

# “Reach the right people”: The politics of “interests” in Facebook’s classification system for ad targeting

Big Data & Society  
January–June: 1–16  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2053951721996046  
journals.sagepub.com/home/bds

Kelley Cotter<sup>1</sup> , Mel Medeiros<sup>2</sup>, Chankyung Pak<sup>3</sup> and Kjerstin Thorson<sup>2</sup>

## Abstract

Political campaigns increasingly rely on Facebook for reaching their constituents, particularly through ad targeting. Facebook’s business model is premised on a promise to connect advertisers with the “right” users: those likely to click, download, engage, purchase. The company pursues this promise (in part) by algorithmically inferring users’ interests from their data and providing advertisers with a means of targeting users by their inferred interests. In this study, we explore for whom this interest classification system works in order to build on conversations in critical data studies about the ways such systems produce knowledge about the world rooted in power structures. We critically analyze the classification system from a variety of empirical vantage points—via user data; Facebook documentation, training, and patents; and Facebook’s tools for advertisers—and through theoretical concepts from a variety of domains. In this, we focus on the ways the classification system shapes possibilities for political representation and voice, particularly for people of color, women, and LGBTQ+ people. We argue that this “big data-driven” classification system should be read as political: it articulates a stance not only on what issues are or are not important in the U.S. public sphere, but also on who is considered a significant enough public to be adequately accounted for.

## Keywords

Algorithmic classification, classification, political communication, political advertising, platforms, Big Data

Following the 2016 U.S. presidential election, Donald Trump’s digital director, Brad Parscale, stated that Facebook helped Donald Trump win the White House (Beckett, 2017). By reportedly targeting more than 50,000 ad variations to highly specific clusters of voters each day, Parscale claimed the Trump campaign was able to turn the tide of support. Illustrating this strategy, Parscale explained, “I started making ads that showed the bridge crumbling. I can find the 1,500 people in one town that care about infrastructure. Now, that might be a voter that normally votes Democrat.” Microtargeting has long been a staple of political campaigns (Brady et al., 1999; Kreiss, 2012), although, as Parscale’s claims demonstrate, the Big Data era has invited new (and contentious) mythologizing about its power (Karpf, 2018).

The allure of Facebook ad targeting lies in the popular belief that the platform’s use of sophisticated machine learning algorithms for processing data at scale opens up avenues for reaching extraordinarily granular segments of the populace. Facebook

microtargeting is said to afford precision. In this, Facebook microtargeting is driven not by a goal of making *all* users available to advertisers, but of making the “*right*” individuals available. Facebook advises that advertisers “Implement a targeting strategy that focuses on reach and precision and eliminates waste” (Facebook, n.d.-a).

The purpose of this article is to interrogate the way users are classified for sale to (political) advertisers on Facebook. We draw on theoretical concepts from a variety of domains to illustrate how this classification system articulates a stance not only on what issues are

<sup>1</sup>School of Social and Behavioral Sciences, Arizona State University, Tempe, AZ, USA

<sup>2</sup>Department of Advertising and Public Relations, Michigan State University, East Lansing, MI, USA

<sup>3</sup>United International College, Guangdong, China

## Corresponding author:

Kelley Cotter, Arizona State University, Tempe, AZ 85287-1004, USA.  
Email: [kelley.cotter@gmail.com](mailto:kelley.cotter@gmail.com)



or are not important in the U.S. public sphere, but also on *who* is considered a significant enough public to be adequately accounted for. The user classification process is a process of commodification, which involves (algorithmically) constructing “interest” categories for ad targeting and sorting users into them. As in market segmentation more generally, the production of audience segments (or interest categories) depends upon a presumption of value, as defined by market dynamics (Gandy, 2000). For online platforms like Facebook, the social, political, and economic relations embedded in user data (Noble, 2018) further shape which interest categories come to be. At the end of this process, advertisers may use the corpus of constructed interest categories as a tool for identifying and appraising audiences as more less valuable for ad targeting (Turow, 2011).

This “datastructuring” process (Flyverbom and Murray, 2018), or the ways populations are made legible to and represented in Facebook’s classification system, establishes conditions for strategic political speech on the platform—who can be reached and how. In this article, we unpack the logics of the classification system and explore its political implications. For this, we focus on ad categories broadly relevant to democratic engagement—campaign and policy politics, but also activism and identity politics. We concentrate in particular on categories relevant to the representation and visibility of people of color (POC), women, and LGBTQ+ people. We accomplish this by critically analyzing the system from a variety of empirical vantage points. We review Facebook’s documentation, training materials for marketers, and patents. We also enter the classification system via a sample of user data that includes which categories individuals’ have been assigned to, and via Facebook’s tools for marketers.

Based on these investigations and informed by past work in critical data studies (e.g., Bowker and Star, 1999; Gillespie, 2014; Noble, 2018), we suggest that Facebook’s classification system is not neutral. Facebook’s classification system constructs a particular vision of the public sphere from an assemblage of actors—algorithms, users, advertisers, and Facebook itself—based on a logic of economic value and profit. We argue that ultimately the classification system’s (mis)representation of marginalized communities reaffirms existing political power structures, which has implications for how well this system serves different groups.

### “To classify is human”<sup>1</sup>

Classifications order our worlds. A classification system constitutes “a set of boxes (metaphorical or

literal) into which things can be put to then do some kind of work—bureaucratic or knowledge production” (Bowker and Star, 1999: 10). Although they are ubiquitous and mundane, classification systems are not neutral. Their construction entails ethical and political choices, which often reproduce and reinforce power relations (Bowker and Star, 1999; Olson, 1998). This is evident in historical struggles over classifications—for example, efforts to remove homosexuality from the Diagnostic and Statistical Manual (DSM), in which it had been classified as a mental disorder (Drescher, 2015), and campaigns to replace the Library of Congress subject heading for “Illegal aliens” with “Undocumented immigrants” (Aguilera, 2016). Such cases reveal that classifications often imply “ideal” ways of being and doing that prioritize dominant groups. Classification systems are sites of struggle for control over identity and one’s place in the world.

The rise of Big Data has prompted an increasing reliance on algorithms for classifying people (Cheney-Lippold, 2017; Mackenzie, 2015). With the capacity to process vast streams of data, classification algorithms are used to infer dimensions of identity like gender and race, as well as more specific classifiers like “terrorist” or “citizen” (Cheney-Lippold, 2017). In this process, algorithms produce “measurable types,” or *approximations* of who users are based on the extent to which their data fit ever-shifting models of identity (Cheney-Lippold, 2017). Like traditional classification systems, algorithmic classification involves ethical and political choices. Algorithmic classification mediates social, political, and economic relations, frequently benefiting dominant groups along dimensions like gender, race and ethnicity, and sexuality (Noble, 2018). Developers make decisions to conceptualize and operationalize prediction models and, in doing so, “place a particular philosophical frame on the world that renders it amenable to the work of code and algorithms” (Kitchin and Dodge, 2011: 247). Developers determine who is represented in datasets and how data is “made algorithm ready” (Gillespie, 2014). Developers also make decisions about which categories to identify, what they mean, who and what should be sorted into them, and what to do with them (Gillespie, 2014). At a higher level, platform owners’ values and interests guide the development of algorithms and post hoc intervention in their outputs and outcomes. Thus, algorithmic classifications represent “interested readings of datafied reality” (Rieder, 2017: 110).

For Facebook and other online platforms, the principal “interest” in algorithmic classification is optimizing for profit. Developing a process for algorithmically inferring users’ interests allowed Facebook to fill in gaps when users did not or could not declare their interests on the platform sufficiently to meet the

company's needs. This process eventually resolved into a series of ad "interest" categories into which users can be sorted, which represent market segments for ad targeting (Thorson et al., 2019). As in market segmentation more generally, the production of Facebook's ad interest categories depends upon market dynamics, as audiences are treated as a commodity to be bought and sold (Gandy, 2000). When audience segments are under-valued in the market, demand among advertisers for the ability to reach them will be relatively low, which decreases the likelihood that a corresponding segment will be produced (Gandy, 2000). On Facebook, these market dynamics shaping the production of audience segments are automated and enacted via the continuous extraction of behavioral data from users. Dimensions of identity gain value to the extent that they correlate with consumptive behavior: clicking, downloading, engaging, purchasing, etc. (Zuboff, 2018). This allows marketers to differentiate between "targets" and "waste," defined, respectively, as individuals likely or unlikely to make a purchase (Turow, 2011). In Facebook's classification system, the market dynamics that guide data capture (Zuboff, 2018) work alongside the deeply rooted sociocultural biases embedded in data (Noble, 2018) to shape the valuation and, so, production of audiences via ad interest categories.

### Data-driven marginality in politics

Political strategizing has been gradually organizing around data collection and analysis for decades (Kreiss, 2012). Political campaigns use data collected from various sources, like public records, private firms, and surveys, to discern details about constituents, such as sociodemographics, location, interests, voting history, and partisanship (Gorton, 2016; Kreiss, 2012). With this information, campaigns engage in "rational prospecting" wherein they focus recruitment efforts on individuals likely to say "yes" to requests for political action (e.g., donating, writing a congress member, attending a protest, etc.) (Brady et al., 1999). To do this, campaigns create predictive models of people's political preferences, ideologies, and behaviors based on data available to them (Hersh, 2015). For example, George W. Bush's 2004 presidential campaign produced microtargeting segments of Michigan voters, such as "Archie in the Bunker" and "Wageable Weak Democrats" (Gorton, 2016). As Facebook, Google, and others have accrued user data at an unprecedented scale, political campaigns increasingly engage these platforms to target desired audience segments (Kreiss and McGregor, 2019).

In a positive sense, the trend towards microtargeting in politics could offer tools for increasing the political power of marginalized groups. Microtargeting has the

potential to make visible a group's common interests (e.g., issues, ideologies, values) that render them a discrete public. Microtargeting allows politicians and campaigns to speak directly to marginalized groups and bring their interests to the fore. It also helps groups connect and organize (Turow, 2011) and provides a resource for shaping their collective identity in the public eye (Sender, 2018). Yet, in practice, microtargeting relies on "activating" targeted segments of the public, rather than encouraging broad, inclusive engagement (Barocas, 2012; Kim et al., 2018; Schier, 2000). Consequently, new concerns about "political redlining" have emerged. As Kreiss (2012) explained, "Campaigns routinely 'redline' the electorate, ignoring individuals they model as unlikely to vote, such as unregistered, uneducated, and poor voters" (n.p.) Indeed, past work has documented gaps in rates of appeals to POC and women, which can exacerbate disparities in access, turnout, and representation (Kim 2016; Verba et al., 1995). In particular, political campaigns tend to avoid including POC in their messages and addressing issues of importance to them when messages are more likely to reach white audiences (Nteta and Schaffner, 2013). Likewise, in electoral campaigns, issues of importance to POC are frequently ignored or downplayed in favor of appeals to white voters deemed more likely to be swing voters (Frymer, 2010). Thus, microtargeting tools have the potential to create new ways of *neglecting* the political interests of marginalized communities. Microtargeting tools also may be used to suppress voting among minoritized communities, as in the 2016 U.S. election (Pybus, 2019).

Facebook has in fact received criticism for providing targeting options that could result in discrimination based on race, gender, ethnicity, and other legally protected characteristics (Tobin, 2019). Yet, importantly, political campaigns need not explicitly exclude marginalized groups for targeted advertising to result in discrimination. Automated ad delivery systems can "profile" users based on their data such that the ads they see reproduce offline discrimination. For example, in one study, men received more Google ads for coaching services for high paying jobs than women (Datta et al., 2015). In another, Google searches for names more commonly assigned to Black babies were more likely to display ads for public arrest records than searches for names more commonly assigned to white babies (Sweeney, 2013). In the context of Facebook advertising, Speicher et al. (2018) revealed multiple ways that targeting options have the potential to facilitate discriminatory outcomes, particularly in ways that evade detection. Together, this body of work suggests that microtargeting often reenacts the extant sociopolitical order, which has important consequences for the representation and visibility of marginalized

communities in political campaigns employing Facebook advertising.

In what follows, we unpack the logics of Facebook's classification system through a synthesis of literature on Facebook ad targeting and catalog a series of cases that lay bare the political negotiations shaping the representation of POC, women, and the LGBTQ+ people in this system. We argue that Facebook's classification system constructs a particular vision of the U.S. public sphere that reaffirms political power structures as a result of the system's technical and commercial logics.

## Method and approach

Studying algorithmic systems poses many practical and epistemological challenges (Kitchin, 2017). Algorithms are typically kept secret (Pasquale, 2015), can be extraordinarily complex (Burrell, 2016) and are constantly evolving (Kitchin, 2017), particularly as platform affordances continually change (Barrett and Kreiss, 2019). Heeding Kitchin's advice on studying algorithms in light of these challenges, in this study, we focus on the ways that algorithms "do work in and make the world" (2017: 18). We entered Facebook's interest classification system from different angles, taking on the perspective of users and then of advertisers, and examined its outputs. Facebook's classification system makes categories available to marketers via the advertising interface and to users via a downloadable data archive. For this study, we drew on data from a larger project, which included a survey in which participants were asked to download and share files from their Facebook data archive that listed ad categories they were sorted into and Facebook pages they "liked." The sample targeted young adults aged 18–35 in the U.S. The total sample consisted of 1312 participants, of which 404 provided Facebook data. Participants' advertising categories and "liked" pages were matched against a dictionary of more than 160,000 political terms gleaned from external lists of U.S. political figures, journalists, news outlets, activist and advocacy organizations, and by consulting Facebook's Audience Insights tool. The resulting dataset consisted of a list of political ad categories and liked pages.<sup>2</sup> Our use of these data is not intended to generalize our claims about the nature of Facebook interest classification; rather, as describe below, we use these data to identify critical cases in line with our focus on POC, women, and LGBTQ+ communities.

We also used the tools Facebook provides to marketers, namely the Ads Manager, Audience Insights, and Marketing API, to explore the classification system. Ads Manager is the interface used by marketers to assemble ad campaigns. The ad targeting categories

appear in the Ads Manager within a loosely organized classification scheme featuring nine broad, top-level "topics" such as "Business and industry" and "Hobbies and activities" (see Figure 1).<sup>3</sup>

The Audience Insights tool allows marketers to view rough estimates of how many individuals have been sorted into ad targeting categories, as well as other information about audiences, like location, demographics, Facebook pages users have liked, and past purchase behavior (see Figure 2). Facebook's Marketing API similarly allows third-parties to search the classification system and view all categories associated with a keyword, the exact number of individuals sorted into a category, and the higher level topics and "paths" into which categories fall (see Figure 3).

We explored Facebook's classification system using an iterative process, drawing on the tools for advertisers and participants' Facebook data as points of access. We identified categories relevant to our communities of interest in participants' data (e.g., "Racial equality," "Gay rights") and explored these via Facebook's advertiser tools. For example, we identified the category "Social justice" as an interest associated with some participants who donated their Facebook data to us, and then used the Ad Manager to seek related categories. We also used participants' data to inform searches for additional relevant categories. For example, we identified a category for "Feminism" in participant data and brainstormed contrastive and complementary terms, organizations, and figures (e.g., "men's rights," NOW (National Organization of Women)). While brainstorming counterparts to the entry category, we concurrently searched the Ads Manager, Audience Insights, and Marketing API to gain a better sense of how issues were represented in the classification system. In many cases, the categories we found via Facebook's advertiser tools led to additional related categories suggested by the tools. For example, Audience Insights suggested the category "Opposition to immigration"<sup>4</sup> when we searched for the category "Immigration." Finally, we also searched Facebook's tools for various keywords relevant to our communities of interests (e.g., "Latino," "Queer," "Women").

From these explorations, we critically analyzed how the system constructs a particular understanding of U.S. politics vis-a-vis the representation of POC, women, and LGBTQ+ people. We read the classification system with three assumptions in mind, as brought forth by Hope A. Olson:

First, classification, like any map, is constructed by dominant cultural discourses. Second, classification, like any system, has constructed boundaries or limits

**Audience**  
Define who you want to see your ads. [Learn More](#)

**Create New Audience**    Use Saved Audience ▼

**Custom Audiences** ⓘ    Add a previously created Custom or Lookalike Audience

[Exclude](#) | [Create New](#) ▼

---

**Locations** ⓘ    **Location:**  
• United States

**Age** ⓘ    18 - 65+

**Gender** ⓘ    All genders

**Detailed Targeting** ⓘ    Include people who match ⓘ

Add demographics, interests or behaviors    [Suggestions](#) | [Browse](#)

▼ **Interests** ⓘ

- ▶ Business and industry ☐
- ▶ Entertainment ☐
- ▶ Family and relationships ☐
- ▶ Fitness and wellness ☐
- ▶ Food and drink ☐
- ▶ Hobbies and activities ☐
- ▶ Shopping and fashion ☐
- ▶ Sports and outdoors ☐

**Placements**  
Show your ads to the right people

**Figure 1.** Facebook ads manager.

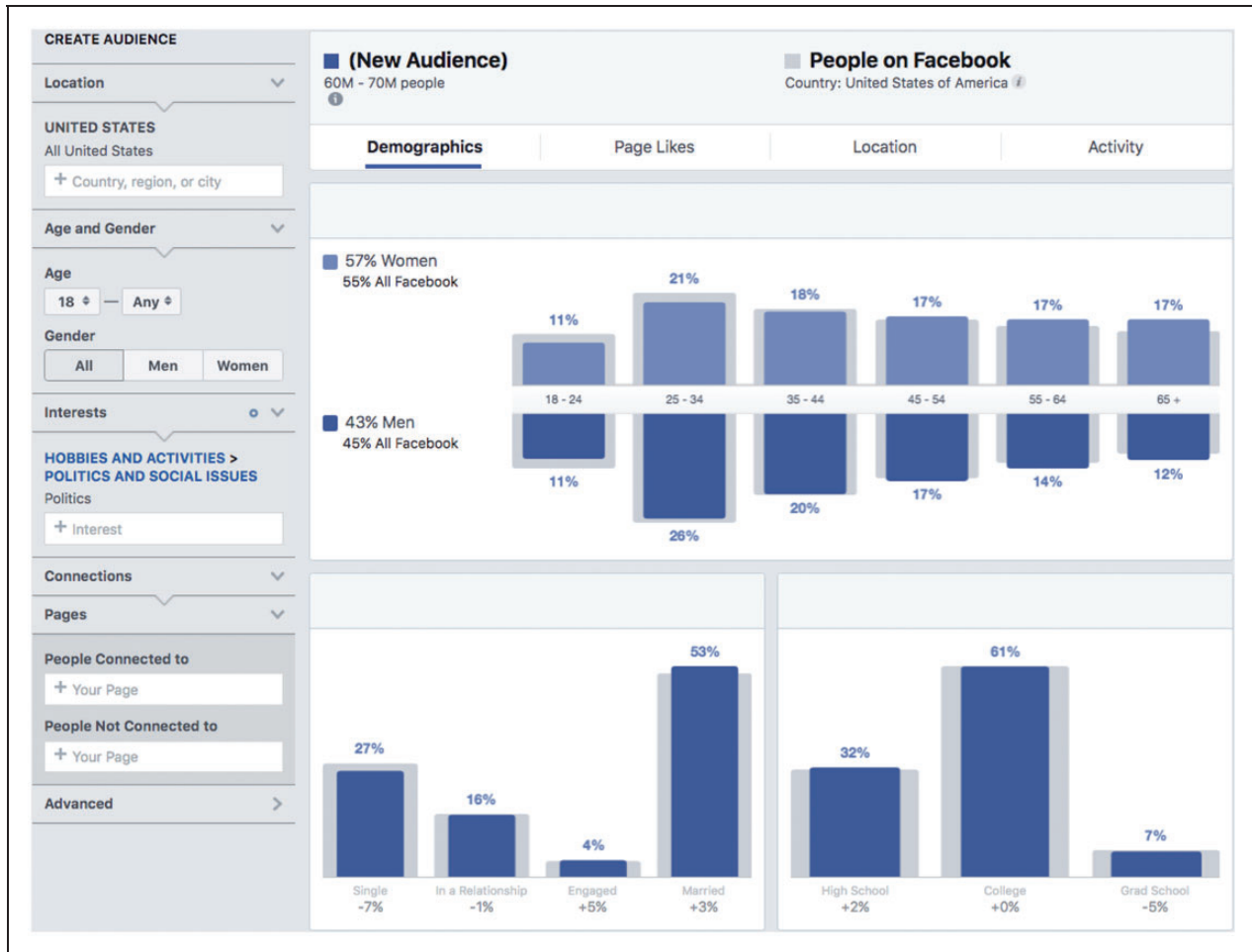
that result in exclusions. Third, the construction of classification is a form of location that defines and sequences what is accepted as knowledge, thus marginalizing as well as excluding. (1998: 252)

Reading the categories in this way, as Olson elaborates, “make[s] what was transparent and invisible opaque and visible, elucidating the biases and the discourses that construct and enforce them” (1998: 252). Thus, we deconstructed elements of language, technical infrastructure, and networks to interpret the political topology of the classification system. Our aim was to problematize this system through a series of example cases. This approach prioritizes thick description of different issues, but also opens up opportunities for investigating their prevalence and the scale of impact.

### How Facebook’s classification system works

It is clear from the small body of work exploring how Facebook’s interest classification works—and from our

own investigations—that the classification system emerges from a nexus of choices by users, advertisers, and Facebook itself. It is also subject to “platform transience,” or “continual and rapid change” over time (Barrett and Kreiss, 2019: 1). The classification process begins with the collection of data about users captured from their profiles and activity on the platform and beyond (Andreou et al., 2018). These data inputs then inform the production of hundreds of thousands of interest categories (Andreou et al., 2018; Speicher et al., 2018). In our investigations, we found that categories can be either *translated*, when a liked Facebook page becomes an ad category, or *imputed*, when an ad interest category is assigned from user behavioral data.<sup>5</sup> Translated interests represent a user-directed, explicit expression of users’ interests: a user “likes” the page for NPR; Facebook translates this user-expressed interest into an algorithmically assigned one.<sup>6</sup> Imputed interests are derived from digital-trace signals of interests beyond liked Facebook pages, including information users include in their profiles, keyword analyses of interactions and



**Figure 2.** Facebook audience insights tool.

posts, and user behavior on and off the platform (Andreou et al., 2018; Rajaram et al., 2014; Yan et al., 2014; Zhou and Moreels, 2014). Using the data Facebook collects and has access to, the platform attempts to “uncover naturally occurring behaviors” (Facebook, 2015a: n.p.) from which they extract keywords and phrases (e.g., “Politics,” “Democracy,” and “Civil rights.”) that become imputed interests. As categories are translated and imputed, users’ choices about what to “like,” read, or talk about informs how they are sorted into categories. Users’ choices in constructing their social networks—whom they connect with, interact with, etc.—and content selection and sharing choices made by those in their social network also shape how they are sorted into categories (Thorson et al., 2019; Yan et al., 2014). Although user choices inform the classification system, the extent to which categories accurately and reliably represent user interests is unclear. Recently revealed internal communications at Facebook call into question the

fidelity of interest categories, with one memo stating that

“‘[I]nterest precision in the US is only 41%—that means that more than half the time we’re showing ads to someone other than the advertisers’ intended audience. And it is even worse internationally . . . We don’t feel we’re meeting advertisers’ interest accuracy expectations today.’” (Biddle, 2020, n.p.)

Besides users, advertisers also play an important role in the classification process. “Bigger” advertisers have access to the Facebook ads team through formal partnerships and support infrastructure, as well as through affordances available to all who advertise on the platform. For example, Facebook worked closely with Trump’s 2016 presidential campaign (and other campaigns) to help his team use Facebook and its advertising tools more effectively (Kreiss and McGregor, 2018).<sup>7</sup> Such support may offer opportunities for

```

{
  "data": [
    {
      "id": "6011515350975",
      "name": "Politics and social issues",
      "audience_size": 1153206700,
      "path": [
        "Interests",
        "Hobbies and activities",
        "Politics and social issues"
      ],
      "description": ""
    },
    {
      "id": "6003654559478",
      "name": "Politics",
      "audience_size": 474560010,
      "path": [
        "Interests",
        "Hobbies and activities",
        "Politics and social issues",
        "Politics"
      ],
      "description": "",
      "topic": "News and entertainment"
    },
    {
      "id": "6003230358588",
      "name": "Political party",
      "audience_size": 178794330,
      "path": [
        "Interests",
        "Additional Interests",
        "Political party"
      ],
      "description": null,
      "topic": "Business and industry"
    },
    {
      "id": "6003209938872",
      "name": "Power (social and political)",
      "audience_size": 140505720,
      "path": [
        "Interests",
        "Additional Interests",
        "Power (social and political)"
      ],
      "description": null,
      "topic": "Business and industry"
    },
    {
      "id": "6003901992266",
      "name": "Political freedom",
      "audience_size": 62143010,
      "path": [

```

**Figure 3.** Marketing API (partial) output for ad interest categories.

advertisers to provide input on existing targeting options and/or request new categories. In practice, Facebook's ad audience optimization algorithms may also nudge advertisers towards interest categories that produce the largest returns in terms of reach or conversions (see Ali et al., 2019). In turn, advertisers' reliance on certain categories may guide which additional categories Facebook creates and imputes, and which it removes (Schiff, 2019). The broader marketing industry has also played a key role in Facebook's classification system. Previously, Facebook partnerships with

marketing firms like Acxiom and Experian allowed for the introduction of targeting categories from third-party data (Andreou et al., 2018; Statt, 2018).<sup>8</sup>

Facebook's policies, values, and goals shape the classification system by informing the design of the system, but also by guiding the removal of categories. Facebook has stated that it routinely removes rarely used categories in order to eliminate "bloat" and enhance usability (Schiff, 2019). Previous reporting by ProPublica that unearthed anti-Semitic categories in the system (e.g., "Jew haters") led to Facebook removing a series of categories (Angwin et al., 2017). Later, facing a lawsuit over charges that the company permitted illegal discrimination in advertising, Facebook removed more than 5,000 ad categories from the system (e.g., "Passover," "Native American culture," "Evangelism"; Schiff, 2018). In explaining its decision, the company wrote "While these options have been used in legitimate ways to reach people interested in a certain product or service, we think minimizing the risk of abuse is more important" (Facebook, 2018: n.p.). Here, as Barrett and Kreiss (2019) noted, Facebook seems to be acting in response to external pressure, but also a "desir[e] to be in line with social values, expectations, and ideals" (15).

Although Facebook's past actions indicate that it can and does intervene in the classification system, the company's procedures and policies regarding when, how often, at what scale, and why it moderates categories are not clear. We do know Facebook allows users to report categories deemed "inappropriate" in the Ad Manager tool. As is the case with reporting content on the platform (Gillespie, 2018), it is likely that Facebook also uses reporting to inform its moderation of ad categories. Beyond this, although Facebook has cited underuse and potential for abuse as explanations shaping many of its decisions, these broad rationales do not logically pair with many of the interests we discovered were unavailable for targeting—for example, high-profile politicians (e.g., Ilhan Omar), social movements (e.g., #MeToo), and ideologies (e.g., Socialism). Given that these unavailable interests are prominent within U.S. political discourse and media coverage, it is unlikely that they are absent due to inappropriateness, lack of demand, or insufficient data signaling interest.

### Representational politics in Facebook's classification system

Having outlined the operational logics of the classification system, we now highlight exemplars of the political nature of this system. In particular, we focus on how the system represents marginalized groups,

focusing on POC, women, and the LGBTQ+ community. In doing so, we discuss implications for a level playing field for political participation on the platform.

### The coded gaze

Through our investigations, we encountered categories that demonstrated that Facebook's representation of communities does not always match how a community sees itself. This is best exemplified through two connected categories: "Transgenderism" and "Passing (gender)." We identified these categories after observing that, while the user dataset featured many categories referencing the LGBTQ+ community, only one included the word "transgender" ("Gay, Lesbian, Bisexual, Transgender, Straight Alliance"). In searching Audience Insights for "transgender," we noticed the category for "Transgenderism." A search via the Marketing API for "transgender" returned the category for "Passing (gender)." "Transgenderism" is considered an offensive term not commonly used by transgender people (GLAAD, 2011). As GLAAD explains, "transgenderism" "is a term used by anti-transgender activists to dehumanize transgender people and reduce who they are to 'a condition'" (2011: n.p.). Yet, using Facebook's Audience Insights tool, we can see that those with an imputed interest in "Transgenderism" may well be members of or allies to the LGBTQ+ community, based on the top pages associated with this category (e.g., LGBTQ Nation, No H8 Campaign, Lizzy the Lezzy).

Similar to "Transgenderism," "Passing (gender)" views transgender people through a cisnormative lens. While Facebook does not provide information about which pages users assigned to the category "Passing (gender)" have liked due to the relatively small size of the audience, that the category emerged from searching the Marketing API for "transgender" suggests a connection between the categories and, potentially, users assigned to them. "Passing" in relation to gender identity refers to transgender people's ability to move about the world without being misgendered. Describing someone as "passing" assumes that they are in some way misrepresenting themselves. By including this category, Facebook's classification system centers cisgender normativity and marks being transgender as aberrant and something to hide. While the category acknowledges some fluidity in gender expression, its existence prioritizes passing—gender presentation matching socially constructed definitions of gender—because it is discoverable alongside other categories returned by a keyword search of the Marketing API for "transgender."

These two categories exemplify that, though marginalized groups may be "seen" by Facebook's

classification system, they are sometimes seen through what Joy Buolamwini terms the *coded gaze*—the "embedded views that are propagated by those who have the power to code systems" (2016: n.p.). Under the coded gaze, members of dominant groups are made more recognizable to algorithms—and recognized with a greater degree of accuracy—than those from marginalized groups. In Facebook's classification system, the coded gaze manifests in categories like "Transgenderism" and "Passing (gender)" that seem to represent nondominant groups via an algorithm trained on data privileging dominant perspectives. The coded gaze hearkens to the notion of the *gaze* in critical theory, which recognizes that we understand ourselves in relation to how others see us, which we can never fully control. In media studies, scholars draw on the concept of the gaze to acknowledge how media tend to be created for an idealized viewer from whom certain responses are intended to be provoked (Sturken and Cartwright, 2001). Marginalized groups seeing themselves "through the implied gaze of others" (Sturken and Cartwright, 2001: 81), thus, experience the symbolic violence of being "othered." With "Transgenderism" and "Passing (gender)," the coded gaze similarly characterizes transgender people within a framework of subordination and conformity. The transgender community on Facebook is recognized, but naturalized as outside of the norm.

### The unmarked user

In examining ad categories in the user data, we observed multiple categories including the word "women," but no categories including the word "men." Keyword searches via the Marketing API replicated this pattern: Facebook's classification system features 474 categories containing "women," "woman," "girl" (e.g., "Women and video games," "Women in business," "Women in science"), but 177 containing "men," "man," or "boy" (e.g., "Men's Style & Grooming," "Men Fashion," "Men's accessories")—the vast majority of which refer to sports teams (e.g., "Washington Huskies men's basketball") or TV or film titles (e.g., "Spider-Man," "Iron Man").<sup>9</sup> POC are represented in a similarly (relatively) abundant—and explicitly encoded—fashion through a series of categories associated with minoritized racial and ethnic groups, for example "African American vernacular English," "African American Expressions," "Asian American culture," "Latino culture," "Hispanic culture," "Chinese American culture," and "Native American culture in the United States." Analogous categories for white ethnic groups do not exist in the classification system, as far as we can tell. For example, there are no "Irish American culture," "Polish

American culture,” or “Italian American culture” categories, although keyword searches for “Irish American,” “Polish American,” and “Italian American” Facebook pages return many results. The classification system also includes categories for users’ “multicultural affinity” that further evoke questions around representation. According to Facebook, these categories refer to “peoples’ affinity to cultures they’ve demonstrated an interest in through their behaviors on Facebook” (2019: n.p.). “White American” (or “European American”) is *not* one of the six targetable “affinities”: “African American (US),” “Asian American (US),” “Hispanic (US - All),” “Hispanic (US - Bilingual),” “Hispanic (US - English dominant),” and “Hispanic (US - Spanish dominant).”

While the above-described categories render certain groups visible in the classification system, their relative prevalence simultaneously traces the shape of a central unmarked user representative of “normal” or “average” in the system (Bowker and Star, 1999). Bowker and Star explain the notion of an unmarked user in relation to the international classification of diseases (ICD), in which the adult male body is the unmarked category. As they explain, the disproportionate number of diseases restricted to women only (“there are sixteen categories or clusters of categories that apply only to males and forty-two that apply only to females” (Bowker and Star, 1999: 90)) amounts to the “the relative pathologizing of the female body” (Bowker and Star, 1999: 90). Because “adult male” is the norm, conditions considered normal for adult males do not require categories. As such, when individuals are classified, they may become “marked” as subordinate to the idealized unmarked user, which typically represents the intersection of powerful social positions—white, Western, man, cisgender, heterosexual, able-bodied, and so on. Importantly, the unmarked user is implicated not simply by the absence of categories “for” dominant groups, but rather *asymmetries* in representation. For example, the classification system has interest categories for major party candidates in the 2020 U.S. presidential election, but no ad categories for third-party candidates. In this case, the selective inclusion of candidates reflects a default user only interested in major parties.

The rendering of an unmarked user through a relative excess of categories referring to women and POC offers an example of what Tressie McMillan Cottom calls “predatory inclusion.” Predatory inclusion occurs when inclusion interventions appear to create new, favorable opportunities for marginalized people, but do so on extractive and often harmful terms—namely, by generating value from (and not necessarily for) them. Making women and POC disproportionately prominent in Facebook’s classification facilitates the

extraction of value from them as part of a longer history of market segmentation in which capitalism has intertwined with the logics of white male supremacy. The growing association of women with consumerism during the early 20th century positioned women—namely, white middle-class women—as a valuable market to exploit (Peiss, 1998). Later, African American and Hispanic consumers gradually came to be seen as profitable markets for reasons including geographic distribution, which created efficiencies in ad targeting via legacy media, and market research on value and population size and growth (Gandy, 2000). Facebook seems to inherit this history. For example, of the multicultural affinity categories, Facebook writes in a training module:

Smart marketers recognize the business imperative behind connecting with people who have diverse interests based around cultural beliefs, traditions, music, aesthetics, or languages. Marketing to these groups is becoming more important as multicultural consumers grow in number, influence, and spending power. (Facebook, 2019: n.p.)

Here, Facebook frames the inclusion of communities of color in the classification system on extractive terms. The presence of the unmarked user in Facebook’s classification system, borrowing Cottom’s words, “perverts” the value of inclusion by offering a vehicle for extractive schemes and by positioning POC—and, to a lesser extent, women—as outsiders with “diverse” interests distinct from “normal” interests.

### Flattening interests

The classification system alternates between capturing highly specific “interests” and identifying broad categories that elide users’ differences. We refer to the latter as a “flattening” of interests. Flattening occurs when issues and identities are collapsed into a broad higher-level category, or when multiple intersecting dimensions of issues or identities are reduced to a single dimension. Flattening is evident in the previously described categories for “multicultural affinity.” With multicultural affinity, only three broad racial and ethnic categories exist: African American, Hispanic American (broken down by language[s] spoken), and Asian American. Each of these three categories could be subdivided into much longer lists of ethnicities that offer more meaningful designations (e.g., Haitian, Dominican, Mexican, Puerto Rican, Chinese, Indian), but they remain as monoliths, likely because the broad categories capture larger markets. Facebook communicates this flattening of ethnicities in its advice to advertisers—for example, suggesting that advertisers

“Integrate content that reflects *the Asian American experience*. From celebrating holidays like Lunar New Year to viewing movies produced by Bollywood in the US, Asian Americans stay connected to their culture and heritage” (Facebook, 2015b; emphasis added).

When it comes to political issues, we see that most categories also do not subdivide into more specific interests. For example, although in participants’ data and via searching Facebook’s Marketing API we can see categories for “Racial equality,” “Transgender activism,” and “Women’s rights,” there is no category for “Black transgender women’s rights.” On the one hand, it may seem unrealistic to expect such a specific category. On the other hand, the existence of categories like “African American vernacular English” and “Gender-specific and gender-neutral pronouns” demonstrates that a high-level of granularity is a feature of the classification system.

The flattening and inconsistency in the granularity of “interests” in Facebook’s classification system may result from the long tail problem (Park and Tuzhilin, 2008). Political interests in the “long tail,” or issues of interest to a minority of the population, by definition, emanate in trace data from a relatively smaller portion of the Facebook population. Data for such issues from which categories could be inferred is limited and noisy, and, thus, less legible to algorithms. Further, when too few people align with categories, Facebook may not be motivated to devote resources to ensuring the creation of these categories, as they do not guarantee a substantial return on investment. Facebook advises marketers against targeting too narrowly (e.g. 1300 users; Facebook, 2019) and suggests that targeting broader audiences tends to produce a greater “impact” (see Figure 4; Facebook, 2016).

The above discussion illustrates the classification system’s manifestation of a “single-axis framework” (Crenshaw, 1989), in which the multidimensionality of users’ experiences in the world is not always captured. This is similar to how traditional classification systems reduce contingency and local meaning in order to be applicable across multiple contexts, and, consequently, elide “the local and specific with the general” (Bowker and Star, 1999: 80). Such flattening impacts the degree to which minoritized interests can be adequately addressed on the platform.

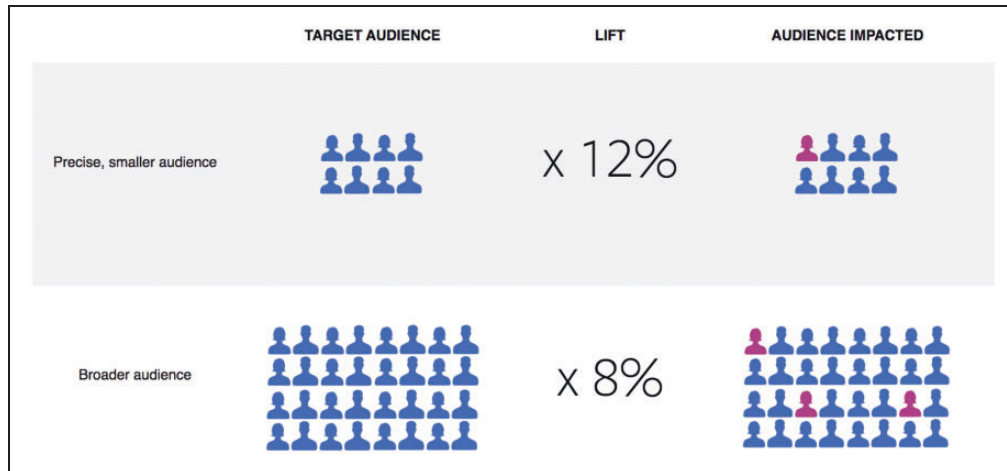
The concept of intersectionality recognizes that individuals’ lives are shaped by the ways they are oppressed and/or privileged along *multiple* socially constructed identities (Crenshaw, 1989). In coining the term “intersectionality,” Kimberlé Crenshaw exemplified the ways that Black women are “multiply-burdened” by discrimination on the basis of race and gender, but that discrimination is identified strictly through

“pure claims” of either racism or sexism. As Crenshaw argued, this “either/or” assessment leaves Black women unprotected when their experiences are distinct from Black men and white women.

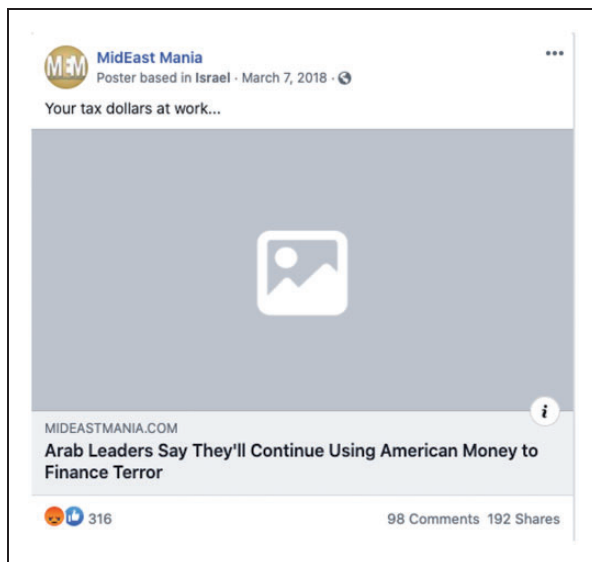
Although Facebook advertisers can attempt to reach individuals interested in an intersectional issue like “Black transgender women’s rights” by combining discrete categories, this is not the same as a single category that captures the issue directly. Intersectionality suggests that oppression along multiple axes of identity cannot be discerned from merely adding together the experiences of separate identities (Collins, 2014). Similarly, targeting individuals assigned to multiple separate categories does not necessarily produce the same audience as targeting individuals assigned to one more specific category. The two audiences may have different interests. When intersectional identities and issues are decomposed into separate components, the classification system overlooks crucial information and fails to adequately represent minoritized communities.

### Both sides-ism

As a result of underlying technical and commercial logics, Facebook’s classification system is, to some extent, agnostic to ideology, treating “both” sides of an issue as equally valid. In this way, it exhibits what Whitney Phillips (2018) calls “both sides-ism.” Phillips uses this concept in the context of journalism to argue that giving equal weight to both sides of an issue often grants credence to and, therefore, “oxygenates” hateful and/or dehumanizing ideologies. Similarly, at times, Facebook’s classification system quietly normalizes harmful ideologies encoded in categories. As an example, we encountered categories for “Immigration to the United States” as well as “Stop Illegal Immigration.” Because not all political perspectives appear in the classification system,<sup>10</sup> the presence of a category for “Stop Illegal Immigration” suggests—to quote Noble (2018)—“a corporate logic of either willful neglect or a profit imperative that makes money from” (p. 5) the ideology that underlies it. While debate on the merits of immigration programs and policies is part and parcel of mainstream political discourse—and, indeed, the classification system includes a category for “Immigration reform”—“Stop Illegal Immigration” seems to represent a more extreme perspective that conflicts with social values of granting respect and dignity to all people. We can see evidence for this in the list of public pages that are associated with this interest category (based on Audience Insights at the time of writing). To offer two examples: pages for the Center for Immigration studies, an organization with a history of associating with white nationalists (SPLC, n.d.-a),



**Figure 4.** Facebook's illustration of the impact of focusing on broader audiences.



**Figure 5.** Facebook post from the page MidEast Mania.

and for MidEast Mania, an Israel-based page that regularly posts Islamophobic content (e.g., see Figure 5), both appear in the list of top pages associated with the category.

Both-sides-ism more commonly occurs implicitly as categories that appear neutral in name encode political ideology by nature of the audiences sorted into them. For example, many of the top pages associated with the category “LGBT Rights in the United States” have taken a stance *against* LGBTQ+ rights (e.g., Liberty Counsel, National Right to Life). Likewise, many of the top pages associated with a category for “Immigration to the United States” express extreme right-wing positions and anti-immigrant sentiment (e.g., Defense of Freedom, The Committee to Defend



**Figure 6.** Facebook post from the page Defense of Freedom.

the President; see Figure 6), including one for the Federation for American Immigration Reform, which reportedly has ties to white supremacist groups (SPLC, n.d.-b).

The encoding of ideology into ad categories is a byproduct of algorithmic inference of users' interests: the classification system merely captures correlations between groups of users and patterns of speech and behavior. Categories inferred in this way encode more than topics. They also encode more abstract information about historically rooted cultural biases and preferences in language (Caliskan et al., 2017). Moreover, the partisan lean and political significance of categories will always be dynamic, changing over time as the model adapts to new data. “Immigration to the United States” might represent users against immigration today, but could represent those for immigration in the future.

The classification system's both sides-ism rests on a cost-benefit calculus, wherein the impetus to provide more targeting options for advertisers may conflict with preventing “toxic technocultures” (Massanari, 2017). A category like “LGBT Rights in the United

States” could represent either users for or against LGBTQ+ rights, and which stance it does represent is not as important as merely inferring “interest” in the issue. Thus, the classification system embeds a cost-effective political apolitical-ness on Facebook’s part that tracks with the company’s desire to remain politically neutral (Gillespie, 2010). Yet, in its both sides-ism, the classification system not only grants legitimacy to harmful perspectives, but provides audiences for amplifying them. This may help sustain a broader online culture that is hostile to POC, women, and the LGBTQ+ community and, so, stymies the possibility of fostering a welcoming space for them (Lenhart et al., 2016).

## Discussion

In October 2020, during the course of revising this article, Facebook announced it would temporarily halt political advertising on the platform after the 2020 U.S. presidential election (Rosen, 2020). This policy change seemed to be influenced by concerns about false information, privacy, manipulation, and election interference (Ghosh and Scott, 2018; Gorton, 2016). Our study introduces an additional concern. We showed that the classifying and labeling of user “interests” should be read as political: Facebook’s classification system articulates “interested readings” of user data (Rieder, 2017) in service of the company’s commercial interests. In particular, we showed how such readings produce political outcomes—namely, a coded gaze, an unmarked user, a flattening of interests, and both sides-ism. These outcomes have consequences for the political visibility and power of many groups, as the system seems to prioritize the interests of the socially and economically powerful. As a result of human choices embedded in datafication processes, the system represents those who have been historically marginalized not on their own terms, but on the terms of those occupying more privileged positions. This is the byproduct of a profit-oriented analysis of “naturally occurring behaviors” (Facebook, 2015a: n. p.) that reflect hierarchical social order. In what follows, we first recap our findings, and then explain their significance and implications for the future of political advertising on Facebook.

Multiple actors shape Facebook’s classification system. Users themselves explicitly report interests by “liking” or “following” pages; these explicit interests may be (but are not always) made available to advertisers for targeting. On the other hand, many interests are imputed: the result of algorithmic inferences based on traces of users’ own data as well as the data representing their social connections. Beyond users, advertisers themselves play a role in shaping the

classification of users, both explicitly through partnerships with Facebook and implicitly based on a shared understanding of what kinds of audiences are valuable. Although the classification system emerges in large part from choices made by users, advertisers, and algorithms, Facebook alone reserves the right to directly modify the classification system, including via moderation efforts. As a fundamental component of the platform’s business model, intervening in the classification system is not only a political decision, but a financial one. With its access to expansive streams of data, Facebook has been able to train algorithms that help the company gain a competitive edge, which has allowed the company to gain market power and accumulate human capital and political influence (Srnicke, 2016). Yet, the company’s commercial interests increasingly come into conflict with its stated commitment to diversity and inclusion (e.g., Facebook, n.d.-b) and to protecting democracy (e.g., Facebook, 2020). When money drives decisions, the result is inevitably the prioritization of those deemed more valuable, as we further explain below. Further, actively intervening in the system via moderation is complex and expensive (Gillespie, 2018). The persistence of categories in the classification system that fail to respectfully represent all users or those that encode hateful ideologies suggests that Facebook has not yet adequately assessed the scale of resources needed for this particular domain of moderation, or developed best practices for when and how to intervene.

Left unchecked, Facebook’s classification system has downstream implications for political discourse and processes to the extent that campaigns and activists can and do employ the platform in the future. In general, Facebook’s classification system grows around Facebook’s economic imperatives and, thus, seems to primarily serve those with “buying power,” in terms of both social and economic capital. In this way, it upholds existing power structures in the political sphere in a number of ways. When communities are visible and, crucially, represented on their own terms, they can better find each other and organize politically. The relative scarcity of categories in Facebook’s classification system addressing interests of POC, women, LGBTQ+ people, as well as intersections of these identities, constrains options for communities to reach each other and for outsider political advertisers to address their interests.

Likewise, if candidates, activists, and advocacy organizations find categories hateful or offensive (personally or on behalf of others), they may choose to avoid them, which eliminates avenues for marginalized communities to use in political organizing and advocacy work. At the same time, it is important to note that rendering marginalized communities visible in

Facebook's classification system, in and of itself, can cause harm as it can open them up to harassment, discrimination, or predatory acts—for example, the targeting of African Americans by Russian disinformation campaigns and by the Trump campaign in the 2016 U.S. election to suppress voter turnout (Wong, 2020). More abstractly, inclusion premised on the extraction of value can contribute to the perpetuation of social hierarchies even when it superficially appears to redress inequities (Cottom, 2020). The question of *when* representation in Facebook's classification system imposes harm and how to balance harm with good in constructing classification systems for ad targeting requires additional attention.

The classification system's implication of an unmarked user defined by characteristics of dominant groups further complicates matters of representation. The disproportionate number of categories referring to nondominant groups has the potential to silo strategic political speech, limiting the extent to which political advertisers can speak to a broad audience about issues that matter to and for these groups. Consider, for example, the case of categories for "Science" versus "Women in Science." According to Audience Insights, the former targets an audience that is 38% men, while the latter targets an audience that is 18% men. This structuring of strategic political speech suggests women in science is treated as a "women's issue," not a matter of importance to society broadly, thus making it harder to bring this issue to the fore of mainstream political discourse. Relatedly, the classification system's both-sides-ism means that political advertisers may find it difficult to reach likeminded, receptive audiences when targeting interest categories that quietly encode ideologies at odds with advertising messages. This problem may lead to higher costs and smaller rewards (see Ali et al., 2019), particularly for advertisers advocating for political issues at the intersection of civil rights and structural power, which tend to be particularly polarizing.

In dissecting these dilemmas, we call into question whether the system can and does serve political campaigns and activists advocating for issues of importance to marginalized communities equally as well as others. We hope that our findings will help us, as a society, more effectively formulate normative questions about what role platform microtargeting should play in political processes. While not all of the dilemmas we described can be resolved, there are steps Facebook could take to make meaningful progress towards a more inclusive classification system. Here, Costanza-Chock's (2020) framework for "design justice" is informative: Facebook should aim for "full inclusion of, accountability to, and ultimately control by people with direct lived experience of the conditions the

design team is trying to change" (99). This means involving and considering advertisers and users of all different races, genders, sexual orientations, disability statuses, immigration statuses, and so on throughout the ongoing process of developing the classification system, in order to best ensure that its benefits and burdens are equitably distributed (Costanza-Chock, 2020). Beyond design, creating regulatory "safeguards" via structures and procedures for greater transparency, accountability, and oversight may also help ward off any potential harms the classification system may pose for political processes (Citron and Pasquale, 2014). Alternately, as the *Feminist Data Manifest-No* argues, sometimes the most just solution to "harmful data regimes," albeit radical, is refusal (Cifor et al., 2019). While wholesale refusal may be considered a nonstarter in the immediate present, refusal of the use of the classification system for political ad targeting may not be given the preponderance of concerns about political advertising on Facebook that intersect with those we raised.

Our findings come with several caveats. Our methodological approach, combined with the limits Facebook imposes on access to its classification system, do not permit us to speak to the scale of the dilemmas we demonstrated. Our principal aim was to provide thick description of the political nature of algorithmic classification for (political) ad targeting. Thus, additional work will be needed to explore how widespread these issues may be and their relative impact on the advertising practices of political candidates as well as advocacy organizations who use Facebook advertising to mobilize support. Early work investigating Facebook ad targeting suggests that interest targeting is widely used, but is used less by larger advertisers, who tend to rely on their own lists of users or algorithmic replication of past campaigns (Edelson et al., 2019; Ghosh et al., 2019). Moreover, this work suggests that different kinds of advertisers rely on interest targeting to different degrees. For example, political advertisers that spend less tend to use interest targeting more (Ghosh et al., 2019). This may be because these advertisers lack the resources for assembling or acquiring extensive voter files or databases. Future research should explore which kinds of (political) advertisers are affected by the dilemmas we brought forward, and to what extent.

There is also more work to be done to consider the ways that Facebook's ad *delivery* system intersects with our findings. This system attempts to deliver ads to audiences based on a variety of inputs, including targeting options selected, but also user attributes not explicitly targeted (Ali et al., 2019), bid prices across advertisers, optimization criteria, and so on (Andreou et al., 2018). It is not clear whether/how the ad delivery

process remediates or exacerbates the dilemmas we identified, though past work demonstrating discriminatory outcomes of ad delivery systems (Datta et al., 2015) seems to point towards the latter.

Our findings provide a window into the subtle but significant ways political choices reverberate through Facebook's infrastructure and enable its profitability. As we attend to the platform's increasingly important role in politics, we should acknowledge the potential harms and limitations associated with algorithmically classifying users for ad targeting. Moving forward, we must continue to reflect on whose interests are classified, why, and how. In particular, we will need to consider how well algorithmic classification for targeted advertising serves activist and advocacy work by people and organizations with varying degrees of power and privilege. In short, we must continue to unpack the complicated politics that underlie the significance of "right" in Facebook's promise to help advertisers "reach the right people."

### Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Kelley Cotter  <https://orcid.org/0000-0003-1243-0131>

Chankyung Pak  <https://orcid.org/0000-0003-4348-6939>

Kjerstin Thorson  <https://orcid.org/0000-0002-5762-4688>

### Notes

1. From Bowker and Star (1999: 1).
2. Additional details on methodology and measures taken to protect participants' privacy are included in Supplementary Materials.
3. Not all of the available targeting categories are grouped under top-level topics (e.g., the category for "Activism" is not assigned to a top-level topic). Some categories are also grouped into top-level topics in unintuitive ways (e.g., the category "Power (social and political)" falls under "Business and industry").
4. Some categories, like this one, no longer appear in the classification system.
5. See Supplementary Materials for additional information.
6. In our donated data, 59% of liked pages were never translated into ad categories.
7. Facebook no longer embeds with political campaigns, though does continue to offer remote support.
8. Facebook curtailed this practice in 2018 (Statt, 2018).
9. At the time of writing.
10. Recall, for example, that Facebook previously eliminated anti-Semitic categories from the classification system.

### Supplemental Material

Supplemental material for this article is available online.

### References

- Aguilera J (2016) Another Word for 'illegal alien' at the Library of Congress: Contentious. *The New York Times*, 22 July. Available at: <https://www.nytimes.com/2016/07/23/us/another-word-for-illegal-alien-at-the-library-of-congress-contentious.html> (accessed 9 February 2021).
- Ali M, Sapiezynski P, Korolova A, et al. (2019) Ad delivery algorithms: The hidden arbiters of political messaging. Available at: <https://arxiv.org/pdf/1912.04255> (accessed 9 February 2021).
- Andreou A, Venkatadri G, Goga O, et al. (2018) *Investigating ad transparency mechanisms in social media: A case study of Facebook's explanations*. Available at: <https://hal.archives-ouvertes.fr/hal-01955309> (accessed 9 February 2021).
- Angwin J, Varner M and Tobin A (2017) Facebook enabled advertisers to reach 'Jew Haters'. Available at: <https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters> (accessed 9 February 2021).
- Barocas S (2012) The price of precision: Voter microtargeting and its potential harms to the democratic process. In: *Proceedings of the first edition workshop on politics, elections and data – PLEAD '12*, Maui, Hawaii, 2 November 2012, p. 31. New York: ACM Press.
- Barrett B and Kreiss D (2019) Platform transience: Changes in Facebook's policies, procedures, and affordances in global electoral politics. *Internet Policy Review* 8(4): 1–22.
- Beckett L (2017) Trump digital director says Facebook helped win the White House. *The Guardian*, 9 October. Available at: <https://www.theguardian.com/technology/2017/oct/08/trump-digital-director-brad-parscale-face-book-advertising> (accessed 9 February 2021).
- Biddle S (2020) Facebook managers trash their own ad targeting in unsealed remarks. *The Intercept*, December 14. Available at: <https://theintercept.com/2020/12/24/facebook-ad-targeting-small-business/> (accessed 9 February 2021).
- Bowker GC and Star SL (1999) *Sorting Things out*. Cambridge, MA: MIT Press.
- Brady HE, Schlozman KL and Verba S (1999) Prospecting for participants: Rational expectations and the recruitment of political activists. *American Political Science Review* 93(1): 153–168.
- Buolamwini J (2016) InCoding – In the beginning was the coded gaze. Available at: <https://medium.com/mit-media-lab/incoding-in-the-beginning-4e2a5c51a45d> (accessed 9 February 2021).
- Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3(1): 1–12.
- Caliskan A, Bryson JJ and Narayanan A (2017) Semantics derived automatically from language corpora contain human-like biases. *Science* 356(6334): 183–186.

- Cheney-Lippold J (2017) *We Are Data*. New York: New York University Press.
- Cifor M, Garcia P, Cowan T, et al. (2019) Feminist Data Manifest-No. Available at: <https://www.manifestno.com> (accessed 9 February 2021).
- Citron D and Pasquale F (2014) The scored society: Due process for automated predictions. *Washington Law Review* 89(1): 1–33.
- Collins PH (2014) *Black Feminist Thought*. New York: Routledge.
- Costanza-Chock S (2020) *Design Justice*. Cambridge, MA: The MIT Press.
- Cottom TM (2020) Where platform capitalism and racial capitalism meet: the sociology of race and racism in the digital society. *Sociology of Race and Ethnicity* 6(4): 441–449.
- Crenshaw K (1989) Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum Article* 8(139): 139–169.
- Datta A, Tschantz Mc and Datta A (2015) Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. *Proceedings on Privacy Enhancing Technologies* 2015(1): 92–112.
- Drescher J (2015) Out of DSM: Depathologizing homosexuality. *Behavioral Sciences* 5(4): 565–575.
- Edelson L, Sakhuja S, Dey R, et al. (2019) An analysis of united states online political advertising transparency. Available at: <https://arxiv.org/abs/1902.04385>
- Facebook (2015a) Creative strategies with topic data. Available at: <https://www.facebook.com/business/news/insights/creative-strategies-with-topic-data> (accessed 9 February 2021).
- Facebook (2015b) Digital diversity: A closer look at Asian Americans in the US. Available at: <https://www.facebook.com/business/news/insights/digital-diversity-a-closer-look-at-asian-americans-in-the-us> (accessed 9 February 2021).
- Facebook (2016) Reach matters: Driving business results at scale. Available at: <https://www.facebook.com/business/news/insights/reach-matters-driving-business-results-at-scale> (accessed 9 February 2021).
- Facebook (2018) Keeping advertising safe and civil. Available at: <https://www.facebook.com/business/news/keeping-advertising-safe-and-civil> (accessed 9 February 2021).
- Facebook (2019) Targeting core audiences. Available at: <https://www.facebookblueprint.com/student/path/196017-targeting-core-audiences> (accessed 9 February 2021).
- Facebook (2020) New steps to protect the US elections. Available at: <https://about.fb.com/news/2020/09/additional-steps-to-protect-the-us-elections/> (accessed 9 February 2021).
- Facebook (n.d.-a) Facebook ad targeting: Online courses. Available at: <https://www.facebook.com/business/learn/categories/targeting> (accessed 9 February 2021).
- Facebook (n.d.-b) At Facebook. Available at: <https://diversity.fb.com/initiatives/at-facebook/> (accessed 9 February 2021).
- Flyverbom M and Murray J (2018) Datastructuring – Organizing and curating digital traces into action. *Big Data & Society* 5(2): 1–12.
- Frymer P (2010) *Uneasy Alliances*. Princeton, N.J.: Princeton University Press.
- Gandy OH (2000) Audience construction: Race, ethnicity and segmentation in popular media. In: *50th annual conference of the international communication association*, Acapulco, Mexico, 1–5 June.
- Ghosh A, Venkatadri G and Mislove A (2019) Analyzing political advertisers' use of Facebook's targeting features. In: *IEEE workshop on technology and consumer protection (ConPro'19)*, San Francisco, CA, 23, May.
- Ghosh D and Scott B (2018) Facebook's new controversy shows how easily online political ads can manipulate you. Available at: <https://time.com/5197255/facebook-cambridge-analytica-donald-trump-ads-data/> (accessed 9 February 2021).
- Gillespie T (2010) The politics of 'Platforms'. *New Media & Society* 12(3): 347–364.
- Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies*. Cambridge, MA: The MIT Press, pp. 167–194.
- Gillespie T (2018) *Custodians of the Internet*. New Haven, CT: Yale University Press.
- GLAAD (2011) GLAAD media reference guide – Transgender. Available at: <https://www.glaad.org/reference/transgender> (accessed 9 February 2021).
- Gorton WA (2016) Manipulating citizens: How political campaigns' use of behavioral social science harms democracy. *New Political Science* 38(1): 61–80.
- Hersh ED (2015) *Hacking the Electorate*. New York: Cambridge University Press.
- Karpf D (2018) *A world without wizards: On Facebook and Cambridge Analytica*. Available at: <https://civichall.org/civicist/world-without-wizards-facebook-cambridge-analytica/> (accessed 9 February 2021).
- Kim YM (2016) Algorithmic opportunity: Digital advertising and inequality in political involvement. *The Forum* 14(4): 471–484.
- Kim YM, Hsu J, Neiman D, et al. (2018) The stealth media? Groups and targets behind divisive issue campaigns on Facebook. *Political Communication* 35(4): 515–541.
- Kitchin R (2017) Thinking critically about and researching algorithms. *Information, Communication & Society* 20(1): 14–29.
- Kitchin R and Dodge M (2011) *Code/Space. Software Studies*. Cambridge, MA: MIT Press.
- Kreiss D (2012) Yes we can (profile you): A brief primer on campaigns and political data. *Stanford Law Review Online*, 64. Available at: <http://www.stanfordlawreview.org/online/privacy-paradox/political-data> (accessed 9 February 2021).
- Kreiss D and McGregor SC (2018) Technology firms shape political communication: The work of Microsoft, Facebook, Twitter, and Google with campaigns during the 2016 U.S. presidential cycle. *Political Communication* 35(2): 155–177.

- Kreiss D and McGregor SC (2019) The “arbiters of what our voters see”: Facebook and Google’s struggle with policy, process, and enforcement around political advertising. *Political Communication* 36(4): 499–522.
- Lenhart A, Ybarra M, Zickuhr K, et al. (2016) Online harassment, digital abuse, and cyberstalking in America. *Data & Society*. Available at: <https://datasociety.net/library/online-harassment-digital-abuse-cyberstalking/> (accessed 9 February 2021).
- Mackenzie A (2015) The production of prediction: What does machine learning want? *European Journal of Cultural Studies* 18(4–5): 429–445.
- Massanari A (2017) #Gamergate and the fapping: How Reddit’s algorithm, governance, and culture support toxic technocultures. *New Media & Society* 19(3): 329–346.
- Noble SU (2018) *Algorithms of Oppression*. New York: New York University Press.
- Nteta T and Schaffner B (2013) Substance and symbolism: Race, ethnicity, and campaign appeals in the United States. *Political Communication* 30(2): 232–253.
- Olson HA (1998) Mapping beyond Dewey’s boundaries: Constructing classificatory space for marginalized knowledge domains. *Library Trends* 47(2): 233–254.
- Park Y-J and Tuzhilin A (2008) The long tail of recommender systems and how to leverage it. In: *Proceedings of the 2008 ACM conference on recommender systems – RecSys ’08*, Lausanne, Switzerland, October 23–25, 2008, p. 11. New York: ACM Press.
- Pasquale F (2015) *The Black Box Society*. Cambridge, MA: Harvard University Press.
- Peiss KL (1998) American women and the making of modern consumer culture. *The Journal for MultiMedia History* 1(1). Available at: <https://www.albany.edu/jmmh/vol1no1/peiss-text.html>
- Phillips W (2018) The oxygen of amplification. *Data & Society*, pp. 1–15.
- Pybus J (2019) Trump, the first Facebook president: Why politicians need our data too. In: Happer C, Hoskins A, and Merrin W (eds) *Trump’s Media War*. Cham, Switzerland: Springer International Publishing, pp. 227–240. DOI: 10.1007/978-3-319-94069-4\_14.
- Rajaram G, Wu G, Yan H, et al. (2014) Providing content using inferred topics extracted from communications in a social networking system. US 2014/0052540 A1. United States.
- Rieder B (2017) Scrutinizing an algorithmic technique: The Bayes classifier as interested reading of reality. *Information, Communication & Society* 20(1): 100–117.
- Rosen G (2020) Preparing for Election Day. Available at: <https://about.fb.com/news/2020/10/preparing-for-election-day/> (accessed 9 February 2021).
- Schier SE (2000) *By Invitation Only*. Pittsburgh, PA: University of Pittsburgh Press.
- Schiff A (2018) Facebook eliminates 5,000 ad targeting options to pull the plug on prejudice. Available at: <https://adexchanger.com/platforms/facebook-eliminates-5000-ad-targeting-options-to-pull-the-plug-on-prejudice/> (accessed 9 February 2021).
- Schiff A (2019) Mamma Mia, Facebook is killing thousands of unused ad targeting options. Available at: <https://adexchanger.com/platforms/mamma-mia-facebook-is-killing-thousands-of-unused-ad-targeting-options/> (accessed 9 February 2021).
- Sender K (2018) The gay market is dead, long live the gay market: From identity to algorithm in predicting consumer behavior. *Advertising & Society Quarterly* 18(4).
- Speicher T, Ali M, Venkatadri G, et al. (2018) Potential for discrimination in online targeted advertising. In: *FAT 2018-conference on fairness, accountability, and transparency*, 23–24 February 2018, pp. 1–15. New York, NY: Proceedings of Machine Learning Research.
- SPLC (n.d.-a) Center for Immigration Studies. Available at: <https://www.splcenter.org/fighting-hate/extremist-files/group/center-immigration-studies> (accessed 9 February 2021).
- SPLC (n.d.-b) Federation for American Immigration Reform. Available at: <https://www.splcenter.org/fighting-hate/extremist-files/group/federation-american-immigration-reform> (accessed 9 February 2021).
- Srnicek N (2016) *Platform Capitalism*. Cambridge, UK: Polity Press.
- Statt N (2018) Facebook will no longer allow third-party data for targeting ads. Available at: <https://www.theverge.com/2018/3/28/17174854/facebook-shutting-down-partner-categories-ad-targeting-cambridge-analytica> (accessed 9 February 2021).
- Sturken M and Cartwright L (2001) Modernity: Spectatorship, power, and knowledge. *Practices of Looking*. Oxford: Oxford University Press, pp. 72–108.
- Sweeney L (2013) Discrimination in online ad delivery. *Communications of the ACM* 56(5): 44–54.
- Thorson K, Cotter K, Medeiros M, et al. (2019) Algorithmic inference, political interest, and exposure to news and politics on facebook. *Information, Communication & Society*, pp. 1–18.
- Tobin A (2019) HUD sues Facebook over housing discrimination and says the company’s algorithms have made the problem worse. Available at: <https://www.propublica.org/article/hud-sues-facebook-housing-discrimination-advertising-algorithms> (accessed 9 February 2021).
- Turov J (2011) *The Daily You*. New Haven: Yale Univ. Press.
- Verba S, Schlozman KL and Brady HE (1995) *Voice and Equality*. Cambridge, MA: Harvard University Press.
- Wong JC (2020) ‘Putin could only dream of it’: How Trump became the biggest source of disinformation in 2020. *The Guardian*, November 2. Available at: <https://www.theguardian.com/us-news/2020/nov/02/trump-us-election-disinformation-russia> (accessed 9 February 2021).
- Yan R, Singla A, Garcia-Martinez AF, et al. (2014) Inferring target clusters based on social connections. U.S. Patent No. 2014/00894.00A1. Washington, DC.
- Zhou D and Moreels P (2014) Inferring user profile attributes from social information. US 8,856.235 B2. Washington, DC.
- Zuboff S (2018) *The Age of Surveillance Capitalism*. New York: PublicAffairs.