A Qualitative Exploration of Perceptions of Algorithmic Fairness

Allison Woodruff[1], **Sarah E. Fox**[2], **Steven Rousso-Schindler**[3], **and Jeff Warshaw**[4]
[1] Google, woodruff@acm.org
[2] Google and Human Centered Design & Engineering, University of Washington, sefox@uw.edu
[3] Department of Anthropology, CSU Long Beach, steven.rousso-schindler@csulb.edu
[4] Google, jeffwarshaw@google.com

ABSTRACT
Algorithmic systems increasingly shape information people are exposed to as well as influence decisions about employment, finances, and other opportunities. In some cases, algorithmic systems may be more or less favorable to certain groups or individuals, sparking substantial discussion of algorithmic fairness in public policy circles, academia, and the press. We broaden this discussion by exploring how members of potentially affected communities feel about algorithmic fairness. We conducted workshops and interviews with 44 participants from several populations traditionally marginalized by categories of race or class in the United States. While the concept of algorithmic fairness was largely unfamiliar, learning about algorithmic (un)fairness elicited negative feelings that connect to current national discussions about racial injustice and economic inequality. In addition to their concerns about potential harms to themselves and society, participants also indicated that algorithmic fairness (or lack thereof) could substantially affect their trust in a company or product.

Author Keywords
Algorithmic fairness; algorithmic discrimination

ACM Classification Keywords
K.4.m. Computers and Society: Miscellaneous.

INTRODUCTION
Scholars and thought leaders have observed the increasing role and influence of algorithms in society, pointing out that they mediate our perception and knowledge of the world as well as affect our chances and opportunities in life [6,8,17,38,54,55,63,76,79]. Further, academics and regulators have long refuted the presumption that algorithms are wholly objective, observing that algorithms can reflect or amplify human or structural bias, or introduce complex biases of their own [4,10,18,33–35,38,46,64]. To raise awareness and illustrate the potential for wide-ranging consequences, researchers and the press have pointed out a number of specific instances of algorithmic unfairness [19,58], for example, in predictive policing [19,43], the online housing marketplace [27,28], online ads [13,17,20,29,82], and image search results [49,64].

Such cases demonstrate that algorithmic (un)fairness is a complex, industry-wide issue. Bias can result from many causes, for example, data sets that reflect structural bias in society, human prejudice, product decisions that disadvantage certain populations, or unintended consequences of complicated interactions among multiple technical systems. Accordingly, many players in the ecosystem, including but not limited to policy makers, companies, advocates, and researchers, have a shared responsibility and opportunity to pursue fairness. Algorithmic fairness, therefore, appears to be a "wicked problem" [72], with diverse stakeholders but, as yet, no clear agreement on problem statement or solution. The human computer interaction (HCI) community and related disciplines are of course highly interested in influencing positive action on such issues [25], having for example an established tradition of conducting research to inform public policy for societal-scale challenges [50,84] as well as providing companies information about how they can best serve their users. Indeed, recent work by Plane et al. on discrimination in online advertising is positioned as informing public policy as well as company initiatives [67].

Building on this tradition, our goal in this research was to explore ethical and pragmatic aspects of public perception of algorithmic fairness. To this end, we conducted a qualitative study with several populations that have traditionally been marginalized and are likely to be affected by algorithmic (un)fairness, specifically, Black or African American, Hispanic or Latinx, and low socioeconomic status participants in the United States. Our research questions centered around participants' interpretations and experiences of algorithmic (un)fairness, as well as their ascription of accountability and their ethical and pragmatic expectations of stakeholders. In order to draw more robust conclusions about how participants interpret these highly contextual issues, we explored a broad spectrum of different types of algorithmic unfairness, using scenarios to make the discussion concrete.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada.

ACM ISBN 978-1-4503-5620-6/18/04.
https://doi.org/10.1145/3173574.3174230

Our findings indicate that while the concept of algorithmic (un)fairness was initially mostly unfamiliar and participants often perceived algorithmic systems as having limited impact, they were still deeply concerned about algorithmic unfairness, they often expected companies to address it regardless of its source, and a company's response to algorithmic unfairness could substantially impact user trust. These findings can inform a variety of stakeholders, from policy makers to corporations, and they bolster the widely espoused notion that algorithmic fairness is a societally important goal for stakeholders across the ecosystem—from regulator to industry practitioner—to pursue. With full recognition of the importance of ethical motivations, these findings also suggest that algorithmic fairness can be good business practice. Some readers may be in search of arguments to motivate or persuade companies to take steps to improve algorithmic fairness. There are many good reasons for companies to care about fairness, including but not limited to ethical and moral imperatives, legal requirements, regulatory risk, and public relations and brand risk. In this paper, we provide additional motivation by illustrating that user trust is an important but understudied pragmatic incentive for companies across the technology sector to pursue algorithmic fairness. Based on our findings, we outline three best practices for pursuing algorithmic fairness.

BACKGROUND

Algorithmic Fairness

In taking up algorithmic fairness, we draw on and seek to extend emerging strands of thought within the fields of science and technology studies (STS), HCI, mathematics, and related disciplines. Research on algorithmic fairness encompasses a wide range of issues, for example, in some cases considering discrete decisions and their impact on individuals (e.g. fair division algorithms explored in [51,52]), and in other cases exploring broader patterns related to groups that have traditionally been marginalized in society. Our focus tends towards the latter, and of particular relevance to our investigation is the perspective taken in critical algorithm studies, which articulates the increasing influence of algorithms in society and largely focuses on understanding algorithms as an object of social concern [6,17,38,54,55,63,76,79]. Countering popular claims that algorithmic authority or data-driven decisions may lead to increased objectivity, many scholars have observed that algorithms can reflect, amplify or introduce bias [4,10,18,33–35,38,46,64].

Articles in academic venues as well as the popular press have chronicled specific instances of unjust or prejudicial treatment of people, based on categories like race, sexual orientation, or gender, through algorithmic systems or algorithmically aided decision-making. For example, Perez reported that Microsoft's Tay (an artificial intelligence chatbot) suffered a coordinated attack that led it to exhibit racist behavior [65]. Researchers have also reported that image search or predictive search results may reinforce or exaggerate societal bias or negative stereotypes related to race, gender, or sexual orientation [4,49,62,64]. Others raised concerns about potential use of Facebook activity to compute non-regulated credit scores, especially as this may disproportionately disadvantage less privileged populations [17,82]. Edelman et al. ran experiments on Airbnb and reported that applications from guests with distinctively African American names were 16% less likely to be accepted relative to identical guests with distinctively White names [28]. Edelman and Luca also found non-Black hosts were able to charge approximately 12% more than Black hosts, holding location, rental characteristics, and quality constant [27]. Colley et al. found Pokémon GO advantaged urban, white, non-Hispanic populations, for example, potentially attracting more tourist commerce to their neighborhoods [15], and Johnson et al. found that geolocation inference algorithms exhibited substantially worse performance for underrepresented populations, i.e., rural users [47].

This public awareness has been accompanied by increased legal and regulatory attention. For example, the upcoming European Union General Data Protection Regulation contains an article on 'automated individual decision-making' [39]. Yet, algorithmic fairness poses many legal complexities and challenges [5] and law and regulation are still in nascent stages in this rapidly changing field (e.g. [9]). To investigate systems' adherence to emerging legal, regulatory, and ethical standards of algorithmic fairness, both testing and transparency have been called for [1,14,77]. A wide range of techniques have been proposed to scrutinize algorithms, such as model interpretability, audits, expert analysis, and reverse engineering [22,42,76,77]. Investigation is complicated however by the myriad potential causes of unfairness (prejudice, structural bias, choice of training data, complex interactions of human behavior with machine learning models, unforeseen supply and demand effects of online bidding processes, etc.) and the sometimes impenetrable and opaque nature of machine learning systems [12,38]. In fact, existing offline discrimination problems may in some cases be exacerbated and harder to investigate once they manifest in online systems [77], and new bigotries based not just on immutable characteristics but more subtle features may arise which are more difficult to detect than traditional discriminatory processes [9].

Not only do opacity and complexity complicate expert analysis, but they may also make it difficult for stakeholders to understand the consequences of algorithmic systems. Many of the proposed mechanisms for scrutinizing algorithms make certain assumptions about the public, regulators, and other stakeholders. However, research has found that perception of algorithmic systems can vary substantially by individual factors as well as platform [21], and that end users often have fundamental questions or misconceptions about technical details of their operation

[11,31,69,85,86], an effect that may be exacerbated for less privileged populations [86]. For example, studies have found that some participants are not aware of algorithmic curation in the Facebook News Feed [31,69] or the gathering of online behavioral data and its use for inferencing [86], or underestimate the prevalence and scale of data gathering and its use in practical applications [85,86]. Further, participants often emphasize the role of human decision-making in algorithmic systems, for example, misattributing algorithmic curation in the Facebook News Feed to actions taken by their friends and family [31], or framing algorithms as calculator-like tools that support human decision-making [86].

Despite this existing research on algorithmic literacy, very little research has explored understandings of algorithmic (un)fairness, and there is currently little insight into how the general public and in particular people affected by algorithmic unfairness might perceive it. In a rare exception, Plane et al. surveyed a broad population in the US, including a near-census representative panel, regarding their responses to online behavioral advertising (OBA) scenarios that used race as a targeting variable for a job ad [67]. Overall, almost half of the respondents viewed the scenarios as a moderate or severe problem, with Black respondents finding them to be of higher severity. We offer a complementary and novel exploration of algorithmic (un)fairness, in that: (1) we explore a much wider range of potential types of algorithmic unfairness; (2) we take a qualitative approach that allows us to deeply explore issues with a smaller population, which is complementary to Plane et al.'s more narrow quantitative exploration with a larger and more representative sample [67]; and (3) we focus on populations that are more likely to be affected by algorithmic unfairness, rather than the general public.

Workshop as Method

In taking up a workshop format, we draw on traditions within and just beyond HCI. This includes programs of participatory action research, participatory design, and living labs. Within the context of HCI and design research, workshop approaches often seek to invite members of the public to engage with practices of design while exploring values and beliefs around technology with each other, positing alternative techniques and outcomes. Noting the collaborative and situated nature of the approach, Rosner et al. describe the design workshop as inviting "a treatment of collaboration and interdisciplinary as a localized and imaginative practice" [74]. These engagements rely on careful collaboration between researcher and subject/partner, across sites like academic or industrial research centers and community groups each with their own goals for the work. Relatedly, research on the public understanding of science argues against assuming a single correct understanding of science and technology, emphasizing that members of the public should not be excluded from democratic decision-making about technology because their interpretations of technology may be different from those of technological experts [87]. Taking this perspective, we orient to our workshop attendees as experts in how technology is experienced in their daily lives—a framing that speaks to their own sets of knowledges that are different, but not any less than, those of technological experts.

In the 1980s, HCI scholars Jungk and Müllert first described the future workshop as a format for social engagement which involved the organization of events with members of the public meant to better address issues of democratic concern [48]. Similar in its political roots, participatory design is a method focused on more actively including members of the public or other under-represented stakeholders in the processes of design. Early examples of this work, from the 1980s, aimed to support worker autonomy and appreciation of traditional expertise in light of the introduction of digitized work practices and, in some cases, automation of labor. For example, Pelle Ehn, a design scholar and longtime proponent of participatory design, collaborated with a Scandinavian graphic designers union to produce a software system meant to better incorporate their skilled practices, compared with management-initiated programs [30].

More contemporary participatory initiatives have taken up concerns outside of work or governmental contexts, from exploring alternative food systems [23,24] to understanding how to promote play among neurodiverse children [80]. Still others have developed the design workshop as a means of examining critical theory through material practice like making and tinkering [70] or used craft to imagine alternative near futures that might yield more equitable social arrangements [3,75].

Here, we build on this legacy of participatory programs by reporting on our use of the workshop format as research instrument toward understanding not only how participants perceive algorithmic (un)fairness, but also how they might elect to construct platforms differently. Due to the potentially sensitive nature of the subject matter we looked to dialogical approaches like participatory design as a helpful technique for collaboratively working through complex ideas (e.g. machine learning) and developing an open environment for sharing feelings and opinions. We see these discussions and subsequent ideas as informing the development of technology and policy as well as communication with diverse users in the future.

METHODOLOGY

In order to better understand how members of marginalized communities perceive algorithmic (un)fairness, we conducted participatory design workshops with members of various communities throughout the San Francisco Bay Area. We then conducted individual follow-up interviews with select participants. The workshops and interviews took place July through September of 2016.

Participants

We recruited 44 adults, all of whom responded to a screener survey administered by a national research recruitment firm with a respondent database including San Francisco Bay Area residents. Participants were compensated for their time, at or above the living wage for their area. Our recruiting focused on inviting individuals who were traditionally marginalized either by categories of socioeconomic status or race, and we organized our participants into five workshops as follows: two workshops based on socioeconomic status as described below; one workshop with participants who identified as Black or African American women; one workshop with Black or African American or mixed race men and women; and one workshop with Hispanic or Latinx men and women. While our work was qualitative and non-representative, we expect the constituencies on which we focused comprise roughly between 40% and 50% of the US population.[1]

Primary factors in considering socioeconomic status were current household income and education level. Selected participants had an annual household income of less than the living wage for their home county—an amount determined from a coarse approximation of Glasmeier's Living Wage Model (livingwage.mit.edu, accessed July-August 2016). In factoring this amount, we considered the total number of adults in the household, the number of adults contributing to the income, the number of dependent children in the household, and the number of children outside the household cared for financially by the respondent. Participants had also earned no more than "some college," defined here as up to 4 years of course taking without receiving an Associate's or Bachelor's degree. As secondary factors contributing to socioeconomic status determination, we also considered the respondent's current occupation and location of residence. With this, the focus was on understanding the respondent's current economic situation as well as near term opportunity for advancement based on proximate resources.

For the remainder of the workshops, our recruitment focused on inviting people of color, based on their responses in the recruitment screener. As a secondary consideration we also looked to respondent's occupation, slightly emphasizing those involved in care or service professions—skills and expertise often underrecognized in technology cultures [57,71].

Most of the participants were from the East Bay and San Francisco, with a wide range of ages (18-65+) and occupations (e.g. public transportation driver, retail manager, special education instructor, community activities coordinator, tasker, line cook, laborer, correctional peace officer, office assistant, theater assistant).

Workshop

Each group participated in a 5-hour workshop, with the following agenda: an icebreaker activity; a group discussion of algorithmic (un)fairness; a meal; a design activity centered around three cases; and a concluding group discussion. In attendance at each workshop were between 6 and 11 participants, 2 researchers who acted as facilitators, and a visual anthropologist who focused on documentation. Participants were aware of Google's involvement in the study, and the workshops took place at a Google location. During the workshops, we took care to encourage collaborative interpretation, problem-solving, and discussion among participants, and to make space for all participants to share their ideas and opinions. Additionally, recognizing the emotional complexity of the topic, we explained that there might be sensitive material, and that participants should feel free to stop participating, sit out on an activity, or step out of the room.

To start the day, we asked participants to take part in an icebreaker activity inspired by anti-racism scholar Peggy McIntosh's Invisible Knapsack exercise [56,78], meant to begin to discuss issues of discrimination, power, and privilege in a non-confrontational manner. After this initial activity, the researchers gave a brief description of algorithms and algorithmic (un)fairness. Broad discussion revolved around participant questions and interpretation of algorithmic (un)fairness, whether participants knew about it prior to the workshop or had ever experienced it, and sharing of general feelings about it. Note that during the workshop we used the term "algorithmic discrimination" rather than "algorithmic (un)fairness." While "algorithmic fairness" is often used as a term in the academic literature, our experience in this study as well as other work at our institution suggests that in a user research context "fairness" may be construed overly narrowly (for example, as emphasizing equality rather than justice) and therefore we preferred to use "algorithmic discrimination" in our conversations with participants.

For the bulk of the day, we focused on a series of three scenario-based design activities. We began each scenario by describing a case that could be understood as an instance of algorithmic unfairness, and then invited participants to share their initial reactions in a brief group discussion. During this discussion, we also occasionally introduced various complexities, for example suggesting different potential causes of unfairness. Then, we asked participants

[1] The US Census Bureau estimates that as of July 2016, the Black or African American population constitutes 13.3% (43 million people) of the total US population (323.1 million people), the Hispanic or Latino population is 17.8% (57.5 million people), and the population with two or more races is 2.6%. (https://www.census.gov/quickfacts, accessed August 2017). While we were not able to find an estimated percentage of the US population that meets the living wage standard, the poverty rate in 2015 was 13.5% (43.1 million people), approximately 51% of whom were Black or Hispanic [68]. Since the living wage exceeds the poverty threshold, we expect that substantially more than 13.5% would not meet the living wage standard [61], and in fact the number seems likely to be closer to the 29% of Americans that Pew identified as living in a lower-class household [37]. Overall this suggests that the populations we focused on (although with only a small, qualitative sample) conservatively comprise nearly 40% of the US population, and more likely slightly over 50%.

to spend 10 minutes working individually to come up with ideas about what they might do if they were a decision-maker at a technology company in charge of responding to the scenario. We told participants they were free to express their ideas using any means of communication they found most comfortable—drawing, story writing, performing were all examples given. After they worked and recorded their ideas, we came back together as a group and went around the table to share and discuss everyone's ideas.

The scenarios we discussed represented a wide range of issues. While the scenarios were based on internet-related products and services, we also encouraged discussion of other domains and the discussion often branched out to other areas in which algorithmic unfairness might occur. The first scenario described a man visiting a newspaper website and seeing ads for high-paying jobs, while a woman visiting the same website saw ads for low-wage work.[2] The second scenario was about results of predictive search (a feature which suggests possible search terms as the user types into a search box) that could be interpreted as stereotyping Black men and children as criminals.[3] With the third and final scenario, we asked participants to consider a practice of excluding businesses in neighborhoods with high crime rates from an online restaurant reviewing and map application.[4] After we completed all three scenarios, we concluded the workshop with a broad group discussion reflecting back on ideas that had emerged throughout the day and the experience of the workshop as a whole.

Interviews

After the workshops were completed, we conducted follow-up interviews approximately one hour in length with 11 participants who appeared particularly engaged during the workshop discussions. Interviews were semi-structured, with questions focused on gaining further understanding of the participant's concerns, opinions, and policy ideas.

Analysis

All interviews were video-recorded and transcribed. In our analysis, we used a general inductive approach [83], which relies on detailed readings of raw data to derive themes relevant to evaluation objectives. In our case, the primary evaluation objective was to inform technical and policy approaches to algorithmic fairness by learning about: (1) participants' interpretation of algorithmic fairness; and (2) participants' ascription of accountability and their ethical and pragmatic expectations of stakeholders, especially companies. Accordingly, we focused on these issues during our time with the participants, and then we jointly analyzed the data from both the workshops and interviews by closely reviewing the text and videos, performing affinity clusterings of textual quotations and video clips to identify emergent themes [7], producing short films synthesizing key themes using a visual ethnographic approach [66], and iteratively revising and refining categories. In keeping with the general inductive approach, our analytic process yielded a small number of summary categories, which we describe in the Findings section below.

Limitations

We note several limitations of our study methodology that should be considered when interpreting this work. First, due to our focus on traditionally marginalized populations, we did not gather data about how more privileged populations think about or experience algorithmic fairness. Second, our sample was not statistically representative of the populations we explored. The findings we report should be viewed as a deep exploration of our sample's beliefs and attitudes, but not as generalizing to those populations as a whole. Third, our choice of scenarios as well as our choice to use the term "algorithmic discrimination," while appropriate given our focus, may have influenced participants and other framings of fairness may have yielded different results. Finally, because we touch on socioeconomic status and ethnicity in this work, we include the detail that the research team consisted only of college-educated, European-American researchers. We describe participants' experiences in their own words, but our interpretations may lack context or nuance that may have been more readily available to a more diverse research team.

FINDINGS

In this section, we describe the main findings that emerged from our analysis.

Unfamiliar, but not Unfathomable

Most participants were not aware of the concept of algorithmic (un)fairness before participating in the study, although once it was described a few reported that they had had personal experiences with it or had heard about it in the media. However, most participants reported extensive experience with discrimination in their daily lives, and they connected their personal stories to the concept of algorithmic (un)fairness.

Personal Experiences with Discrimination

Most participants reported extensive negative experience with discrimination and stereotyping. Unfair treatment or racial profiling by law enforcement was commonly raised, for example, some participants described experiences with "driving while Black" (being pulled over by police because of their race, particularly when driving in affluent neighborhoods with few Black residents) [53]. Participants also raised a number of issues related to social and environmental justice, such as "white privilege" (societal advantages conferred on Caucasians), gentrification forcing people with low incomes out of their homes, food deserts

[2] Inspired by [20], which reported an experiment in which simulated men visiting the Times of India website were more likely than simulated women to see an ad for a career coaching service for $200K+ executive positions.

[3] Inspired by [62].

[4] Inspired by [73].

(lack of access to grocery stores and healthy food in impoverished areas), and the proximity of low income neighborhoods to pollution and environmental hazards. Participants also shared a number of other experiences, such as "shopping while Black" (receiving poor service in retail establishments, or being followed or monitored by staff who suspect they may steal) [36], being targeted by direct mail (unsolicited advertisements sent by physical mail) for predatory lending and other disadvantageous opportunities, being stereotyped as "angry" because they are Black, or employment-related discrimination. Many viewed these as pervasive issues that framed their opportunities and daily experiences, often from a young age.

> "My mother was taking us to daycare. And I remember her getting pulled over in [city] and the police officer arresting her, taking her to jail. Me and my sister had to go to a place where there were other children our age. At the time, we were scared. We didn't know why she was actually in handcuffs. We stayed there all day, and it was because the car was behind in registration… I wasn't even in elementary school yet. We were going to preschool. And it was quite traumatizing and I do believe that it was because she was an African American in [city]. So you learn the roles that you have or what could possibly happen at a very young age. So, some things now are just anticipated. They're not even shocking anymore." — P43[5]

> "I tell my daughter that, 'when you were eight months, in your mom's womb, you were already [racially] profiled [in a traffic stop]'." — P20

> "They're following me around the grocery store like I'm going to steal something." — P11

> "There was a lot of environmental racism in the neighborhood that I grew up in. It was very impoverished. Lots of police brutality... It's just set up that way for us to fail." — P11

Prior Awareness of Algorithmic Unfairness

Once algorithmic unfairness was described to them, a few participants reported that they were aware of times they had experienced it (naturally, participants may also have experienced it and not been aware of it), and a few other participants said they were familiar with the concept from the media. For example, a small number of participants raised concerns about having been targeted for low income ads, and a few discussed turning off location history to avoid racial profiling and "racially motivated advertising." A couple of participants also discussed experiences with computer systems making unfair job and scholarship decisions. Several participants also described stories they had heard about in the press regarding companies such as Airbnb, Facebook, Google, NextDoor, and others.

> "I'm constantly bombarded with 'You can get this low income credit card.' 'You can get this low finance loan.' I didn't ask for no loan. I didn't ask for no credit card… Plus it's a low income loan. It's not like 'Would you like to buy a house?' 'Would you like to buy a boat?' 'Would you like to finance a car?' No. Why can't I have like a Capital One or Discovery or American Express? No, they've already labeled me as the low income person." — P43

> P28: They had to hire Eric Holder to tamp down all the racism of [Airbnb].
> …
> Facilitator: So, what do you think Airbnb should do?
> P28: (laughs)
> P29: Well, something was already done. An African American man creating—
> P28: The Attorney General of the United States. They had to hire the former Attorney General, the biggest lawyer in the United States, to handle the racism of Airbnb.

Reactions to Algorithmic Unfairness

Even though most participants had not been aware of algorithmic unfairness prior to the study, learning about it elicited strong negative feelings, evoking experiences with discrimination in other settings. For example, participants drew connections between algorithmic unfairness and national dialogues about racial injustice and economic inequality, as well as lost opportunities for personal advancement.

> "If I would have searched and those things popped up, I would have been very angry. In fact it makes me angry right now just looking at it. Because what should be is that if somebody wants to know if he was a thug they have to type in, 'was he a thug'. Not have it be suggested to them. Because for people like me who feel like the police are taking advantage of getting away with killing brown and black people all over the country, it's infuriating. So what they should do is no matter what other people have typed in before, when someone types it in, it should show up as certain facts, no adjectives, no judgments, no positive or negative connotations. Just whatever happened that has been factually reported." — P23

> "[To] have your destiny, or your destination in life, based on mathematics or something that you don't put in for yourself… to have everything that you worked and planned for based on something that's totally out of your control, it seems a little harsh. Because it's like, this is what you're sent to do, and because of a algorithm, it sets you back from doing just that. It's not fair." — P04

Participants also drew connections with personal stories and life experiences. For example, they objected heavily to stereotyping, such as negative online characterizations of marginalized groups, or online ads or information being personalized based on demographic characteristics (similar to concerns raised in [67,86]). Similarly, they also felt it was very unfair to personalize ads or information based on the online behavior of other people with similar characteristics. While at first glance this may appear to contrast with Plane et al.'s finding that online behavioral advertising was seen as significantly less problematic than explicit demographic targeting [67], it seems likely that participants' underlying concern in both cases relates to the use of demographic characteristics or other sensitive traits to personalize information.

> P34: It's totally unfair—
> P33: —because not every woman's the same.

> "It's not accurate if you're just basing it on a group." — P22

> "They didn't even base it [what was shown to me] on what I've done in the past, they're just basing it on what they think I am." — P23

[5] For ease of reading, we have followed editing conventions consistent with applied social science research practices as described in [16]. Specifically, we edited quotes to remove content such as filler words and false starts, and in some cases we re-punctuated. We use ellipses to indicate substantial omissions.

Some participants oriented to algorithmic unfairness as a modern incarnation of familiar forms of discrimination, an unwelcome extension of offline discrimination into the online arena.

> "It's setup for not everyone [to win]… Since the beginning of civilization there's always been a hierarchy… technology is just another wheel in that." — P37

> "It seems like in technology, it's fascinating, but at the same time it's alarming because it seems like in every phase…people have taken it and have always done something wicked with it." — P30

> "[Because it's algorithmic] there is some type of system to it. Which means that there is some type of work being put into this certain type of discrimination... that it's actually people in the world that want it to be that way. And it's like, why? … I just don't understand why we have to live under these type of circumstances." — P04

> P12: We deal with this just walking down the street—
> P14: On a daily basis.
> P12: —on a daily basis. We don't need this on our internet, on our sites that we trust the most. We don't need to see the negative connotation come up every time. We have to walk out of our house and wonder if we're going to make it back in, and when we're safe in our homes we need to feel safe…especially if it comes from Google, or a site that we trust.
> P11: Um-hm. You have to draw the line somewhere… When we get home we've already dealt with it all day at work, at school, and it's like I want to come home and I don't want to have to deal with this, too… When I get on the computer…I shouldn't have to be subjected to racial stereotypes.

Although parallels to other life experiences may have driven initial negative responses, participants shared nuanced and pragmatic perspectives as the workshops unfolded, showing an appreciation for the complexity of this topic as they discussed it.

Scale and Impact of Algorithmic Systems

Though a small number of participants expressed a belief that large-scale algorithmic systems underlie many aspects of modern society, many participants viewed algorithmic systems as small in scope and low in both complexity and impact. This was especially apparent in the solutions that many participants proposed to scenarios of algorithmic unfairness, which often emphasized manual work by the end user or employees of technology companies, echoing the types of manual work envisioned by participants in [86]. For example, some participants proposed that filtering or recommendation processes could be made more fair by removing algorithmic processing and allowing the end user to go through the content themselves. Most participants tended to favor and trust human decision-making over algorithmic decision-making (this appears to contrast with Plane et al.'s results [67], which could be due to a variety of factors such as the different populations studied, and bears further investigation).

> "The algorithm is not a person. It's just a mathematical equation. It just has information. Then somebody chooses that information in a certain way and does with it whatever. That could mean choosing whether to use you in a job or where to put the next K-Mart… It's not making human decisions." — P39

> "I think it should stick with suggestions. I mean, what happens if the computer makes a bad decision? Does it just suggest…or is it going to be the final decision maker? … It's all good so that it can help categorize it, suggest. But to be the main decision maker, that would be scary." — P05

Further, for the most part, participants interpreted small percentage biases of algorithmic decisions as low-impact, and indicating natural imperfection rather than subtle bias. While researchers have argued that small statistical differences can have significant cumulative effects on individuals and/or groups, thereby perpetuating or increasing inequality [41], participants appeared to interpret small statistical disparities as benign, largely considering them to be natural, inevitable, and impossible to fix.

> "It sounds fine to me… I don't expect perfection, of course." — P43

High Salience of Representational Consequences

While participants may not have always come in with a previous notion of the wide-reaching implications of the underlying algorithmic systems, they did care deeply about the visible results of these systems and how marginalized groups were portrayed online. Participants were aware of and concerned about skewed representations and negative stereotypes, for example, online sexualization of women or offensive language about particular ethnic groups. Such offenses connected to a broader system of microaggressions [81] and personal stories from their own lives.

> P29: If you type in 'two Black teenagers,' you will see all mugshots of Black boys. But with White teenagers, you will see them playing basketball, boy scout.
> ...
> P28: You have negative connotations for the word black and positive connotations for the word white. That's just the way it is.

> "I'm just really not happy with the way that these words are put out there, these ideas." — P24

> "To see the things that they said [criminalizing] that little boy, that just broke my heart... He didn't do nothing to deserve that, and the fact that that's what society thinks of him, that's not just something that the computer put out there... I got sisters, I got little cousins, little nieces and nephews… they could look that up and see that. That's not right. That is not right at all...that's just sickening. Because that's a whole bunch of human beings that really typed that in … if I had any type of way to filter stuff like that, I would, because that's not cool. I would just erase it all." — P04

Participants were especially concerned with how children might be affected by negative representations.

> "There's lots of images that society already tells young, Black boys, or boys of color, that they're thugs; that they're gangster; and this and that. I wouldn't want my son to look up this teenage boy's name, and those type of images or associations comes up behind his name because my son is a young, Black boy… I don't think people should be stereotyped. And I don't want my son to think that society—even though it's the truth—society does label you because you're a young, Black boy." — P11

Participants also felt that popularity algorithms are not benign mirrors of the world, pointing out that social media can amplify societal biases and increase the reach of stereotyping messages.

> "I was just talking to my girlfriend about this last night. It's ridiculous how every time you click on Facebook or turn on news, radio station, or just the internet in general, there's some type of discrimination

> going on… and the main reason why it's gotten this big is because social media is in the middle of it all…" — P04

> "Feeding into that stuff, to me, is going backwards. Even encouraging people to read about that stuff and feeding into those thoughts, there's no need to feed." — P22

Accountability

Participants proposed a number of different parties might be responsible for algorithmic unfairness, and sometimes had differing opinions about the likely underlying cause of unfairness. Three of the most commonly proposed causes were: (1) a non-diverse population of programmers; (2) prejudiced online behavior by members of society; and (3) the news media. While a number of these ideas suggest an understanding of algorithmic fairness that goes beyond the technical, it is worth noting that many potential causes commonly raised in technical circles, such as lack of diverse training data or inequitable accuracy in classifying members of different categories [44], were raised rarely or not at all.

Many participants held the programmer accountable for an algorithm's discrimination, not necessarily because they thought programmers were ill-intended, but rather because their perception was that programmers are predominantly privileged white males who do not understand the perspective of more diverse users. They felt more diverse hiring practices would help.

> "People create the technology to do these things, so that's why I say it stems from the writer." — P29

> "When you lack that diversity, they may not be able to input certain things into that equation...because they don't know that reality…because the people that are writing these apps are probably not from our community... You need to be more selective, diverse or whatever in who you're hiring." — P20

> Facilitator: Does anybody else have any thoughts about who's writing algorithms?
> P24: I think it's kind of assumed that it is white males.
> P17: Ivy League people.
> P21: (laughs) I was going to say rich white men.
> ...
> P24: I mean who else? (laughs)
> P21: Does that make us racist when we say that?

Participants also often thought that much of the stereotyping or racism was coming from outside of technology companies, frequently calling out the role society played in creating the problem. Some participants also emphasized that the news media is a source of bias.

> "It's not really like a company being racist… it's really just a machine, it's stats... It's counting numbers, it's counting what we are all looking at. It's based on what we're looking at, not what Google wants you to look at… The problem is us, and what we have in our minds, so we can't really turn around and be like, 'oh, Google did it.'" — P02

> P06: I hear what you're saying, and I'm totally against everything that's going on, but the only reason it's so popular is because everybody's clicking on it, and people are making it popular… people have put that in there. Doesn't mean it's true…
> P02: Yeah, the problem's not really the search engine, it's the people searching. I wouldn't blame Google or anything because…it's just… going on clicks. The machine's not deciding whether it's right or wrong. People are entitled to their opinions… I guess that's their way of going online and free speeching too. Whether it's right or wrong, the search engine's not at fault. It's humanity… I wouldn't blame a company for that.

Even when they believed that the cause was external, most still saw technology companies as having some responsibility and a role to play in addressing the issue (this is consistent with and extends Plane et al.'s finding that many participants held both the advertiser and the ad network responsible, regardless of which was explicitly named as the perpetrator [67]). Further, they believed that companies could readily resolve many of the problems if they chose to do so.

> "I think that people that work for these companies…they can make the change tonight if they wanted to. It's just a matter of how are they going to meticulously put everything so it will still benefit them in some aspect." — P29

Occasionally, in specific contexts, some participants indicated that they did not feel companies could or should take action. The most prevalent arguments for inaction were: freedom of expression; concern about censoring content from credible news sources; a belief that a user is personally responsible for making good choices in their online activity, in order to shape what they see; or a belief that there was not a feasible technological solution.

> "As a company like Google, you'd have to respect the free speech. What could you do? It would be a very difficult decision for me to have to make." — P44

> "Sometimes that's what people want to see. You kind of got to give them what they want to see, unfortunately. It's scary." — P24

> "Unless Google owns the news companies, I think it's kind of out of their hands." — P37

> "I don't know who's going to really go and actually keep up with each controversial racial issue that comes up… How would you regulate? How would you know that these things would eventually come up? You just check every damn time something happened. You just kind of look and you kind of monitor? I don't even know if that's actually feasible." — P43

However, these positions were less common, tended to arise for fairly specific situations, and were often in opposition to much more commonly expressed positions that companies can and should act to reduce unfairness.

Curation

As mentioned in the previous section, participants expressed certain expectations of companies, regardless of the source of unfairness. In this section, we discuss the most prominent themes regarding expectations: a curatorial position on representation and the voice of the company.

Journalistic Standards

Participants tended to hold technology companies such as search engines to journalistic standards. For instance, they expected them to perform careful, manual fact checking (although resonating with the findings above regarding underestimation of scale, participants tended to propose manual, human-scale approaches), and show proven facts rather than opinions or biased content. Some participants

indicated the news media do not always meet this standard but rather sometimes shows harmful biased representations of marginalized populations, and some felt that technology companies could compensate for this.

> P10: I would only allow what is a actual fact. I don't need to know your cousin, your momma, said this that and the other, just include—
> P07: The truth.
> P10: —the facts.

> "The media responsibility. Google has that responsibility." — P28

> "I just need the news on it… It makes you upset when you see that all the time about any person pretty much that has been in the news for being brutalized or killed…I would prefer for it to be just official news…I would like to try to explore on my own, make my own opinion. But it seems like my opinion is already kind of being made before I can even search for answers." — P43

> "I think it's their responsibility to not do that. They don't have to report it like that, just because the news reports it like that." — P12

On a related note, many participants suggested that a predictive search feature should not suggest negative information for individuals, particularly minors. A few also suggested that negative information should be counterbalanced with positive information so the reader could learn about both sides of an argument and reach their own conclusions.

Voice of the Company

Participant responses suggest that in-product information processed by algorithms can give the impression that a company generated or endorses a message. For example, predictive search actively suggests content within the user interface, and some participants felt this gave the appearance that the content originated with the company that produced the feature. Participants also felt that the feature could make it too easy for users to find such content, or even encourage searching for it, and suggested that users should have to generate the negative searches themselves.

> "I feel like encouraging this type of searching is just toxic." — P22

> "If there were any negative connotations then it wouldn't pop up at all, so if you wanted to see something negative, you would have to spell it out." — P08

> "I would clear off all the negative...and just let them actually type in what they wanted to know about the person. Instead of offering things." — P38

Inaction posed the risk of appearing to endorse others' discrimination by signal boosting it.

> "You guys [Facebook] are pretty much promoting this hate and promoting this deceit... That's not doing nothing but making everybody mad." — P04

Impact on User Trust

As illustrated in the preceding sections, algorithmic fairness connects to strong emotions and in many cases participants have high expectations of how companies will ensure fairness in their products. Consistent with the philosophy of relationship marketing [59], participants linked algorithmic fairness to their relationships with companies, expressing feelings of betrayal, disappointment, or anger when companies they trusted surfaced societal bias or prejudice.

> "I've used Google a lot, it's been my lifeline almost… Maybe that's why I'm even more offended... It's like, come on, Google. I thought we were better than that." — P24

> "When I go on Google, I like the company and I expect great things from them, and I expect facts and I expect not to see stuff like that and don't want my child to see it because it's such a great company." — P12

However, when participants perceived companies were protecting them from unfairness or discrimination, it greatly enhanced user trust and strengthened their relationships with those companies.

> "I think that it's a very good decision that Google decided to stop running tobacco ads and stop doing the payday loans[6] because it lets me know that as a consumer…they are taking my feelings into consideration... I tell my son to search Google all the time and so now I feel more confident I may not have to watch over his shoulder… Very good. I'm very pleased." — P43

DISCUSSION

As human-computer interaction researchers, we often make arguments to stakeholders about how and why they can change technology to better serve users and/or improve society. In the case of algorithmic fairness, stakeholders such as regulators, lawmakers, the press, industry practitioners, and many others have the opportunity to take positive action. Technology companies in particular have tremendous leverage to improve algorithmic fairness because they are immediately proximate to many of the technical issues that arise, and they are uniquely positioned to diagnose and develop effective solutions to complex problems that would be difficult for outsiders to address. Accordingly, while we hope it is apparent that our findings can be directly leveraged by a wide variety of stakeholders, especially for decisions relating to product categories such as social media and search engines, we focus here on three best practices that our findings suggest apply to companies across the technology sector.

#1: Include fairness as a value in product design and development. Similar to considerations such as privacy, fairness can be included as a consideration throughout the product life cycle. Many positive steps can be taken, such as ensuring diverse training data for machine learning models, ensuring that designers are aware of inequalities in their systems so they can consider appropriate action [15,49], and including diverse populations in user testing.

In support of this point, our participants cared about fairness, had strong ethical expectations of companies, were disappointed when companies did not act (regardless of the source of the unfairness), and greatly valued efforts on the part of companies to ameliorate societal bias and make their products as inclusive as possible. Therefore, it is likely that measureable gains in user trust and engagement can result

[6] Earlier in the interview, we told the participant that Google had established a policy that banned ads for payday loans [40,45].

from incorporating algorithmic fairness in product design. Our findings suggest this is an opportune time for companies to act proactively, while public perception of this complex topic is still evolving. Algorithmic fairness issues are challenging both technically and organizationally and can take a long time to address, particularly if mechanisms are not already in place, so it is strategically wise to take positive steps before additional pressures apply. Due to the complexity of these issues, it is also wise to proceed thoughtfully with user research and to engage stakeholders to represent diverse perspectives. We discuss these in turn in the next two points.

#2: Design user studies that accommodate diverse perspectives, and include members of traditionally marginalized populations in user testing. The workshop format supported and encouraged participants' exploration and development of diverse, nuanced, and at times conflicting positions, and participants reported that it was empowering to take the perspective of a decision-maker at a technology company. At the same time, our experience reflects both the value and challenges of user research on complex computational topics. Complementing other work, our findings suggest that participants' opinions on this topic were highly contextual, often varying in response to situational factors (e.g. specific details of given scenarios), individual factors (which appears to resonate with variation reported in [69,86]), different stakeholder perspectives (as discussed for example in [51,52]), and different framings of fairness (for example, an emphasis on fair division as in [51,52] versus social justice). This contextual nature may help explain why research on this topic yields results that may sometimes appear inconsistent; for example, while many of our findings are broadly consistent with Plane et al. (e.g. objections to personalization based on demographic characteristics, and the expectation that technology companies play a role in addressing issues caused by external forces), our findings differed in other regards such as the fact that our participants appeared to favor and trust human decision-making over algorithmic decision-making. Additional research could yield further insights that account for such variation. Relatedly, we caution that decontextualized user research on this topic may yield misleading results. We recommend that researchers prepare and account for the beliefs and knowledge that participants may bring to the research environment, in order to provide an inclusive research environment for all participants. In some situations it will also be valuable to use ethnographic approaches to explore participants' underlying values and extrapolate from those values to technological implications (see [26] for additional discussion of the nature of analytic knowledge that can be gained in ethnographic studies).

#3: Engage with community groups and advocates to collaboratively develop solutions. As is common with wicked problems, stakeholders should not work in isolation to address the complex issues posed by algorithmic fairness [72]. A robust understanding of the goals and the best path forward will result from strong participation of multiple players, a point reinforced by Lee et al.'s argument that algorithmic service design support multiple stakeholder perspectives [52]. For example, companies can partner with community groups and community leaders to address particular challenges, as Airbnb did when addressing racism on its platform [2,60], as Facebook did when addressing concerns about ethnic affinity marketing [29], and as Google did when developing its policy about payday lending ads [40,45]. Our research underscores the importance of such efforts, since it shows that traditional methods of user testing may not yield a complete picture of different groups' perspectives on this computationally and socially complex issue. Community groups and leaders are experienced in considering societal-scale consequences and representing their constituencies on a range of issues, and are well-positioned to contribute to such discussions.

CONCLUSIONS AND FUTURE WORK

One way to make social change is to bolster pragmatic arguments for corporations to do good, by demonstrating that societally positive actions are also good business practice. Consider for example how Green to Gold effectively argued that sustainable business practices not only benefit the environment but can yield significant financial profit [32]. In this paper, we presented a novel exploration of how traditionally marginalized populations perceive algorithmic fairness. While our findings can inform a range of stakeholders, we highlight the insight that company handling of algorithmic fairness interacts significantly with user trust. We hope this insight may provide additional motivation for companies across the technology sector to actively pursue algorithmic fairness.

Future work could fruitfully explore these findings with a broader population, noting that Plane et al.'s study offers evidence that at least some of these issues may resonate widely [67]. We also suggest further exploring concrete actions that companies can take regarding algorithmic fairness, such as making specific improvements to product experiences, to build and maintain user trust. Finally, we suggest further research on how stakeholders across the ecosystem can work collectively to leverage their different perspectives and skills to pursue algorithmic fairness.

ACKNOWLEDGMENTS

We thank the following for their thoughtful comments and contributions to this work: Paul Aoki, Ed Chi, Charina Choi, Mark Chow, Rena Coen, Sunny Consolvo, Jen Gennai, Lea Kissner, Brad Krueger, Ali Lange, Irene Tang, Lynette Webb, Jill Woelfer, and the anonymous reviewers.

REFERENCES

1. ACM US Public Policy Council. 2017. Statement on Algorithmic Transparency and Accountability. Retrieved September 16, 2017 from https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

2. Airbnb, Inc. Airbnb's Nondiscrimination Policy. Retrieved September 14, 2017 from https://www.airbnb.com/terms/nondiscrimination_policy
3. Mariam Asad, Sarah Fox, and Christopher A. Le Dantec. 2014. Speculative Activist Technologies. *Proceedings iConference 2014*. https://doi.org/10.9776/14074
4. Paul Baker and Amanda Potts. 2013. 'Why do white people have thin lips?' Google and the perpetuation of stereotypes via auto-complete search forms. *Critical Discourse Studies* 10, 2: 187–204. https://doi.org/10.1080/17405904.2012.744320
5. Solon Barocas and Andrew D. Selbst. 2016. Big Data's Disparate Impact. *California Law Review* 104, 3: 671–732.
6. David Beer. 2009. Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society* 11, 6: 985–1002. https://doi.org/10.1177/1461444809336551
7. Hugh Beyer and Karen Holtzblatt. 1998. *Contextual Design: Defining Customer-centered Systems*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
8. danah boyd and Kate Crawford. 2012. Critical Questions for Big Data. *Information, Communication & Society* 15, 5: 662–679. https://doi.org/10.1080/1369118X.2012.678878
9. danah boyd, Karen Levy, and Alice Marwick. 2014. The Networked Nature of Algorithmic Discrimination. In *Data and Discrimination: Collected Essays*, Seeta Peña Gangadharan, Virginia Eubanks and Solon Barocas (eds.). Open Technology Institute, New America Foundation, Washington, D.C., 53–57.
10. Engin Bozdag. 2013. Bias in algorithmic filtering and personalization. *Ethics and Information Technology* 15, 3: 209–227. https://doi.org/10.1007/s10676-013-9321-6
11. Taina Bucher. 2017. The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society* 20, 1: 30–44.
12. Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3, 1: 1–12. https://doi.org/10.1177/2053951715622512
13. Kathleen Chaykowski. 2016. Facebook To Ban "Ethnic Affinity" Targeting For Housing, Employment, Credit-Related Ads. *Forbes*.
14. Danielle Keats Citron and Frank Pasquale. 2014. The Scored Society: Due Process for Automated Predictions. *Washington Law Review* 89, 1.
15. Ashley Colley, Jacob Thebault-Spieker, Allen Yilun Lin, Donald Degraen, Benjamin Fischman, Jonna Häkkilä, Kate Kuehl, Valentina Nisi, Nuno Jardim Nunes, Nina Wenig, Dirk Wenig, Brent Hecht, and Johannes Schöning. 2017. The Geography of PokéMon GO: Beneficial and Problematic Effects on Places and Movement. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 1179–1192. https://doi.org/10.1145/3025453.3025495
16. Anne Corden and Roy Sainsbury. 2006. *Using Verbatim Quotations in Reporting Qualitative Social Research*. University of York, York, UK.
17. Tressie McMillan Cottom. 2015. Credit Scores, Life Chances, and Algorithms. Retrieved September 15, 2017 from https://tressiemc.com/uncategorized/credit-scores-life-chances-and-algorithms/
18. Kate Crawford. 2014. The Anxieties of Big Data. *The New Inquiry*.
19. Kate Crawford. 2016. Artificial Intelligence's White Guy Problem. *The New York Times*.
20. Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated Experiments on Ad Privacy Settings. In *Proceedings on Privacy Enhancing Technologies* (PETS 2015), 92–112. https://doi.org/10.1515/popets-2015-0007
21. Michael A. DeVito, Jeremy Birnholtz, and Jeffery T. Hancock. Platforms, People, and Perception: Using Affordances to Understand Self-Presentation on Social Media. In *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '17), 740–754.
22. Nicholas Diakopoulos. 2015. Algorithmic Accountability. *Digital Journalism* 3, 3: 398–415. https://doi.org/10.1080/21670811.2014.976411
23. Carl DiSalvo, Thomas Lodato, Laura Fries, Beth Schechter, and Thomas Barnwell. 2011. The collective articulation of issues as design practice. *CoDesign* 7, 3–4: 185–197. https://doi.org/10.1080/15710882.2011.630475
24. Carl DiSalvo, Illah Nourbakhsh, David Holstius, Ayça Akin, and Marti Louw. 2008. The Neighborhood Networks Project: A Case Study of Critical Engagement and Creative Expression Through Participatory Design. In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008* (PDC '08), 41–50.
25. Carl DiSalvo, Phoebe Sengers, and Hrönn Brynjarsdóttir. 2010. Mapping the Landscape of Sustainable HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 1975–1984. https://doi.org/10.1145/1753326.1753625
26. Paul Dourish. 2006. Implications for Design. In *Proceedings of the SIGCHI Conference on Human*

Factors in Computing Systems (CHI '06), 541–550. https://doi.org/10.1145/1124772.1124855

27. Benjamin Edelman and Michael Luca. 2014. Digital Discrimination: The Case of Airbnb.com. *Harvard Business School Working Paper* 14-054.
28. Benjamin Edelman, Michael Luca, and Daniel Svirsky. 2017. Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics* 9, 2: 1–22.
29. Erin Egan. 2016. Improving Enforcement and Promoting Diversity: Updates to Ethnic Affinity Marketing. *Facebook Newsroom Blog*. Retrieved September 14, 2017 from https://newsroom.fb.com/news/2016/11/updates-to-ethnic-affinity-marketing/
30. Pelle Ehn. 1990. *Work-Oriented Design of Computer Artifacts*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA.
31. Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I Always Assumed That I Wasn't Really That Close to [Her]": Reasoning About Invisible Algorithms in News Feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 153–162. https://doi.org/10.1145/2702123.2702556
32. Daniel C. Esty and Andrew Winston. 2006. *Green to Gold: How Smart Companies Use Environmental Strategy to Innovate, Create Value, and Build Competitive Advantage*. Yale University Press.
33. Executive Office of the President. 2016. *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*.
34. Federal Trade Commission. 2016. *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues*.
35. Batya Friedman and Helen Nissenbaum. 1996. Bias in Computer Systems. *ACM Transactions on Information Systems* 14, 3: 330–347. https://doi.org/10.1145/230538.230561
36. Shaun L. Gabbidon. 2003. Racial Profiling by Store Clerks and Security Personnel in Retail Establishments: An Exploration of "Shopping While Black." *Journal of Contemporary Criminal Justice* 19, 3: 345–364. https://doi.org/10.1177/1043986203254531
37. Marilyn Geewax. 2015. The Tipping Point: Most Americans No Longer Are Middle Class. *NPR*.
38. Tarleton Gillespie. 2014. The relevance of algorithms. In *Media Technologies: Essays on Communication, Materiality, and Society - University Press Scholarship*, Tarleton Gillespie, Pablo Boczkowski and Kirsten Foot (eds.). MIT Press.
39. Bryce Goodman and Seth Flaxman. 2016. European Union regulations on algorithmic decision-making and a "right to explanation." In *ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*.
40. David Graff. 2016. An update to our AdWords policy on lending products. *Google Public Policy Blog*. Retrieved September 15, 2017 from https://www.blog.google/topics/public-policy/an-update-to-our-adwords-policy-on/
41. Anthony G. Greenwald, Mahzarin R. Banaji, and Brian A. Nosek. 2015. Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology* 108, 4: 553–561. https://doi.org/10.1037/pspa0000016
42. Maya Gupta, Andrew Cotter, Jan Pfeifer, Konstantin Voevodski, Kevin Canini, Alexander Mangylov, Wojciech Moczydlowski, and Alexander van Esbroeck. 2016. Monotonic Calibrated Interpolated Look-Up Tables. *Journal of Machine Learning Research* 17, 109: 1–47.
43. Bernard E. Harcourt. 2007. *Against prediction: Profiling, policing, and punishing in an actuarial age*. University of Chicago Press.
44. Moritz Hardt, Eric Price, and Nathan Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems* (NIPS 2016), 3315–3323.
45. Shin Inouye. 2016. Advocates Applaud Google's Ban on Payday Loan Advertisements. *The Leadership Conference on Civil and Human Rights*. Retrieved September 15, 2017 from http://civilrights.org/advocates-applaud-googles-ban-on-payday-loan-advertisements/
46. Lucas D. Introna and Helen Nissenbaum. 2000. Shaping the Web: Why the Politics of Search Engines Matters. *The Information Society* 16, 3: 169–185. https://doi.org/10.1080/01972240050133634
47. Isaac Johnson, Connor McMahon, Johannes Schöning, and Brent Hecht. 2017. The Effect of Population and "Structural" Biases on Social Media-based Algorithms: A Case Study in Geolocation Inference Across the Urban-Rural Spectrum. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 1167–1178. https://doi.org/10.1145/3025453.3026015
48. Robert Jungk, Norbert Müllert, and Institute for Social Inventions. 1987. *Future workshops: how to create desirable futures*. Institute for Social Inventions, London.
49. Matthew Kay, Cynthia Matuszek, and Sean A. Munson. 2015. 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 (CHI '15), 3819–3828. https://doi.org/10.1145/2702123.2702520

50. Jonathan Lazar, Julio Abascal, Simone Barbosa, Jeremy Barksdale, Batya Friedman, Jens Grossklags, Jan Gulliksen, Jeff Johnson, Tom McEwan, Loïc Martínez-Normand, Wibke Michalk, Janice Tsai, Gerrit van der Veer, Hans von Axelson, Ake Walldius, Gill Whitney, Marco Winckler, Volker Wulf, Elizabeth F. Churchill, Lorrie Cranor, Janet Davis, Alan Hedge, Harry Hochheiser, Juan Pablo Hourcade, Clayton Lewis, Lisa Nathan, Fabio Paterno, Blake Reid, Whitney Quesenbery, Ted Selker, and Brian Wentz. 2016. Human–Computer Interaction and International Public Policymaking: A Framework for Understanding and Taking Future Actions. *Foundations and Trends in Human-Computer Interaction* 9, 2: 69–149. https://doi.org/10.1561/1100000062
51. Min Kyung Lee and Su Baykal. 2017. Algorithmic Mediation in Group Decisions: Fairness Perceptions of Algorithmically Mediated vs. Discussion-Based Social Division. In *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '17), 1035–1048.
52. Min Kyung Lee, Ji Tae Kim, and Leah Lizarondo. 2017. A Human-Centered Approach to Algorithmic Services: Considerations for Fair and Motivating Smart Community Service Management that Allocates Donations to Non-Profit Organizations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 3365–3376.
53. Richard J. Lundman and Robert L. Kaufman. 2003. Driving While Black: Effects of Race, Ethnicity, and Gender on Citizen Self-Reports of Traffic Stops and Police Actions. *Criminology* 41, 1: 195–220. https://doi.org/10.1111/j.1745-9125.2003.tb00986.x
54. Caitlin Lustig and Bonnie Nardi. 2015. Algorithmic Authority: The Case of Bitcoin. In *48th Hawaii International Conference on System Sciences* (HICSS 2015), 743–752. https://doi.org/10.1109/HICSS.2015.95
55. Caitlin Lustig, Katie Pine, Bonnie Nardi, Lilly Irani, Min Kyung Lee, Dawn Nafus, and Christian Sandvig. 2016. Algorithmic Authority: The Ethics, Politics, and Economics of Algorithms That Interpret, Decide, and Manage. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '16), 1057–1062. https://doi.org/10.1145/2851581.2886426
56. Peggy McIntosh. 1990. White Privilege: Unpacking the Invisible Knapsack. *Independent School* 49, 4: 31–5.
57. Amanda Menking and Ingrid Erickson. 2015. The Heart Work of Wikipedia: Gendered, Emotional Labor in the World's Largest Online Encyclopedia. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 207–210. https://doi.org/10.1145/2702123.2702514
58. Claire Cain Miller. 2015. When Algorithms Discriminate. *The New York Times*.
59. Robert M. Morgan and Shelby D. Hunt. 1994. The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing* 58, 3: 20–38. https://doi.org/10.2307/1252308
60. Laura W. Murphy. 2016. *Airbnb's Work to Fight Discrimination and Build Inclusion: A Report Submitted to Airbnb*. Retrieved September 15, 2017 from http://blog.atairbnb.com/wp-content/uploads/2016/09/REPORT_Airbnbs-Work-to-Fight-Discrimination-and-Build-Inclusion.pdf?3c10be
61. Carey Nadeau and Amy K. Glasmeier. 2016. *Minimum Wage: Can an Individual or a Family Live on It?* Retrieved September 15, 2017 from http://livingwage.mit.edu/articles/15-minimum-wage-can-an-individual-or-a-family-live-on-it
62. Safiya Umoja Noble. 2014. Teaching Trayvon: Race, Media, and the Politics of Spectacle. *The Black Scholar* 44, 1: 12–29. https://doi.org/10.5816/blackscholar.44.1.0012
63. Cathy O'Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown, New York.
64. Jahna Otterbacher, Jo Bates, and Paul Clough. 2017. Competent Men and Warm Women: Gender Stereotypes and Backlash in Image Search Results. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 6620–6631. https://doi.org/10.1145/3025453.3025727
65. Sarah Perez. 2016. Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism. *TechCrunch*.
66. Sarah Pink. 2014. *Doing Visual Ethnography*. Sage Publications.
67. Angelisa Plane, Elissa Redmiles, Michelle Mazurek, and Michael Tschantz. 2017. Exploring User Perceptions of Discrimination in Online Targeted Advertising. In *Proceedings of the 2017 USENIX Security Symposium*.
68. Bernadette D. Proctor, Jessica L. Semega, and Melissa A. Kollar. 2016. *Income and Poverty in the United States: 2015*. The United States Census Bureau. Retrieved September 15, 2017 from https://www.census.gov/library/publications/2016/demo/p60-256.html
69. Emilee Rader and Rebecca Gray. 2015. Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 173–182. https://doi.org/10.1145/2702123.2702174

70. Matt Ratto. 2011. Critical Making: Conceptual and Material Studies in Technology and Social Life. *The Information Society* 27, 4: 252–260. https://doi.org/10.1080/01972243.2011.583819

71. Noopur Raval and Paul Dourish. 2016. Standing Out from the Crowd: Emotional Labor, Body Labor, and Temporal Labor in Ridesharing. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16), 97–107. https://doi.org/10.1145/2818048.2820026

72. Horst W. J. Rittel and Melvin M. Webber. 1973. Dilemmas in a general theory of planning. *Policy Sciences* 4, 2: 155–169. https://doi.org/10.1007/BF01405730

73. Rosemary Rodriguez. 2015. Discovery. *The Good Wife*.

74. Daniela K. Rosner, Saba Kawas, Wenqi Li, Nicole Tilly, and Yi-Chen Sung. 2016. Out of Time, Out of Place: Reflections on Design Workshops as a Research Method. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16), 1131–1141. https://doi.org/10.1145/2818048.2820021

75. Elizabeth B.-N. Sanders and Pieter Jan Stappers. 2008. Co-creation and the new landscapes of design. *CoDesign* 4, 1: 5–18. https://doi.org/10.1080/15710880701875068

76. Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2015. Can an Algorithm be Unethical? In *65th Annual Meeting of the International Communication Association*.

77. Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*: May 2014.

78. Sassafras Tech Collective. 2016. Icebreaker. In *Exploring Social Justice, Design, and HCI Workshop at CHI 2016*.

79. Clay Shirky. 2011. A Speculative Post on the Idea of Algorithmic Authority. Retrieved September 15, 2017 from http://www.shirky.com/weblog/2009/11/a-speculative-post-on-the-idea-of-algorithmic-authority/

80. Kiley Sobel, Katie O'Leary, and Julie A. Kientz. 2015. Maximizing Children's Opportunities with Inclusive Play: Considerations for Interactive Technology Design. In *Proceedings of the 14th International Conference on Interaction Design and Children* (IDC '15), 39–48. https://doi.org/10.1145/2771839.2771844

81. Derald Wing Sue. 2010. *Microaggressions in Everyday Life: Race, Gender, and Sexual Orientation*. Wiley.

82. Astra Taylor and Jathan Sadowski. 2015. How Companies Turn Your Facebook Activity Into a Credit Score. *The Nation*.

83. David R. Thomas. 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation* 27, 2: 237–246. https://doi.org/10.1177/1098214005283748

84. Vanessa Thomas, Christian Remy, Mike Hazas, and Oliver Bates. 2017. HCI and Environmental Public Policy: Opportunities for Engagement. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 6986–6992. https://doi.org/10.1145/3025453.3025579

85. Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security* (SOUPS '12), 4:1–4:15. https://doi.org/10.1145/2335356.2335362

86. Jeff Warshaw, Nina Taft, and Allison Woodruff. 2016. Intuitions, analytics, and killing ants: Inference literacy of high school-educated adults in the US. In *Proceedings of the Twelfth Symposium on Usable Privacy and Security* (SOUPS '16).

87. Brian Wynne. 1991. Knowledges in Context. *Science, Technology, & Human Values* 16, 1: 111–121.