertiser's target audiences, Facebook has a number of ways to collect or infer racial and ethnic data about its users. One option is similar to the North Carolina voter registration form, which asks a user to voluntarily provide their race and ethnicity. Facebook currently requests gender and date of birth on its account sign up page and includes an option for a "Custom" gender where a user can select their preferred pronouns and textbox for a preferred gender label. Facebook could adapt this approach to collect racial and ethnic data directly from its users. It is possible that many would find it unappealing to give Facebook more data about themselves, given past controversies over how Facebook has handled user data, such as the Cambridge Analytica scandal [111]. Another option is for Facebook to infer the data indirectly by following other examples in the tech industry such as Airbnb's Project Lighthouse, which was launched in 2020 to study the racial experience gap for guests and hosts on Airbnb [112]. Project Lighthouse used a third-party contractor to assess the perceived race of an individual based on their profile picture and name [112]. Another approach is to use the name alone to infer race and ethnicity by using algorithmic approaches such as the ethnicolr Python library [78] or the U.S. Census Bureau's Frequently Occurring Surnames dataset [79]. In fact, there is precedent at Facebook for doing exactly this type of analysis. In December 2009, Facebook researchers Lars Backstrom, Jonathan Chang, Cameron Marlow, and Itamar Rosenn published a paper on the race and ethnicity of Facebook users from January 2006 to January 2009 by comparing the last names of users to the U.S. Census's Frequently Occurring Surnames dataset [113]. They found that Facebook was becoming increasingly diverse over time by having more African-American and Hispanic users, a result cited in a *The Wall Street Journal* article titled "Facebook Touts Diversity of Its Members" [114].

How "Fairness Through Unawareness" Doesn't Prevent Algorithmic Discrimination

Finally, this study has ramifications beyond Facebook in terms of how to detect and address the issue of algorithmic discrimination in an increasingly digital world. Many of the

anti-discrimination changes that Facebook implemented in recent years to its advertising platform are examples of trying to achieve "fairness through unawareness" [63], the idea that discrimination is prevented by eliminating the use of protected class variables or close proxies. For example, Facebook explicitly created the Special Ad Audiences tool – as an alternative to Lookalike Audiences – to block the use of sensitive attributes such as "age, gender or ZIP code" in considering which users are similar enough to the source audience to be included [2]. However, this study demonstrates that Special Ad audiences based on African-Americans or Whites can be biased toward the race that is more dominant in the customer list used to create the audience, just like the corresponding Lookalike audiences. In fact, even though Facebook designed its Special Ad Audience tool to explicitly not use ZIP codes as part of its algorithm, in 2021, the sample shares of African-American voters were 12 percentage points higher for Special Ad audiences based on African-Americans with stereotypically African-American ZIP codes versus African-Americans from anywhere in North Carolina (Figure 28). Likewise, in 2021, the sample shares of White voters were 10 percentage points higher for Special Ad audiences based on Whites from >90% White ZIP codes versus Whites from anywhere in North Carolina (Figure 29).

Statistics research labels this phenomenon as the Rashomon effect or the multiplicity effect [115]. This means that given a large dataset with many variables, there exists a large number of potential models that can perform approximately equally as well as a prohibited model that uses protected class variables [63]. Thus, even though the Special Ad Audiences algorithm for finding similar users to a customer list does not use demographic attributes in the same way as the Lookalike Audiences algorithm, the two algorithms may end up making functionally comparable decisions on which users are considered similar enough to be included.

In recent years, there have been regulatory efforts to promote "fairness through unawareness" as a means of protecting companies from the liability of a discrimination lawsuit. This study illustrates that Facebook's "fairness through unawareness" changes, such as its Special Ad Audiences tool, do not necessarily prevent discrimination on the platform, though the tool may have removed the ability for plaintiffs to successfully sue Facebook for discrimination if the proposed federal policies were implemented. The initial language of the "Implementation of the Fair Housing Act's Disparate Impact Standard" rule by the U.S. Department of Housing and Urban Development's (HUD), which was drafted in response to the Supreme Court's *Texas v. Inclusive Communities* decision and released on August 19, 2019, stated that a defendant may successfully argue its model is not discriminatory if the model does not rely on "factors that are substitutes or close proxies for protected classes under the Fair Housing Act" [116]. After public comments criticized this language, it was removed in the final rule published on September 24, 2020 [110]. This study found that Facebook's Special Ad Audience tool can enable racial and ethnic biases in ad audiences even though it does not rely on sensitive attributes that are likely related to protected classes, which is the defense criteria established in the initial language of HUD's disparate impact rule.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

References

1. Wagner K. Digital advertising in the US is finally bigger than print and television. Vox. February 20, 2019.

2. Facebook. Updates To Housing, Employment and Credit Ads in Ads Manager. 2019.

3. Angwin J and Parris Jr T. Facebook Lets Advertisers Exclude Users by Race. ProPublica. October 28, 2016. https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race.

4. Angwin J, Tobin A, and Varner M. Facebook (Still) Letting Housing Advertisers Exclude Users by Race. ProPublica. November 21, 2017. https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin.

5. Sapiezynski P, Ghosh A, Kaplan L, Mislove A, and Rieke A. Algorithms that "Don't See Color": Comparing Biases in Lookalike and Special Ad Audiences. 2019. http://arxiv.org/abs/1912.07579.

6. NATIONAL FAIR HOUSING ALLIANCE; FAIR HOUSING JUSTICE CENTER, INC.; HOUSING OPPORTUNITIES PROJECT FOR EXCELLENCE, INC.; FAIR HOUSING COUNCIL OF GREATER SAN ANTONIO v. Facebook. United States District Court, Southern District of New York. https://nationalfairhousing.org/wp-content/uploads/2019/03/2018-06-25-NFHA-v.-Facebook.-First-Amended-Complaint.pdf.

7. Sherwin G. How Facebook Is Giving Sex Discrimination in Employment Ads a New Life. ACLU. 2018. https://www.aclu.org/blog/womens-rights/womens-rights-workplace/how-facebook-giving-sex-discrimination-employment-ads-new.

8. Communications Workers of America vs. T-Mobile US, Inc. Amazon.com, Inc., Cox Communications, Inc., Cox Media Group, LLC, and similarly situated employers and employment agencies, DOES 1 through 1,000. United States District Court, Northern District of California. https://doi.org/https://www.onlineagediscrimination.com/sites/default/documents/og-cwa-complaint.pdf.

9. HUD Public Affairs. HUD CHARGES FACEBOOK WITH HOUSING DISCRIMINATION OVER
    COMPANY'S TARGETED ADVERTISING PRACTICES. Department of Housing and Urban
    Development. 2019.
    https://www.hud.gov/press/press_releases_media_advisories/HUD_No_19_035.

10. Statt N. Facebook signs agreement saying it won't let housing advertisers exclude
    users by race. The Verge. July 24, 2018.
    https://www.theverge.com/2018/7/24/17609178/facebook-racial-dicrimination-ad-tar
    geting-washington-state-attorney-general-agreement.

11. Timber C and Stanley-Becker I. Black voters are being targeted in disinformation
    campaigns, echoing the 2016 Russian playbook. The Washington Post. August 26,
    2020.
    https://www.washingtonpost.com/technology/2020/08/26/race-divisions-highlighted-
    disinformation-2016/.

12. Frenkel S and Barnes J. Russians Again Targeting Americans With Disinformation,
    Facebook and Twitter Say. The New York Times. September 1, 2020.
    https://www.nytimes.com/2020/09/01/technology/facebook-russia-disinformation-ele
    ction.html.

13. Horwitz J. Political Groups Elude Facebook's Election Controls, Repost False Ads. The
    Wall Street Journal. November 1, 2020.
    https://www.wsj.com/articles/political-groups-elude-facebooks-election-controls-rep
    ost-false-ads-11604268856.

14. Stanley-Becker I. Disinformation campaign stokes fears about mail voting, using
    LeBron James image and boosted by Trump-aligned group. The Washington Post.
    August 20, 2020.
    https://www.washingtonpost.com/politics/disinformation-campaign-stokes-fears-abo
    ut-mail-voting-using-lebron-james-image-and-boosted-by-trump-aligned-group/2020
    /08/20/fcadf382-e2e2-11ea-8181-606e603bb1c4_story.html?itid=lk_inline_manual_2.

15. DiResta R et al. The Tactics & Tropes of the Internet Research Agency, New Knowledge.
    2019.
    https://int.nyt.com/data/documenthelper/533-read-report-internet-research-agency/7
    87lea6d5bafbf19/optimized/full.pdf#page=1.

16. Wagner K and Nix N. Facebook Scorned by Advocacy Groups After Zuckerberg Meeting.
    Bloomberg. July 7, 2020.
    https://www.bloomberg.com/news/articles/2020-07-07/facebook-denounced-by-civil-
    rights-group-over-speech-policies.

17. Hsu T and Lutz E. More Than 1,000 Companies Boycotted Facebook. Did It Work? The New York Times. 1. August 2020. https://www.nytimes.com/2020/08/01/business/media/facebook-boycott.html.

18. Murphy L and Relman Colfax. Facebook's Civil Rights Audit – Final Report. 2020.

19. Sandberg S. Making Progress on Civil Rights – But Still a Long Way to Go. Facebook. 2020. https://about.fb.com/news/2020/07/civil-rights-audit-report/.

20. Wagner K. Facebook Limits Ad Targeting That Some Linked to Race. Bloomberg. August 11, 2020. https://www.bloomberg.com/news/articles/2020-08-11/facebook-further-limits-advertisers-ability-to-target-by-race.

21. Newitz A. Facebook's ad platform now guesses at your race based on your behavior. Ars Technica. March 18, 2016.

22. Weise E. Facebook sued for housing and employment bias. USA Today. November 7, 2016. https://www.usatoday.com/story/tech/news/2016/11/07/facebook-sued-housing-and-employement-bias/93424824/.

23. SUZANNE-JULIETTE MOBLEY, KAREN SAVAGE, VICTOR ONUOHA, on behalf of themselves and all others similarly situated v. Facebook. United States District Court, Northern District of California. 2016. https://pdfserver.amlaw.com/nlj/FacebookFairHousing.pdf.

24. Angwin J and Tobin A. Fair Housing Groups Sue Facebook for Allowing Discrimination in Housing Ads. ProPublica. March 27, 2018. https://www.propublica.org/article/facebook-fair-housing-lawsuit-ad-discrimination.

25. Gosselin P. New Allegations Added to Lawsuit on How Facebook's Targeting Tools Helped Advertisers Exclude Older Workers. ProPublica. May 30, 2018. https://www.propublica.org/article/new-allegations-lawsuit-on-how-facebook-targeting-tools-exclude-older-workers.

26. Tobin A. Facebook Promises to Bar Advertisers From Targeting Ads by Race or Ethnicity. Again. ProPublica. July 25, 2018. https://www.propublica.org/article/facebook-promises-to-bar-advertisers-from-targeting-ads-by-race-or-ethnicity-again.

27. Rosenberg M. Facebook must stop advertisers from excluding people from viewing ads for housing, jobs and more. Seattle Times. July 24, 2018. https://www.seattletimes.com/business/technology/facebook-must-end-discriminatory-ad-practice-under-deal-with-washington-attorney-general/.

28. Boyd R et al. Characterizing the Internet Research Agency's Social Media Operations During the 2016 U.S. Presidential Election using Linguistic Analyses. 1–9. 2018. https://doi.org/10.31234/osf.io/ajh2q.

29. Ribeiro F N et al. On microtargeting socially divisive ads: A case study of Russia-linked Ad campaigns on Facebook. FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency. 140–149. 2019. https://doi.org/10.1145/3287560.3287580.

30. Lecher C. Trump campaign using targeted Facebook posts to discourage black Americans from voting. The Verge. October 27, 2016. https://www.theverge.com/2016/10/27/13434246/donald-trump-targeted-dark-facebook-ads-black-voters.

31. Sanz C. Misinformation targeted Latino voters in the 2020 election. ABC News. November 21, 2020. https://abcnews.go.com/Politics/latino-voters-misinformation-targets-election-2020/story?id=74189342.

32. Romero L. Florida's Latino voters being bombarded with right-wing misinformation, experts and advocates say. ABC News. October 20, 2020. https://abcnews.go.com/Politics/floridas-latino-voters-bombarded-wing-misinformation-advocates/story?id=73707056.

33. Ryan-Mosley T. "It's been really, really bad": How Hispanic voters are being targeted by disinformation. MIT Technology Review. October 12, 2020. https://www.technologyreview.com/2020/10/12/1010061/hispanic-voter-political-targeting-facebook-whatsapp/.

34. Rodriguez S and Caputo M. "This is f---ing crazy": Florida Latinos swamped by wild conspiracy theories. Politico. September 14, 2020. https://www.politico.com/news/2020/09/14/florida-latinos-disinformation-413923.

35. Wong Q. Facebook ad boycott: Why big brands "hit pause on hate." CNET. July 30, 2020. https://www.cnet.com/news/facebook-ad-boycott-how-big-businesses-hit-pause-on-hate/.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

36. Guynn J. What civil rights groups want from Facebook boycott: Stop hate speech and harassment of Black users. USA Today. July 7, 2020. https://www.usatoday.com/story/tech/2020/07/07/facebook-ad-boycott-racism-harassment-hate-african-americans/5385514002/.

37. Fryer R G and Levitt S D. The causes and consequences of distinctively Black names. Quarterly Journal of Economics. Vol 119. No 3. 767–806. 2004. https://doi.org/10.1162/0033553041502180.

38. Imai K and Khanna K. Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records. Political Analysis. Vol 24. No 2. 263–272. 2016. https://doi.org/DOI: 10.1093/pan/mpw001.

39. Bertrand M and Mullainathan S. Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. American Economic Review. Vol 94. No 4. 991–1013. 2004. https://doi.org/10.1257/0002828042002561.

40. Kang S K, DeCelles K A, Tilcsik A, and Jun S. Whitened Résumés: Race and Self-Presentation in the Labor Market. Administrative Science Quarterly. Vol 61. No 3. 469–502. March 17, 2016. https://doi.org/10.1177/0001839216639577.

41. Quillian L, Pager D, Hexel O, and Midtbøen A H. Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. Proceedings of the National Academy of Sciences. Vol 114. No 41. 10870 LP – 10875. October 10, 2017. https://doi.org/10.1073/pnas.1706255114.

42. Gaddis S M. An Introduction to Audit Studies in the Social Sciences BT - Audit Studies: Behind the Scenes with Theory, Method, and Nuance. Gaddis S M, Editor. Springer International Publishing. Cham. 2018. 3–44.

43. WHITE A R, NATHAN N L, and FALLER J K. What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials. American Political Science Review. Vol 109. No 1. 129–142. 2015. https://doi.org/DOI: 10.1017/S0003055414000562.

44. Hanson A and Hawley Z. Where does racial discrimination occur? An experimental analysis across neighborhood and housing unit characteristics. Regional Science and Urban Economics. Vol 44. 94–106. 2014. https://doi.org/https://doi.org/10.1016/j.regsciurbeco.2013.12.001.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

45. Capps K and Rabinowitz K. How the Fair Housing Act Failed Black Homeowners. City Lab. April 11, 2018.

46. Richardson J, Mitchell B, and West N. Home Mortgage and Small Business Lending in Baltimore and Surrounding Areas. 2015.

47. Glantz A and Martinez E. For people of color, banks are shutting the door to homeownership. Reveal News. February 15, 2018.

48. Chetty R, Hendren N, Jones M, and Porter S. Race and Economic Opportunity in the United States: An Intergenerational Perspective. 24441. 2019.

49. Turner Lee N, Resnick P, and Barton G. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. 2019.

50. Zarsky T. Understanding Discrimination in the Scored Society. Washington Law Review. Vol 89. No 4. 2014.

51. Sweeney L. Discrimination in online Ad delivery. Communications of the ACM. Vol 56. No 5. 44–54. 2013. https://doi.org/10.1145/2447976.2447990.

52. Buolamwini J and Gebru T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. 2018. Vol 81. 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html.

53. Hannak A, Soeller G, Lazer D, Mislove A, and Wilson C. Measuring Price Discrimination and Steering on E-Commerce Web Sites. in *Proceedings of the 2014 Conference on Internet Measurement Conference*. 2014. 305–318. https://doi.org/10.1145/2663716.2663744.

54. Kay M, Matuszek C, and Munson S A. Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015. 3819–3828. https://doi.org/10.1145/2702123.2702520.

55. Kulshrestha J et al. Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media. in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2017. 417–432. https://doi.org/10.1145/2998181.2998321.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

56. Robertson R E, Jiang S, Joseph K, Friedland L, Lazer D, and Wilson C. Auditing Partisan Audience Bias within Google Search. Proc. ACM Hum.-Comput. Interact. Vol 2. No CSCW. November 2018. https://doi.org/10.1145/3274417.

57. Datta A, Datta A, Makagon J, Mulligan D K, and Tschantz M C. Discrimination in Online Advertising: A Multidisciplinary Inquiry. in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. 2018. Vol 81. 20–34. http://proceedings.mlr.press/v81/datta18a.html.

58. Chen L, Ma R, Hannák A, and Wilson C. Investigating the Impact of Gender on Rank in Resume Search Engines. in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018. 1–14. https://doi.org/10.1145/3173574.3174225.

59. Hannák A, Wagner C, Garcia D, Mislove A, Strohmaier M, and Wilson C. Bias in Online Freelance Marketplaces: Evidence from TaskRabbit and Fiverr. in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2017. 1914–1933. https://doi.org/10.1145/2998181.2998327.

60. Dastin J. Amazon scraps secret AI recruiting tool that showed bias against women. Reuter. October 10, 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

61. Bolukbasi T, Chang K-W, Zou J, Saligrama V, and Kalai A T. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. in *NIPS*. 2016. NIPS. https://www.microsoft.com/en-us/research/publication/quantifying-reducing-stereotypes-word-embeddings/.

62. Obermeyer Z, Powers B, Vogeli C, and Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. Vol 366. No 6464. 447 LP – 453. October 25, 2019. https://doi.org/10.1126/science.aax2342.

63. Xiang A. Reconciling Legal and Technical Approaches to Algorithmic Bias. Tennessee Law Review. Vol 88. No 3. 2021.

64. Federal Reserve. Consumer Compliance Handbook. 2017.

65. Feldman M, Friedler S A, Moeller J, Scheidegger C, and Venkatasubramanian S. Certifying and Removing Disparate Impact. in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2015. 259–268. https://doi.org/10.1145/2783258.2783311.

66. Zafar M B, Valera I, Gomez-Rodriguez M, and Gummadi K P. Fairness Constraints: A Flexible Approach for Fair Classification. Journal of Machine Learning Research. Vol 20. No 75. 1–42. 2019. http://jmlr.org/papers/v20/18-262.html.

67. Facebook Newsroom. Improving Enforcement and Promoting Diversity: Updates to Ads Policies and Tools. Facebook. 2017. https://about.fb.com/news/2017/02/improving-enforcement-and-promoting-diversity-updates-to-ads-policies-and-tools/.

68. Tobin A and Merrill J. Besieged Facebook Says New Ad Limits Aren't Response to Lawsuits. ProPublica2. August 23, 18AD. https://www.propublica.org/article/facebook-says-new-ad-limits-arent-response-to-lawsuits.

69. Facebook Business. Keeping Advertising Safe and Civil. Facebook2. 2018. https://www.facebook.com/business/news/keeping-advertising-safe-and-civil.

70. Sherwin G and Bhandari E. Facebook Settles Civil Rights Cases by Making Sweeping Changes to Its Online Ad Platform. ACLU. 2019. https://www.aclu.org/blog/womens-rights/womens-rights-workplace/facebook-settles-civil-rights-cases-making-sweeping.

71. National Fair Housing Alliance. Facebook Settlement: Civil Rights Advocates Settle Lawsuit with Facebook: Transforms Facebook's Platform Impacting Millions of Users. 2019. https://nationalfairhousing.org/facebook-settlement/.

72. ACLU. Facebook Agrees to Sweeping Reforms to Curb Discriminatory Ad Targeting Practices. 2019. https://www.aclu.org/press-releases/facebook-agrees-sweeping-reforms-curb-discriminatory-ad-targeting-practices.

73. Berman G. STATEMENT OF INTEREST OF THE UNITED STATES OF AMERICA in National Fair Housing Alliance v. Facebook, Inc. UNITED STATES DISTRICT COURT SOUTHERN DISTRICT OF NEW YORK. New York. 2018. https://www.justice.gov/crt/case-document/file/1089231/download.

74. Tse A and Winslow S. UNITED STATES' STATEMENT OF INTEREST in ONUOHA v. FACEBOOK, INC. UNITED STATES DISTRICT COURT NORTHERN DISTRICT OF CALIFORNIA. San Jose. 2018. https://www.justice.gov/crt/case-document/file/1112561/download.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

75. Abril D. The Facebook ad boycott ended months ago. But some big companies continue the fight. Fortune. November 7, 2020. https://fortune.com/2020/11/07/facebook-ad-boycott-big-brands-lego-clorox-verizon-microsoft-hp/.

76. Newton C and Schiffer Z. What a damning civil rights audit missed about Facebook. The Verge. July 10, 2020. https://www.theverge.com/interface/2020/7/10/21318718/facebook-civil-rights-audit-critique-size-congress.

77. Dwoskin E, Tiku N, and Kelly H. Facebook to start policing anti-Black hate speech more aggressively than anti-White comments, documents show. The Washington Post. December 3, 2020. https://www.washingtonpost.com/technology/2020/12/03/facebook-hate-speech/.

78. Sood G and Laohaprapanon S. Predicting Race and Ethnicity From the Sequence of Characters in a Name. May 5, 2018. https://arxiv.org/abs/1805.02109.

79. U.S. Census Bureau. Frequently Occurring Surnames from the 2010 Census. Genealogy Data. 2010. https://www.census.gov/topics/population/genealogy/data/2010_surnames.html.

80. Yoon S, Falzon L, Anderson N B, and Davidson K W. A look at the increasing demographic representation within behavioral medicine. Journal of Behavioral Medicine. Vol 42. No 1. 57–66. 2019. https://doi.org/10.1007/s10865-018-9983-y.

81. Yeung N, Lai J, and Luo J. Face Off: Polarized Public Opinions on Personal Face Mask Usage during the COVID-19 Pandemic. October 31, 2020. http://arxiv.org/abs/2011.00336.

82. Shiffer-Sebba D and Behrman J. Gender and Wealth in Demographic Research: A Research Brief on a New Method and Application. Population Research and Policy Review. 2020. https://doi.org/10.1007/s11113-020-09603-w.

83. Nations J M and Martin I W. Racial Context and Political Support for California School Taxes. Social Science Quarterly. Vol 101. No 6. 2220–2237. October 1, 2020. https://doi.org/https://doi.org/10.1111/ssqu.12869.

84. Bertolero M A et al. Racial and ethnic imbalance in neuroscience reference lists and intersections with gender. bioRxiv. 2020.10.12.336230. January 1, 2020. https://doi.org/10.1101/2020.10.12.336230.

85. U.S. Census Bureau. American Community Survey (ACS). Our Surveys & Programs.
    2019. https://www.census.gov/programs-surveys/acs.

86. U.S. Census Bureau. U.S. Census Bureau QuickFacts: North Carolina. QuickFacts. 2020.
    https://www.census.gov/quickfacts/fact/table/NC/PST045219.

87. U.S. Census Bureau. Annual Estimates of the Resident Population by Sex, Age, Race
    Alone or in Combination, and Hispanic Origin for the United States: April 1, 2010 to July
    1, 2019. Newsroom. 2019.
    https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.ht
    ml.

88. Wiggers K. IBM releases Diversity in Faces, a dataset of over 1 million annotations to
    help reduce facial recognition bias. VentureBeat. January 29, 2019.
    https://venturebeat.com/2019/01/29/ibm-releases-diversity-in-faces-a-dataset-of-over
    -1-million-annotations-to-help-reduce-facial-recognition-bias/.

89. Lohr S. Facial Recognition Is Accurate, if You're a White Guy. The New York Times.
    February 9, 2018.
    https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-int
    elligence.html.

90. Emerging Technology from the ArXiv. Racism is Poisoning Online Ad Delivery, Says
    Harvard Professor. MIT Technology Review. February 4, 2013.
    https://www.technologyreview.com/2013/02/04/253879/racism-is-poisoning-online-a
    d-delivery-says-harvard-professor/.

91. Corbett-Davies S, Pierson E, Feller A, Goel S, and Huq A. Algorithmic Decision Making
    and the Cost of Fairness. in *Proceedings of the 23rd ACM SIGKDD International
    Conference on Knowledge Discovery and Data Mining*. 2017. 797–806.
    https://doi.org/10.1145/3097983.3098095.

92. Kleinberg J, Ludwig J, Mullainathan S, and Rambachan A. Algorithmic Fairness. AEA
    Papers and Proceedings. Vol 108. 22–27. 2018.
    https://doi.org/10.1257/pandp.20181018.

93. Hardt M, Price E, and Srebro N. Equality of Opportunity in Supervised Learning. in
    *Proceedings of the 30th International Conference on Neural Information Processing
    Systems*. 2016. 3323–3331.

94. Zafar M B, Valera I, Gomez Rodriguez M, and Gummadi K P. Fairness Beyond Disparate
    Treatment & Disparate Impact: Learning Classification without Disparate

Mistreatment. in *Proceedings of the 26th International Conference on World Wide Web*.
2017. 1171–1180. https://doi.org/10.1145/3038912.3052660.

95. Angwin J, Larson J, Mattu S, and Kirchner L. Machine Bias. ProPublica. May 23, 2016.
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sente
ncing.

96. Dieterich W, Mendoza C, and Brennan T. COMPAS Risk Scales: DemonstratingAccuracy
Equity and Predictive Parity. 2016.

97. Chouldechova A. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism
Prediction Instruments. Big Data. Vol 5. No 2. 153–163. June 1, 2017.
https://doi.org/10.1089/big.2016.0047.

98. Kleinberg J. Inherent Trade-Offs in Algorithmic Fairness. in *Abstracts of the 2018 ACM
International Conference on Measurement and Modeling of Computer Systems*. 2018. 40.
https://doi.org/10.1145/3219617.3219634.

99. Dwork C and Ilvento C. Fairness Under Composition. in *ITCS*. 2019.

100. Sharad Goel J M H M I S. Who Does What on the Web: A Large-Scale Study of
Browsing Behavior. Proceedings of the Sixth International AAAI Conference on Weblogs
and Social Media. 1–8. 2012.
http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/viewPDFInterstitial/4660/
4975%5Cnpapers2://publication/uuid/F3306966-AB04-4CF1-ACCE-A7EF49E1282C.

101. De Bock K and Van Den Poel D. Predicting website audience demographics forweb
advertising targeting using multi-website clickstream data. Fundamenta Informaticae.
Vol 98. No November. 49–70. 2010. https://doi.org/10.3233/FI-2010-216.

102. Ying J J, Chang Y, Huang C, and Tseng V S. Demographic Prediction Based on User's
Browsing Behaviors. Research.Nokia.Com. Vol 2012. No 1. 1–6. 2012.
http://research.nokia.com/files/public/mdc-final241-ying.pdf%5Cnfile:///Users/orasan
en/Documents/Papers/Ying/research.nokia.com.

103. Sweeney L. Online ads roll the dice. Tech@FTC. 2014.
https://www.ftc.gov/news-events/blogs/techftc/2014/09/online-ads-roll-dice.

104. Messias J, Vikatos P, and Benevenuto F. White, Man, and Highly Followed: Gender
and Race Inequalities in Twitter. in *Proceedings of the International Conference on Web
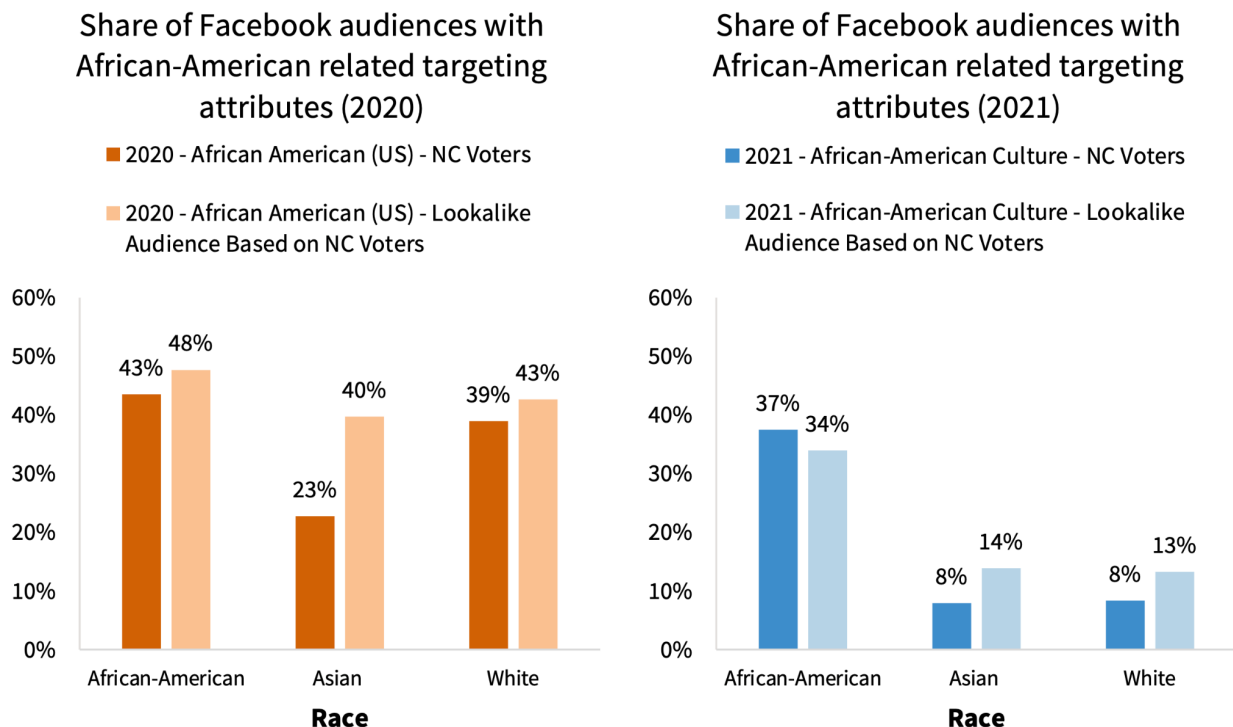Intelligence*. 2017. 266–274. https://doi.org/10.1145/3106426.3106472.

105.   Vikatos P, Messias J, Miranda M, and Benevenuto F. Linguistic Diversities of Demographic Groups in Twitter. in *Proceedings of the 28th ACM Conference on Hypertext and Social Media*. 2017. 275–284. https://doi.org/10.1145/3078714.3078742.

106.   Fairlie R. Have we finally bridged the digital divide? Smart phone and Internet use patterns by race and ethnicity. First Monday. Vol 22. 2017.

107.   Tsetsi E and Rains S A. Smartphone Internet access and use: Extending the digital divide and usage gap. Mobile Media & Communication. Vol 5. No 3. 239–255. June 13, 2017. https://doi.org/10.1177/2050157917708329.

108.   Cox D, Navarro-Rivera J, and Jones R. Race, Religion, and Political Affiliation of Americans' Core Social Networks. PRRI. Vol 5. 1–5. 2014. https://www.prri.org/research/poll-race-religion-politics-americans-social-networks/.

109.   Facebook. Ad Library. 2020.

110.   Department of Housing and Urban Development. HUD's Implementation of the Fair Housing Act's Disparate Impact Standard. 2020. https://www.federalregister.gov/documents/2020/09/24/2020-19887/huds-implementation-of-the-fair-housing-acts-disparate-impact-standard.

111.   Confessore N. Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. The New York Times. April 4, 2018. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

112.   Airbnb. A new way we're fighting discrimination on Airbnb. Airbnb Resource Center. 2020. https://www.airbnb.com/resources/hosting-homes/a/a-new-way-were-fighting-discrimination-on-airbnb-201.

113.   Chang J, Rosenn I, Backstrom L, and Marlow C. ePluribus: Ethnicity on Social Networks. 2010. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1534.

114.   Taylor M. Facebook Touts Diversity of Its Members. The Wall Street Journal. December 18, 2009. https://www.wsj.com/articles/BL-DGB-9541.

115.   Semenova L, Rudin C, and Parr R. A study in Rashomon curves and volumes: A new perspective on generalization and model simplicity in machine learning. August 5, 2019. https://arxiv.org/abs/1908.01755.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

116.   Department of Housing and Urban Development. HUD's Implementation of the Fair Housing Act's Disparate Impact Standard. 2019. https://www.federalregister.gov/documents/2019/08/19/2019-17542/huds-implementation-of-the-fair-housing-acts-disparate-impact-standard.

117.   State of Washington King County Superior Court. In re: Facebook, Inc. Assurance of Discontinuance. 2018. https://agportal-s3bucket.s3.amazonaws.com/uploadedfiles/Another/News/Press_Releases/2018_07_23%20AOD.pdf

Appendix

Appendix A – The Share of Multicultural Affinity Groups (2020) vs. Cultural Interest Groups (2021) in Lookalike Audiences

In 2020, I found that Lookalike audiences had similar shares of users with the "African-American (US)" attribute at 40–48% regardless of whether a list of only African-American, Asian, or White voters was used to create the Lookalike audience (Figure 32). In 2021, Lookalike audiences based on Asian or White voters had shares of users interested in "African-American Culture" that were far lower, at 13-14%, than the 34% share of the Lookalike audience based on African-American voters (Figure 32).

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Figure 32. Share of Facebook Audiences with African-American Related Targeting Attributes in 2020 and 2021.

I found the opposite pattern for Facebook's Asian-American related targeting attributes. In 2020, 2.1% of the Lookalike audience based on Asian voters could be reached by targeting "Asian American (US)" while only 0.2% of the Lookalike audiences based on African-American or White voters (Figure 33) can be reached using this target selection. In 2021, approximately 2-3% of the Lookalike audiences based on African-American, Asian, or White voters could be targeted by Facebook's "Asian American Culture" option (Figure 33).
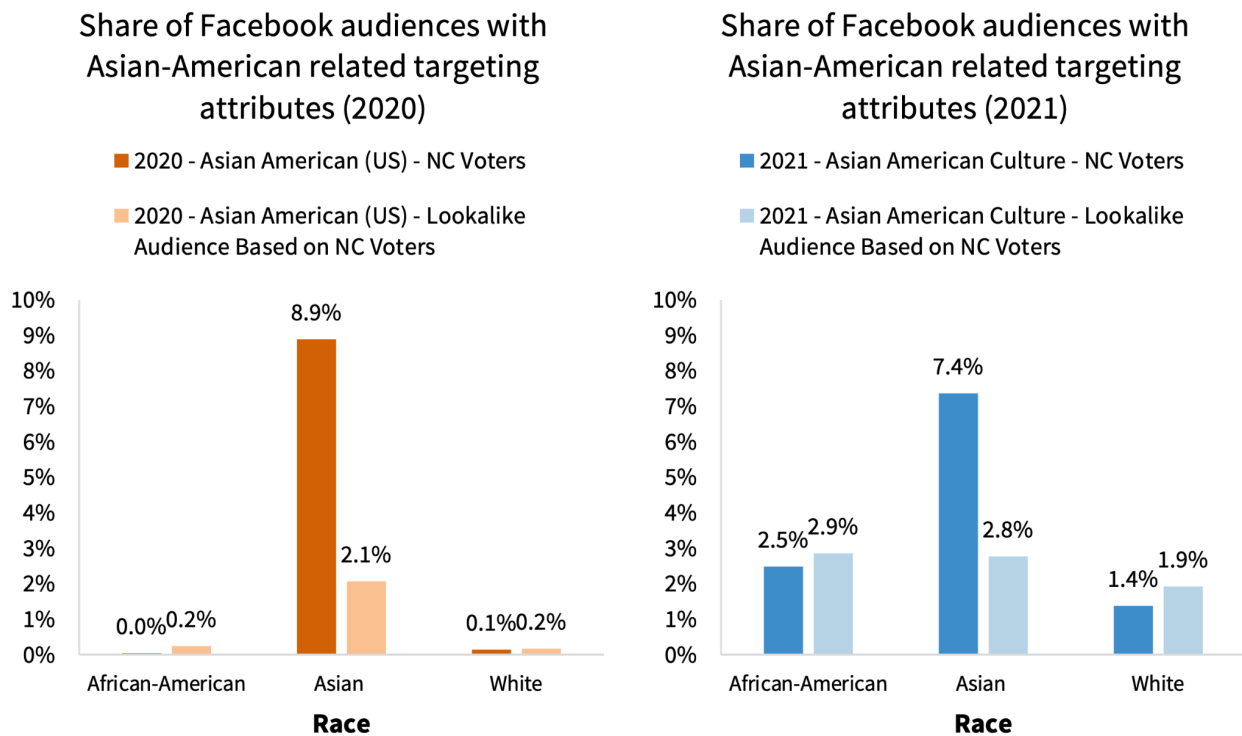
Figure 33. Share of Facebook Audiences with Asian-American Related Targeting Attributes in 2020 and 2021.

In both 2020 and 2021, Lookalike audiences based on Hispanic voters were significantly more likely to be reached by Facebook's Hispanic-American targeting attributes than Lookalike audiences based on Non-Hispanic voters. In 2020, 13% of the Lookalike audience based on Hispanic voters could be targeted with the "Hispanic (US – All)" option, compared to 2% of the Lookalike audience based on Non-Hispanic voters (Figure 34). In 2021, those rates were 7% and 2% for the "Hispanic American Culture" targeting option (Figure 34).
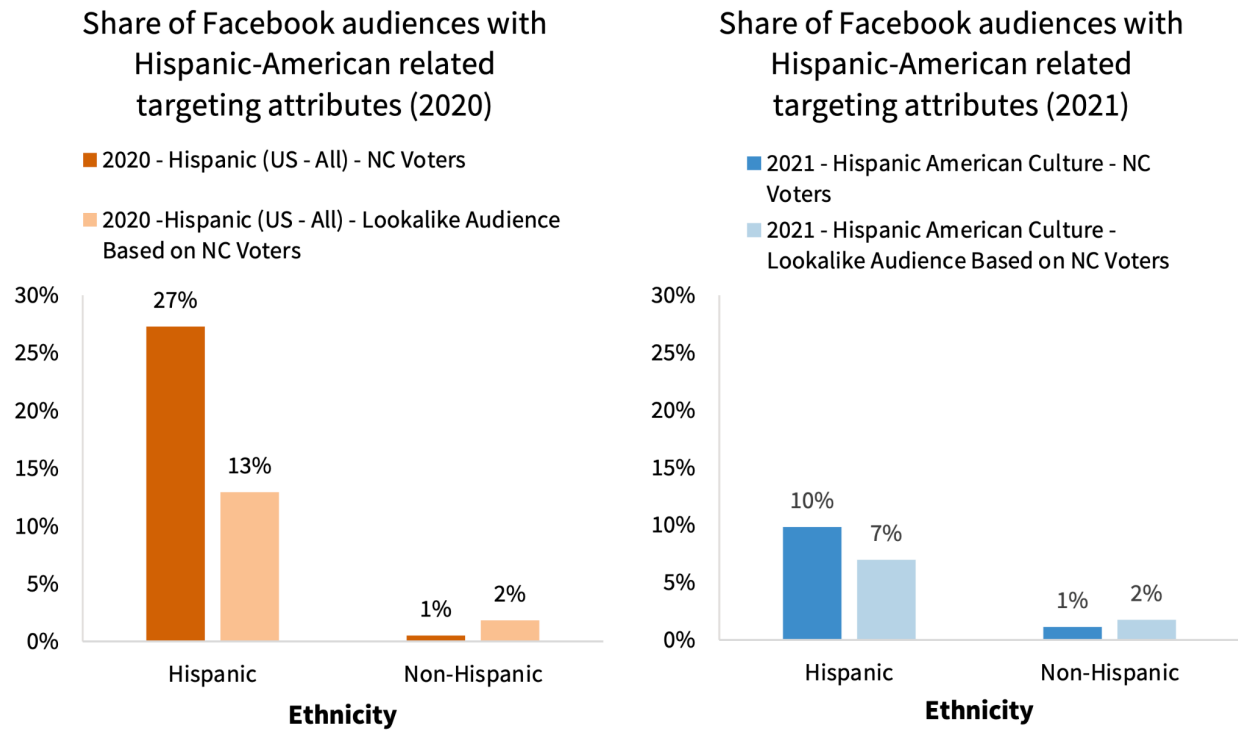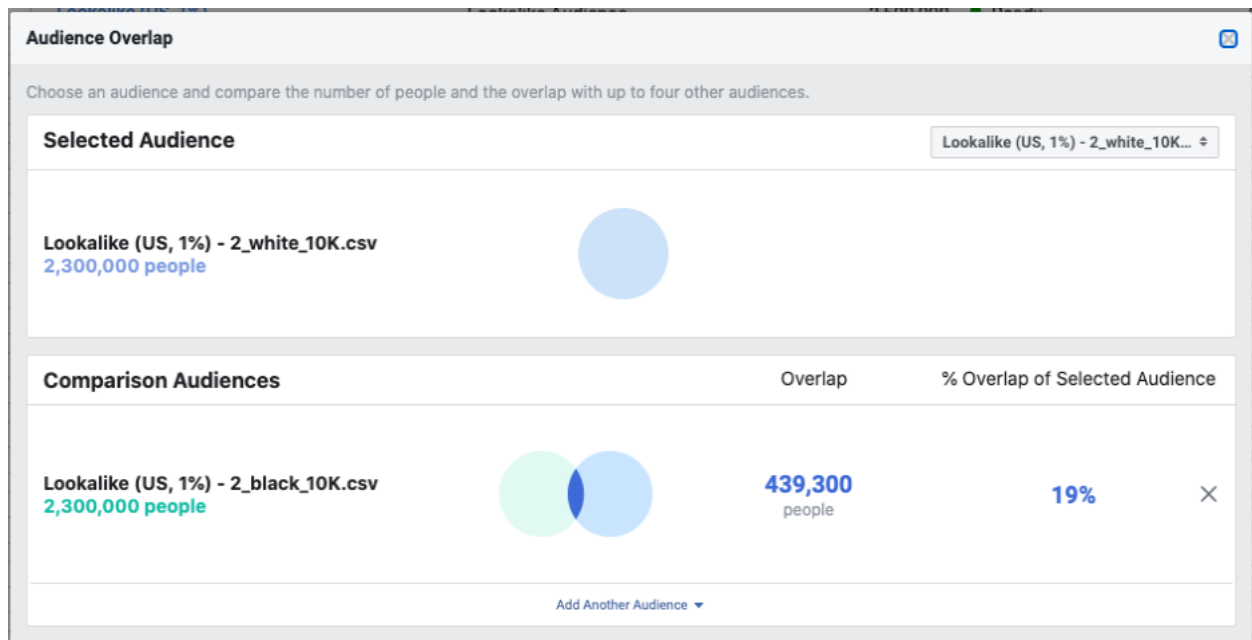
Share of Facebook audiences with Hispanic-American related targeting attributes (2020)

- 2020 - Hispanic (US - All) - NC Voters
- 2020 - Hispanic (US - All) - Lookalike Audience Based on NC Voters

Share of Facebook audiences with Hispanic-American related targeting attributes (2021)

- 2021 - Hispanic American Culture - NC Voters
- 2021 - Hispanic American Culture - Lookalike Audience Based on NC Voters

Figure 34. Share of Facebook Audiences with Hispanic-American Related Targeting Attributes in 2020 and 2021.

Appendix B – Lookalike Audiences Overlap Analysis

I used Facebook's audience overlap tool to study how many users were shared between two Lookalike audiences created using different racially or ethnically biased lists of voters. For example, Figure 35 shows that the Lookalike audience based on White voters and the Lookalike audience based on African-American voters shared 19% of the same Facebook users in 2021. Facebook didn't allow the audience overlap tool to be used for Special Ad audiences.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Figure 35. Example of Audience Overlap Between a Lookalike Audience Based on White Voters Vs. a Lookalike Audience Based on African-American Voters in 2021.

When two Lookalike audiences were created based on lists of voters of different races, generally 1/3 or less of the users were in both groups. The overlap rates decreased from 2020 to 2021. In 2020, 25-26% of the Lookalike audiences that were based on African-American versus White voters or African-American versus Asian voters overlapped (Figure 36). In 2021, those rates decreased to 18-19% (Figure 36). Lookalike audiences based on White versus Asian voters had higher overlap rates of 36% in 2020 and 29% in 2021 (Figure 36). Finally, the Lookalike audience based on Hispanic voters shared 46% of its users with the Lookalike audience based on Non-Hispanic voters in 2020 and 36% in 2021 (Figure 36).
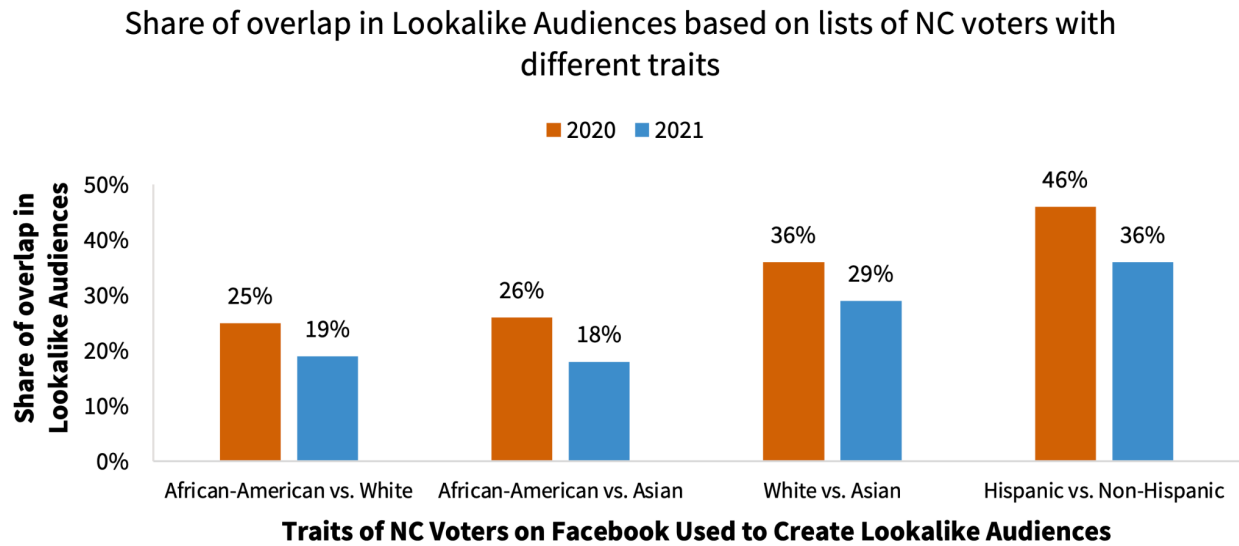
Share of overlap in Lookalike Audiences based on lists of NC voters with different traits

Figure 36. Share of Overlap in Lookalike Audiences Based on Lists of NC Voters with Different Traits.

Appendix C – Facebook Ad Library Example Screenshots and Restrictions and Notices for Special Ads Related to Housing, Employment, or Credit

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/
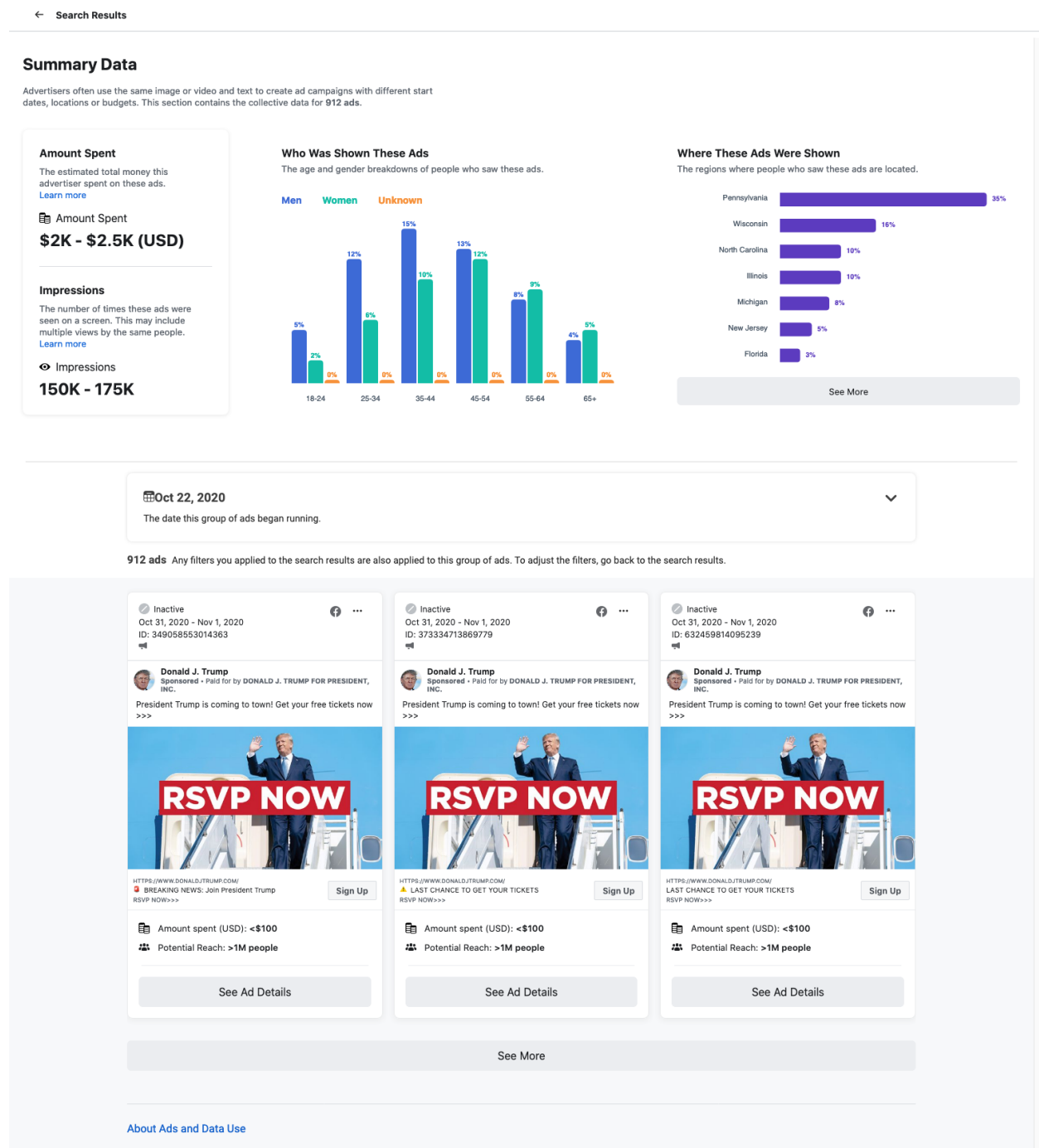
Figure 37. Example Political Ad on Facebook's Ad Library.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Figure 38. Example Housing Ad on Facebook's Ad Library.

Zang J. How Facebook's Advertising Algorithms Can Discriminate By Race and Ethnicity. Technology Science. 2021101901. October 19, 2021. https://techscience.org/a/2021101901/

Figure 39. Example Anti-Discrimination Restrictions and Notices on Facebook's Ad Planning Tool for Housing, Employment, or Credit-Related Ads.