Exploring User Perceptions of Discrimination in Online Targeted Advertising

Angelisa C. Plane, Elissa M. Redmiles, and Michelle L. Mazurek, *University of Maryland;*
Michael Carl Tschantz, *International Computer Science Institute*

https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/plane

This paper is included in the Proceedings of the
26th USENIX Security Symposium

August 16–18, 2017 • Vancouver, BC, Canada

ISBN 978-1-931971-40-9

Open access to the Proceedings of the
26th USENIX Security Symposium
is sponsored by USENIX

Exploring User Perceptions of Discrimination in Online Targeted Advertising

Angelisa C. Plane
University of Maryland

Elissa M. Redmiles
University of Maryland

Michelle L. Mazurek
University of Maryland

Michael Carl Tschantz
International Computer Science Institute

Abstract

Targeted online advertising now accounts for the largest share of the advertising market, beating out both TV and print ads. While targeted advertising can improve users' online shopping experiences, it can also have negative effects. A plethora of recent work has found evidence that in some cases, ads may be discriminatory, leading certain groups of users to see better offers (e.g., job ads) based on personal characteristics such as gender. To develop policies around advertising and guide advertisers in making ethical decisions, one thing we must better understand is what concerns users and why. In an effort to answer this question, we conducted a pilot study and a multi-step main survey (n=2,086 in total) presenting users with different discriminatory advertising scenarios. We find that overall, 44% of respondents were moderately or very concerned by the scenarios we presented. Respondents found the scenarios significantly more problematic when discrimination took place as a result of explicit demographic targeting rather than in response to online behavior. However, our respondents' opinions did not vary based on whether a human or an algorithm was responsible for the discrimination. These findings suggest that future policy documents should explicitly address discrimination in targeted advertising, no matter its origin, as a significant user concern, and that corporate responses that blame the algorithmic nature of the ad ecosystem may not be helpful for addressing public concerns.

1 Introduction

Online advertising revenue is projected to reach $83 billion in 2017, an increase of $20 billion and 40% since 2015 [27, 36]. It has surpassed T.V. and print advertising, accounting for 37% of the media market share [2]. The growth of online advertising can be attributed both to growth in digital users and the ability to do unprecedentedly specific targeting of ads: individually customizing advertisements to users. Targeted advertising is often driven by inferencing: the process of using collected information about a user's digital habits to infer beliefs about her demographics and preferences [8]. Targeted advertising—also known as online behavioral advertising, or OBA—has a number of consumer benefits (e.g., seeing more interesting or relevant ads) [19, 34, 40] but it has also raised serious concerns [5, 6, 13, 21, 28, 29, 32, 37, 47], including threats to consumer privacy and the potential for discrimination.

Consumer privacy issues related to targeted advertising have received considerable attention from researchers, media, and government agencies for several years [5, 9, 14, 16, 18, 19, 22, 33, 34, 40, 44]. More recently, the issue of algorithmic discrimination in targeted advertising has begun to attract similar attention [5, 6, 13, 21, 28, 29, 32, 37, 47]. In one example, Datta et al. found that Google showed ads promoting certain high-paying jobs more frequently to men than women [13].

Consumer opinions about general privacy threats from targeted advertising have been fairly well documented [30, 33, 39, 40]. Recent work has also begun to examine how well users understand the process of inferencing [45] and how inference accuracy affects attitudes and perceptions [11, 38]. To the best of our knowledge, however, little to no investigation has focused on people's attitudes toward discriminatory practices that arise, possibly unintentionally, from inferencing and OBA.

We argue that better understanding of such attitudes is critical, because the instances of discrimination in targeted advertising touch on complicated legal and moral issues. While consumer preferences are far from the only important factor to consider, they do help us to understand the current landscape. Companies might use information about consumer attitudes to avoid particularly egregious mistakes that can lead to bad press and even lawsuits [22, 44]. Knowledge of people's attitudes can also aid advocates of algorithmic fairness in understanding how to focus their public awareness efforts. Finally, data about consumer attitudes may prove valuable to policymakers, who can take these attitudes—and resulting corporate incentives—into account (as two of many important factors) when developing a regulatory framework for this increasingly controversial ecosystem.

As a first step toward achieving this understanding, we conducted three surveys (two smaller pilots and then a main survey) comparing respondents' attitudes to different discriminatory advertising scenarios, with the aim of understanding which specific scenarios people find most problematic and why. In particular, we varied factors such as which player in the ecosystem was responsible, whether targeting decisions were made by an algorithm or a human, and whether the targeting was based explicitly on demographic factors or arose from behavioral factors. To ensure we encountered a range of attitudes, we recruited a broad array of respondents, both from Amazon's Mechanical Turk crowdsourcing site (MTurk) and from a web panel with quota sampling to closely match the demographics of the U.S. population. We used the two pilot surveys to develop a final set of questions and a candidate regression model, which we applied in our main survey.

In our main survey (n=891), a large portion (44%) of respondents viewed our scenarios of discrimination in targeted advertising as a moderate or severe problem. The severity of the problem, however, depended primarily on how the discrimination occurred—based on explicit targeting of demographic factors or behavioral inferencing—and who was discriminated against. Respondents tended to rate scenarios in which differences in behavioral patterns led to discriminatory effects as less problematic and more ethical than scenarios in which discrimination was explicitly based on demographics. To our surprise, however, whether a human or an algorithm made the targeting decision had no statistically significant impact on perceptions of problem severity or ethics. Responses on severity also did not appear to differ based on the entity responsible for the discrimination (e.g., the ad network or the advertiser), and many participants held both entities responsible, regardless of which was explicitly named as the perpetrator. Based on these results, we suggest implications for companies and policymakers and suggest future work to deepen understanding of attitudes toward discrimination in targeted advertising.

2 Related Work

We review related work in two key areas: empirically observed discrimination in online targeted advertising and end-user perceptions of inferencing and behavioral advertising.

2.1 Discrimination in Online Targeting

Since the inner workings of the ad process are opaque, most knowledge of behavioral advertising has been derived through black-box observation.

Researchers have designed tools that create profiles with specific attributes (e.g., age, gender) to scrape ads seen with this profile and compare to other profiles' ads, providing insight into how often targeted ads are displayed and which attributes influence targeting [6, 13, 28, 29, 32, 35, 47]. Mikians et al. found early evidence of price and search discrimination based on user characteristics [35]. In another measurement, up to 65% of the ads seen across the ad categories tested were targeted based on some behavioral or profile aspect, such as browsing patterns [32].

Some of the identified targeting can be considered discriminatory. In one of the earliest examples, Sweeney found that ads displayed during search were more likely to associate stereotypically African American names than stereotypically white names with claims about arrest records [37]. Carrascosa et al. [10] found that health and religion were used in assigning advertisements to consumers, even though this is prohibited by E.U. law and may be prohibited by U.S. law for certain advertisements [42]. Finally, using the AdFisher tool, Datta et al. determined that ads promoting the seeking of high-paying executive jobs were shown significantly more often to simulated men than women [13].

2.2 Perceptions of Inferencing and Behavioral Advertising

Significant research has explored users' perceptions of targeted advertising, including both their understanding of the process and their attitudes and opinions.

There are strong indications that the process of targeted advertising is poorly understood. McDonald and Cranor found in surveys and interviews that people did not understanding the mechanisms or frequency of tracking [34]. Ur et al. identified a mismatch between participants' mental models and actual OBA implementations [40]. Warshaw et al. interviewed high-school-only-educated adults and found that they did not understand or believe in strong behavioral inferencing; instead, participants believed that targeting decisions were based on stereotypes or on straightforward intuitions [45].

Reaction to behavioral advertising has been mixed, with some appreciation of potential benefits but also concern for potential harms. Ur et al. found that people informed about online behavioral advertising express interest in receiving more-relevant ads, but also strong concerns about data collection and privacy [40]. Similarly, Agarwal et al. found that people expressed interest in relevant ads but were concerned about personal or intimate advertisements being shown, particularly when other people might also see them [3]. Turow et al. found that many users are resigned to privacy violations, and therefore accept benefits such as discounts or relevant ads as some consolation for unavoidable tracking [39].

In a lab experiment, Malheiros et al. concluded that when ads were more personalized to the user they were more noticeable, but that the users also became less comfortable as the degree of personalization increased [33]. More recently, Coen et al. found that people were less concerned about inferencing when they believed the results were accurate [11]. Tschantz et al. found no statistically significant associations between profile accuracy and people's concern about tracking or confidence in avoiding it [38].

3 Overview of Studies

To examine peoples' perceptions of discriminatory advertising, we first performed an exploratory pilot study, Pilot 1 (Section 4), which looked broadly at a wide variety of possible discrimination situations, with the goal of identifying a smaller set of relevant constructs and relationships to further examine. In our main study, we used the resulting smaller set of questions in a two-step regression analysis. First, we conducted a second pilot study, Pilot 2 (Section 5.2), in order to collect training data. Using this data, we conducted an exploratory regression analysis and distilled a set of parsimonious models to evaluate. Finally, we collected a final larger data set to validate these models and generate our final results (Section 5.3).

The structure of the survey questions was similar in both Pilot 1 and the final survey. In each case, the participant was given a scenario about discrimination in targeted advertising, together with a brief *explanation* of how the discrimination occurred. In each case, the scenario consisted of a fictional technology company, Systemy, placing a job ad using the fictional ad network Bezo Media. The job ad, which in the scenario appeared on a local newspaper's website, was shown more frequently to people in some *target* group than to people in other groups. This scenario was loosely based on real-life findings from Datta et al. about discriminatory ads [13].

Explanations included information about how the decision to target a specific group was made: whether an algorithm or a human made the decision, which company in the scenario made the decision, and what behavioral or demographic cues led to the targeting decision.

The participant then answered Likert-scale questions about how responsible various entities (e.g., the advertiser, the ad network) were for the discrimination, whether each entity had acted ethically, and whether the overall situation constituted a problem. We deliberately asked the responsibility questions before the question about how problematic the scenario was, to avoid priming the responsibility answers with an assumption that the scenario was problematic. In addition, we asked the participant how believable they found the scenario they had read. We then asked respondents to optionally provide free-text feedback on the scenario. Finally, we collected standard demographic information, including age, gender, education level, and ethnicity. The full set of questions is shown in Appendix A. All surveys were deployed using the Qualtrics web survey tool.

All three studies were approved by the University of Maryland's Institutional Review Board (IRB).

4 Pilot 1: Evaluating a Broad Range of Discriminatory Factors

We designed the first pilot study to explore a broad range of factors that might prove important to respondents' perceptions of discrimination in targeted online advertising.

4.1 Scenarios

As described in Section 3, in our survey respondents were presented with a scenario describing an online targeted advertising situation that resulted in discrimination. They were then asked questions about their opinion of the scenario. Respondents in Pilot 1 were assigned randomly to one of 72 total scenarios. The scenarios varied along two axes. The first was the *target* of the discriminatory ads, that is, one of eight groups of people who saw the job ad more frequently. The second was the *explanation* for how the targeting came about. We drew the eight explanations we considered in part from suggested explanations posited by the authors of an ad-discrimination measurement study [12] with the intent to span a range of both real-life plausibility and discriminatory intent. We also used a ninth condition, in which no explanation was provided, as a control. The targets and explanations used in Pilot 1 are listed in Table 1.

Because we used racial, political, and health characteristics in the target sets, we included questions about race/ethnicity, political affiliation, and health status in the demographic portion of the survey.

4.2 Cognitive Interviews

We anticipated that the explanations of discriminatory targeting provided in our scenarios might be complex and unfamiliar to our respondents. As such, we carefully pretested the wording of our explanations and subsequent questions using *cognitive interviews*, a standard technique

Targets:

• Are/be over 30 years old	• Are/be under 30 years old
• Are/be a registered Democrat	• Are/be a registered Republican
• Are/be white	• Are/be Asian
• Have a pre-existing health condition	• Have no pre-existing health condition

Explanations:

- *No explanation given (control).*
- An HR employee at Systemy chooses to target individuals who [TARGET].
- An employee at Bezo Media chooses to target individuals who [TARGET].
- An advertising sales employee at the local news site chooses to target Systemy's ads to individuals who [TARGET].
- An HR employee at Systemy chooses to advertise on the local news site specifically because its readers are known to mostly [TARGET].
- Individuals who [TARGET] tend to click on different ads than [OPPOSITE OF TARGET]. Bezo Media's automated system has observed this difference and automatically assigns the Systemy ads to individuals who [TARGET].
- Systemy requests that this ad be shown to viewers who have recently visited technology-interest websites. People who [TARGET] tend to visit more technology-interest websites than individuals [OPPOSITE OF TARGET].
- Bezo Media charges less to reach individuals who [TARGET] than individuals who [OPPOSITE OF TARGET], and a Systemy marketing employee chooses the less expensive option.
- Bezo Media charges less to reach individuals who [TARGET] than individuals who [OPPOSITE OF TARGET], and Systemy's marketing computer program automatically selects the less expensive option.

Table 1: Scenarios for Pilot 1. Each respondent viewed one explanation, with one targeted group filled in as receiving more of the targeted ads.

Gender	Age	Race	Education
Female	52 yrs	Black	High School
Female	34 yrs	White	M.S.
Male	22 yrs	Black	B.S.
Female	22 yrs	White	B.S.
Female	20 yrs	Black	Some College
Female	39 yrs	Black	High School
Male	31 yrs	Black	High School
Male	44 yrs	White	B.S.

Table 2: Cognitive Interview Demographics

for evaluating the intelligibility and effectiveness of survey questions by asking respondents to think aloud while answering the survey questions [46]. We conducted eight in-person cognitive interviews with respondents from a variety of demographic groups (Table 2). As a result of these interviews, we made the scenario descriptions more narrative, clarified the wording of some questions, and added the question about believability.

4.3 Respondents

The targets and explanations in this pilot study were deliberately designed to cover a broad range of possible topics, to help us identify the most salient and relevant issues to explore further. As such, we wanted to ensure that we sampled from a broad range of respondents, so that issues important to different demographic groups would be potentially salient in our results. This goal seemed particularly critical in light of prior work suggesting that people with less educational attainment have important misconceptions about targeted advertising [45]. To achieve these broad demographics, we contracted Survey Sampling International (SSI) to obtain a near-census-representative sample.

In August and September of 2016, 988 respondents completed our Qualtrics questionnaire, which took on average four to five minutes. Respondents were paid according to their individual agreements with SSI; this compensation could include a donation to a charity of their choosing, frequent flier miles, a gift card, or a variety of other options. We paid SSI $3.00 per completion. The demographic makeup of the respondents was close to the U.S. population as seen in table 3, with slightly more educated individuals. Between 15 and 16 respondents were assigned to each of the 72 scenarios.

4.4 Results

We examined the results using exploratory statistics and data visualizations to identify themes of most interest.

Metric	SSI	Census
Male	47.6%	48.2%
Female	52.4%	51.8%
Caucasian	67.0%	64.0%
Hispanic	12.0%	16.0%
Asian	5.0%	5.4%
African American	13.1%	12.0%
Other	2.9%	2.6%
up to H.S.	18.5%	41.3%
Some college	40.0%	31.0%
B.S. or above	41.6%	27.7%
18–29 years	23.7%	20.9%
30–49 years	38.8%	34.7%
50–64 years	23.5%	26.0%
65+ years	14.1%	18.4%

Table 3: Respondent demographics, Pilot 1, compared to 2015 U.S. Census figures [41]

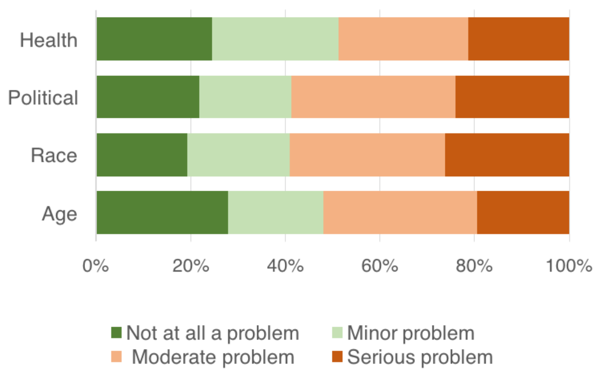

Figure 1: Problem severity, organized by target (Pilot 1).

One key goal was to develop a smaller set of issues to focus on in the follow-up studies.

First, we considered the issue of who was targeted in the scenario, that is, which group of people benefited from or was shortchanged by the discriminatory advertising. We found that the scenarios that targeted race were more likely to be considered problematic than the other targets that we considered: age, political affiliation, and health condition (see Figure 1). Opinions about which groups are targeted touch on a range of cultural and sociological issues that are not likely to be unique to online targeted advertising; as such, these opinions were not of primary interest to our research question, which mainly concerns how different explanations for discriminatory outcomes affect people's attitudes. Therefore, we decided to limit future scenarios to targeting race, in the interest of provoking more dramatic reactions that might allow us to identify interesting explanation-based differences.

Second, we considered respondents' responses regard-

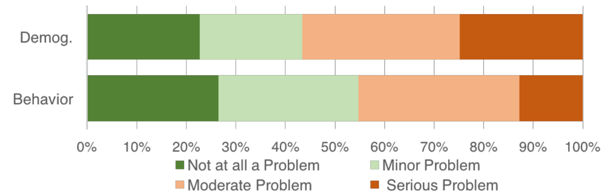

Figure 2: Problem severity, organized by targeting mechanism (Pilot 1).

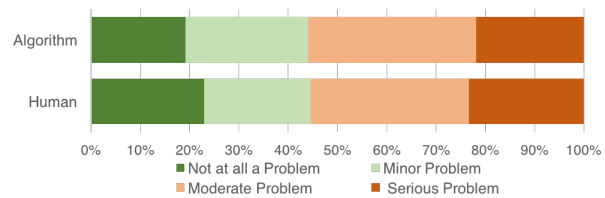

Figure 3: Problem severity, organized by human or algorithmic decision (Pilot 1).

ing the severity of the various scenarios. The most noticeable pattern was that scenarios that targeted based on behavior (e.g., browsing history), rather than explicit demographics, were generally rated less problematic (see Figure 2).

Third, we had hypothesized that whether a human or an algorithm made the decision to target the advertisement would play an important role in respondents' perceptions of the scenario. We were surprised that we did not find evidence for this in the pilot, but we decided to include it in our subsequent studies in hopes of confirming (or not) its lack of importance (see Figure 3).

5 Main Study

Based on the results from Pilot 1, we designed our final survey. Below, we detail the content of this final survey and the results of our generation and validation of regression models for analysis of this data.

5.1 Final Survey Instrument

Our final survey contrasted demographic and behavioral explanations, as well as human and algorithmic decisions. Because there is confusion about which entity in the complex advertising ecosystem makes decisions that can have discriminatory outcomes, and because we were explicitly interested in asking questions about responsibility, we included a factor locating the decision-making either at Systemy (the company placing the ad) or Bezo (the ad network). We did not include the local news site as a potential decision-maker because it did not seem to provide particularly interesting results in Pilot 1. As discussed in

Target	Mechanism	Decider	Entity
White	Behavior	Human	Advertiser
Asian	Demographics	Algorithm	Ad Network
Black			

Table 4: Variables included in the scenarios for the final survey instrument.

Section 4, we limited the targeted groups to only consider race.

The final set of 24 scenarios (demographic vs. behavioral × human vs. algorithmic × two entities × three target groups) is detailed in Table 4.

The text of the scenario shown to the respondents was:

> Systemy is a local technology firm that develops software. They are expanding and want to hire new employees. Systemy contracts with Bezo Media, an online advertising network, which places Systemy's job ad on a local news website. [EXPLANATION]. As a result, the ad is shown more frequently to [TARGET] individuals than [OPPOSITE OF TARGET] individuals.

The explanations shown to the respondents can be found in Table 5.

Because the scenario wording remained very close to the wording as used in Pilot 1, we did no further cognitive interviews.

5.2 Pilot 2: Training Data Generation

Before running the final collection of data with this survey, we conducted one additional pilot survey. This pilot generated training data that we used to test a variety of potential regression models without worrying about erosion of statistical confidence due to multiple testing. Such testing allowed us to narrow down the breadth of potential covariates to only the most relevant.

5.2.1 Respondents

As the goal of Pilot 2 was to create training data for selecting a final set of regression models to be confirmed with a larger data collection, we considered it sufficient to collect a smaller, less diverse—and also less expensive—sample. We deployed our four- to five-minute survey to 191 respondents using Amazon's Mechanical Turk crowdsourcing service (MTurk).[1] MTurk has been shown to provide adequate data quality, but also to be younger and more educated than the general population [24, 26]. We required respondents to have an approval rate of at least 85% on the MTurk service and reside in the U.S., and we compensated them $0.75 each. To avoid duplicate respondents, each participant's unique MTurk identification number was recorded and duplicate IDs were prevented from completing the survey again. Detailed demographics can be found in Table 6.

[1] https://www.mturk.com

Metric	MTurk	Census
Male	48.2%	48.2%
Female	51.8%	51.8%
Caucasian	81.2%	64.0%
Hispanic	4.7%	16.0%
Asian	4.7%	5.4%
African American	7.3%	12.0%
Other	2.1%	2.6%
Up to H.S.	13.6%	41.3%
Some college	32.5%	31.0%
B.S. or above	53.9%	27.7%
18–29 years	26.6%	20.9%
30–49 years	53.1%	34.7%
50–64 years	16.7%	26.0%
65+ years	3.6%	18.4%

Table 6: Respondent demographics for Pilot 2, compared to figures from the 2015 U.S. Census [41].

5.2.2 Analysis and Results

Because the majority of our survey questions were Likert scales, we primarily analyze our data using logistic regression, which measures how several different input factors correlate with a step increase in the output Likert variable being studied [23]. This allows us to examine how both our experimental factors and demographic covariates correlate with respondents' reactions to the presented scenario.

For the degree of responsibility and problem questions, we generated an initial model including the experimental factors (the target, mechanism, decider, and entity variables from Table 4); participant demographic covariates including age, gender, ethnicity, and education level; and pairwise interactions between various factors. We then compared a variety of models using subsets of these covariates, looking for the best fit according to the lowest Akaike Information Criterion (AIC) [4]. (We included the experimental factors in every model we considered.)

For each question, multiple models were very close in AIC value. From among those with near-minimal AIC for each of the five questions, we selected a final model that included the four experimental factors—target, mechanism, decider, and entity—along with the demographic covariates that appeared most relevant. No pairwise interactions were included in the final model. The final set of factors and covariates is summarized in Table 7. For each

Targets:		
• Are/be white	• Are/be Asian	• Are/be black

Explanations:

- An employee at Systemy places an order with Bezo Media to show the ad more often to people who have recently visited technology-interest websites. The employee predicts, based on prior experience, that people who recently visited a technology-interest website will be more likely to read and click on the ad. Individuals who are [TARGET] tend to visit more technology-interest websites than individuals of other races.
- Systemy uses an algorithm to decide how to place its ads. The algorithm places an order with Bezo Media to show the ad more often to people who have recently visited technology-interest websites. The algorithm predicts, based on prior data, that people who have recently visited a technology-interest website will be more likely to read and click on the ad. Individuals who are [TARGET] tend to visit more technology-interest websites than individuals of other races.
- Systemy uses an algorithm to decide how to place its ads. The algorithm places an order with Bezo Media to show the ad more often to people who are [TARGET] than individuals of other races. The algorithm predicts, based on prior data, that [TARGET] people will be more likely to read and click on the ad.
- An employee at Systemy places an order with Bezo Media to show the ad more often to people who are [TARGET] than individuals of other races. The employee predicts, based on prior experience, that [TARGET] people will be more likely to read and click on the ad.
- Bezo Media uses an algorithm to decide when to show which ads. The algorithm shows the ad more often to people who have recently visited technology-interest websites. The algorithm predicts, based on prior data, that people who had recently visited a technology-interest website will be more likely to read and click on the ad. Individuals who are [TARGET] tend to visit more technology-interest websites than individuals of other races.
- An employee at Bezo Media decides to show the ad more often to people who have recently visited technology-interest websites. The employee predicts, based on prior experience, that people who recently visited a technology-interest website will be more likely to read and click on the ad. Individuals who are [TARGET] tend to visit more technology-interest websites than individuals of other races.
- An employee at Bezo Media decides to show the ad more often to people who are [TARGET] than individuals of other races. The employee predicts, based on prior experience, that [TARGET] people will be more likely to read and click on the ad.
- Bezo Media uses an algorithm to decide when to show which ads. The algorithm shows the ad more often to people who are [TARGET] than individuals of other races. The algorithm predicts, based on prior data, that [TARGET] people will be more likely to read and click on the ad.

Table 5: Scenarios in the final survey instrument. Each participant viewed one explanation, with one targeted group filled in as receiving more of the targeted ads.

question, we excluded respondents who gave "don't know" responses to that question from the associated regression analysis.

5.3 Final Survey Results

To validate the regression model developed during Pilot 2, we conducted a final, larger-scale data collection with our final survey instrument. To promote both high data quality and broad generalizability in our results, with reasonable cost, we deployed our survey with both MTurk and SSI. We again required Turkers to have 85% approval and compensated them $0.75; we again paid SSI $3.00 per completion. Respondents from both the first and second pilot study were excluded from participation in this survey. To account for differences in the two samples, we added sample provider as a covariate to our regression model (shown at the bottom of Table 7).

Table 8 summarizes the results.

5.3.1 Respondents

We collected responses from 535 MTurk respondents and 372 SSI respondents, for a total of 907. Demographics for the two samples are shown in Table 9, with U.S. Census data for comparison [41].

The 16 respondents who reported their race as "other" were excluded from the dataset, because the small sample frequently prevented the regression model from converging. All further results are therefore reported for the remaining 891 respondents, or for slightly fewer when respondents answered "don't know" to certain questions.

Factor	Description	Baseline
Target	The ethnicity receiving more ads in the scenario. White, Asian, or Black.	White
Mechanism	Decision made based on either the demographics or the behavior of the targeted group.	Demographics
Decider	Whether the targeting decision was made by an algorithm or a human.	Algorithm
Entity	Entity making the decision: Either the ad network or the advertiser.	Ad network
Age	Of respondent. Continuous.	n/a
Education	Of respondent. High school diploma or less, Some college (HS+), Bachelor's Degree and up (BS+)	High school or less
Ethnicity	Of respondent. White, Black, Hispanic or Latino, Asian, or Other	White
Sample Provider	Amazon's Mechanical Turk and SSI	MTurk

Table 7: Factors used in the regression models for problem, responsibility, ethics, and believability. The sample provider factor was used in the main study only, not in Pilot 2.

		Ad network		Advertiser		News site		End user	
Factor	**Severity**	**Respons.**	**Ethical**	**Respons.**	**Ethical**	**Respons.**	**Ethical**	**Respons.**	**Ethical**
T-Asian	🟢	🟠	–	🟠	–	–	–	–	–
T-Black	🟢	–	–	–	–	–	–	–	–
Behavior	🟢	🟢	🟢	🟢	🟢	–	🟢	–	–
Human	–	–	–	–	–	–	–	–	–
Advertiser	–	🟢	–	🟠	–	–	–	–	–
Age of respondent	🟢	🟢	–	–	–	🟢	–	🟢	–
HS+	🟠	–	–	–	–	–	–	–	–
BS+	🟠	–	🟠	–	–	–	🟠	🟠	–
R/E-Asian	–	–	–	–	–	–	–	–	–
R/E-Black	🟠	🟠	–	–	–	–	–	🟠	–
R/E-Hisp. or Lat.	–	–	–	–	–	–	–	–	–
SSI	🟠	–	–	–	–	🟠	–	🟠	–

Table 8: Summary of regression results. 🟠 indicates a significant increase in severity, in responsibility, or in unethical behavior, compared to baseline, as appropriate. 🟢 indicates a significant decrease, and – indicates no significant effect. T- indicates the race of the targeted group, while R/E indicates the race or ethnicity of the respondent.

5.3.2 Model Validation

To verify that the set of factors and covariates we selected in Pilot 2 were also reasonable for our final data, we verified that the error rate when applying this regression to the final dataset was within the confidence interval of the error rate observed on our training data (e.g. the Pilot 2 data). More specifically, we bootstrapped [15] root mean square error (RMSE) [31] confidence intervals from the Pilot 2 data and verified that the RMSE after applying the models to the final data were within these confidence intervals. This enabled us to verify that the models selected based on our training data were appropriately fit to the final data. All of the models *except* the model for user responsibility and the model for local responsibility were appropriately fit. We retain these two models for analysis continuity, while acknowledging that a different model might have been a better fit.

5.3.3 Severity of Problem

Respondents were asked, on a four-point scale from "not a problem" (1) to "a serious problem" (4), to rate how problematic they found the discrimination scenario with which they were presented. The ordering and phrasing of the scale was taken from a commonly used set of Likert-type items developed by Vagias [43]. Across all scenarios, 44% of respondents selected a "moderate" (3) or "serious" (4) problem.

Overall, respondents gave a median rating of "somewhat of a problem" (2) to scenarios in which the discriminatory advertising occurred as a result of the users' behavior (e.g., Asian people visit technology job sites more often and thus Asian people saw the ad more often), compared to a median rating of "moderate problem" for scenarios in which discrimination occurred due to direct demographic targeting. In the demographic scenario, 53% of respondents indicated a moderate or severe problem, compared to 34% in the behavioral scenario. Figure 4

Metric	SSI	MTurk	Total	Census
Male	41.4%	50.5%	46.7%	48.2%
Female	58.3%	49.5%	53.1%	51.8%
Caucasian	63.2%	83.2%	75.0%	64.0%
Hispanic	12.9%	3.9%	7.6%	16.0%
Asian	5.4%	4.9%	5.1%	5.4%
African American	16.9%	6.2%	10.6%	12.0%
Other	1.6%	1.9%	1.8%	2.6%
Up to H.S.	31.7%	11.2%	19.6%	41.3%
Some college	35.8%	33.5%	34.4%	31.0%
B.S. or above	32.5%	55.3%	46.0%	27.7%
18–29 years	20.4%	27.1%	24.4%	20.9%
30–49 years	41.9%	56.4%	50.5%	34.7%
50–64 years	31.5%	14.8%	21.6%	26.0%
65+ years	6.2%	1.7%	3.5%	18.4%

Table 9: Respondent demographics for the main study. The Total column is the demographics of the total sample including both the MTurk and SSI respondents. The census figures are from 2015 U.S. Census [41].

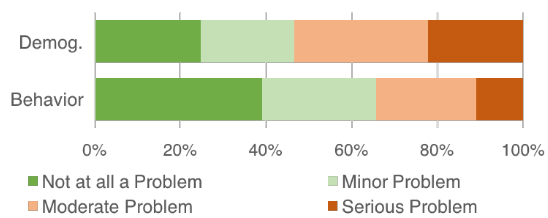

Figure 4: Responses for problem severity, broken down into behavior and demographic conditions.

provides an overview of the scores. If we instead compare scenarios based on whether a human or algorithm decided to do the targeting, we find the respondents gave a median rating of "somewhat of a problem" in both cases.

To assess which factors influence respondents' perceptions of problem severity, we conducted a regression analysis (as described in Section 5.2.2). Results are shown in Table 10. Using this analysis, we find that respondents' perception of the severity of the scenario was significantly affected by how the discrimination took place (e.g., based on users' online behavior vs. explicitly their demographics). Behavior-based ad targeting was only 49% as likely as demographic-based targeting to increase respondents' severity rating. That is, respondents evidenced less concern when user behavior (in this case, web browsing history) led to de-facto discrimination than when explicit demographic targeting yielded the same result.

Respondents also found targeting black and Asian individuals for more job ads significantly less problematic (58% and 60% as likely to increase severity rating, respectively) than targeting white individuals. On the other hand,

Factor	OR	CI	*p*-value
T-Asian	0.60	[0.41, 0.88]	0.010*
T-Black	0.58	[0.40, 0.86]	0.006*
Behavior	0.49	[0.36, 0.67]	<0.001*
Human	1.11	[0.82, 1.51]	0.498
Advertiser	0.94	[0.69, 1.28]	0.689
Age of respondent	0.99	[0.97, 1.00]	0.040*
HS+	1.76	[1.13, 2.75]	0.013*
BS+	1.58	[1.03, 2.43]	0.036*
R/E-Asian	1.34	[0.67, 2.68]	0.413
R/E-Black	2.87	[1.55, 5.34]	<0.001*
R/E-Hispanic or Latino	1.94	[0.99, 3.85]	0.052
SSI	1.66	[1.17, 2.35]	0.005*

Table 10: Regression results for problem severity (n=830). n may not add to the total number of respondents due to item non-response. OR is the odds ratio between the given factor and the baseline: that is, how many times more likely this factor is than the baseline to increase one step on the four-point problem severity scale. CI is the 95% confidence interval for the odds ratio. Statistically significant factors (p <0.05) are denoted with a *. T- indicates the race of the targeted group, while R/E indicates the race or ethnicity of the respondent.

as was the case in both pilots, whether the decision on how to target the advertisement was made by an algorithm or a human did not appear to affect respondents' perceptions. The entity doing the targeting (advertiser or ad network) similarly had no significant effect on perceptions.

Certain respondent demographics also factored into ratings of problem severity. Table 10 shows that older respondents are associated with lower severity ratings; for example, a 10-year age gap is associated with only a 90% ($0.99^{10} = 0.90$) likelihood of increased severity. Black respondents were 2.87× as likely as baseline white respondents to rate the problem as more severe. Results for education level indicate that holding at least a high-school diploma was associated with higher likelihood of increased severity; there was no apparent further distinction based on achievement of a bachelor's degree. Finally, respondents recruited through SSI were 1.66× more likely to increase one step in severity, despite our model separately accounting for age and ethnicity.

5.3.4 Degree of Responsibility

We next consider the responsibility level respondents assign to different entities involved in the discriminatory scenario: the ad network (Bezo Media), the advertiser (Systemy), the local news website on which the advertisement was displayed, and the end user who sees the

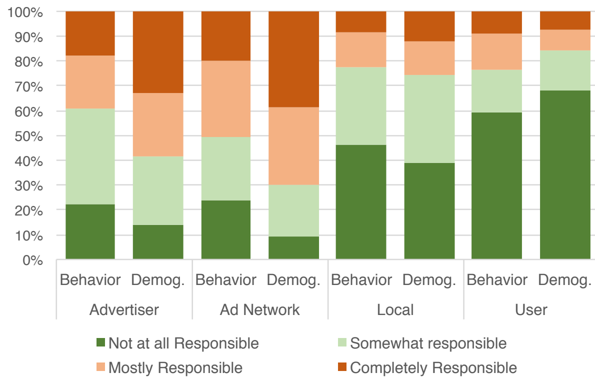

Figure 5: Responsibility scores, per entity, broken down by the behavioral and demographic conditions.

Factor	OR	CI	*p*-value
T-Asian	1.64	[1.01, 2.67]	0.045*
T-Black	1.10	[0.70, 1.73]	0.674
Behavior	0.33	[0.22, 0.49]	<0.001*
Human	1.12	[0.76, 1.66]	0.554
Advertiser	0.44	[0.30, 0.66]	<0.001*
Age of respondent	0.97	[0.96, 0.99]	<0.001*
HS+	0.88	[0.51, 1.52]	0.656
BS+	1.40	[0.81, 2.43]	0.228
R/E-Asian	1.28	[0.50, 3.32]	0.604
R/E-Black	3.24	[1.34, 7.86]	0.009*
R/E-Hispanic or Latino	1.71	[0.72, 4.03]	0.221
SSI	1.01	[0.66, 1.56]	0.946

Table 11: Regression results for ad network responsibility (n=867), where OR > 1 is associated with more responsibility. See Table 10 caption for more detailed explanation.

ad. Respondents provided their responsibility ratings on a four-point scale including “not at all responsible” (1), “somewhat responsible” (2), “mostly responsible” (3), and “completely responsible” (4) [43].

Across scenario types, 63% of respondents rated the user as “not at all responsible” for the outcome; this was also the median value. Respondents also did not attribute a high level of responsibility to the local news network: the median responsibility score in this case was “somewhat responsible,” with 42% of respondents selecting “not at all responsible.” On the other hand, only 17% and 18%, respectively, of respondents rated the ad network and the advertiser “not at all responsible,” with the median score for the ad network “mostly responsible” and for the advertiser “somewhat responsible.” Respondents’ ratings of responsibility for each entity are shown in Figure 5.

We also applied regression analysis to determine what factors influenced respondents ratings of responsibility for each of these entities. Tables 11–14 illustrate the results of the regressions for each entity.

For all entities, except for end user, the mechanism by which the advertisement was targeted (demographics vs. behavior) is significant. The advertiser, ad network, and local news site all accrue less responsibility when behavior is used. This effect is strongest for the ad network; respondents are only 33% as likely to rate the ad network as responsible for the discrimination when demographic targeting rather than behavioral targeting is used. The advertiser and ad network also accrue more responsibility when Asian people are targeted as compared to white people.

As might be expected, responsibility aligns with the details of the scenarios seen by the respondents: the advertiser is assigned more responsibility when the scenario provided implicates the advertiser than when it implicates the ad network, and the same holds for the ad network’s responsibility when the scenario implicates the network. The implicated entity does not significantly affect how responsibility is assigned to the local news site or end user. These results, while unsurprising, do help to validate that our respondents read and understood their assigned scenarios. As with problem severity, whether a human or algorithm made the targeting decision continues to have no significant impact.

Also similarly to problem severity, age proved a significant factor for three of the four responsibility questions (not advertiser). In all three cases, older respondents were correlated with lower responsibility scores. Finally, respondents recruited from SSI assigned greater responsibility to the local news site and the end user than MTurk respondents. Unlike with problem severity, the race of the respondent appears to have limited correlation with responsibility assignment in most cases.

5.3.5 Ethical Behavior

Next, we consider respondents’ opinions about whether each of the four entities behaved ethically. Specifically, respondents were asked to agree or disagree that the entity had behaved ethically, on a five-point Likert scale from strongly agree to strongly disagree. Across all scenarios, 75% of respondents agreed or strongly agreed that the user behaved ethically (median = agree, or 2). Additionally, 57% of respondents reported that the local news network behaved ethically (median = agree). On the other hand, only 49% and 43% agreed or strongly agreed that the advertiser and ad network, respectively, behaved ethically (both medians = neutral (3)). We note that these ratings align well with those observed for responsibility.

The regression results for ethical behavior are shown in Tables 15–18. Consistent with the findings from previous

Factor	OR	CI	*p*-value
T-Asian	1.62	[1.03, 2.58]	0.038*
T-Black	0.87	[0.57, 1.32]	0.518
Behavior	0.54	[0.37, 0.77]	<0.001*
Human	0.70	[0.49, 1.01]	0.055
Advertiser	1.96	[1.36, 2.83]	<0.001*
Age of respondent	0.99	[0.97, 1.00]	0.160
HS+	0.66	[0.38, 1.12]	0.125
BS+	0.80	[0.47, 1.36]	0.403
R/E-Asian	1.98	[0.75, 5.26]	0.170
R/E-Black	1.71	[0.84, 3.49]	0.140
R/E-Hispanic or Latino	1.06	[0.53, 2.12]	0.867
SSI	1.06	[0.71, 1.58]	0.783

Table 12: Regression results for advertiser responsibility (n=857), where OR > 1 is associated with more responsibility. See Table 10 caption for more detailed explanation.

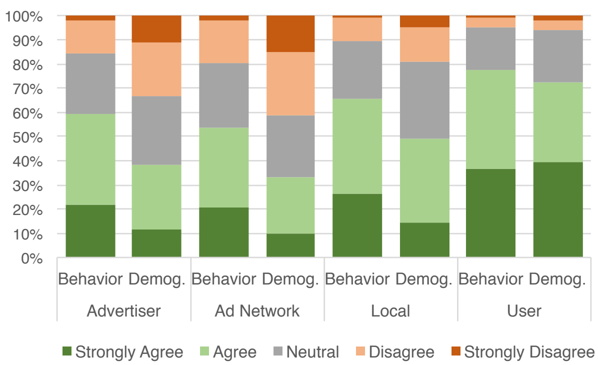

Figure 6: Agreement that each entity was behaving ethically, broken down by the behavioral and demographic conditions.

questions, the mechanism of targeting is significant for advertiser, ad network and local news website; in all three cases, behavior-based targeting is significantly correlated with a lower perception of unethical behavior than the demographic-based targeting. This is illustrated in Figure 6. Human vs. algorithmic decision making continues to show no significant effect.

In contrast to responsibility, the entity making the decision in the provided scenario (the ad network or advertiser) does not appear to have a significant effect on respondents' perceptions of ethical behavior in any case. The targeted group is similarly uncorrelated.

Respondent demographics appear to have little to no correlation with these results. In two cases (ad network and local news site), Asian respondents were more likely to disagree that the entity in question had behaved ethically, but no other demographic covariates were significant.

Factor	OR	CI	*p*-value
T-Asian	0.86	[0.61, 1.22]	0.400
T-Black	1.00	[0.70, 1.41]	0.983
Behavior	0.71	[0.54, 0.95]	0.019*
Human	0.89	[0.67, 1.18]	0.430
Advertiser	1.17	[0.88, 1.55]	0.284
Age of respondent	0.98	[0.97, 1.00]	0.011*
HS+	0.77	[0.51, 1.17]	0.216
BS+	0.80	[0.53, 1.19]	0.271
R/E-Asian	1.67	[0.85, 3.28]	0.140
R/E-Black	1.08	[0.66, 1.75]	0.764
R/E-Hispanic or Latino	1.21	[0.69, 2.14]	0.502
SSI	2.00	[1.46, 2.76]	<0.001*

Table 13: Regression results for local news site responsibility (n=843), where OR > 1 is associated with more responsibility. See Table 10 caption for more detailed explanation.

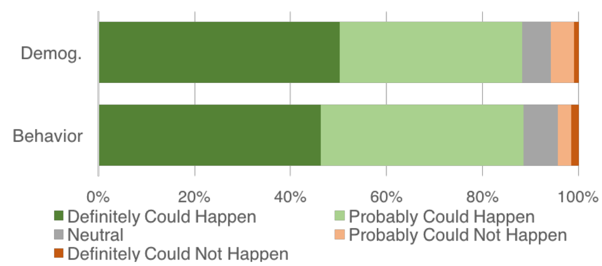

Figure 7: Responses for scenario believability, broken down into behavior and demographic conditions.

5.3.6 Believability

Because several of our cognitive interview respondents expressed skepticism that discriminatory scenarios like the ones we described could be realistic, we added a question about believability at the end of the survey. Respondents were asked to rate the scenario on a five-point scale from "definitely could not happen" to "definitely could happen." Overall, 88% of respondents reported that the scenario "definitely" or "probably" could happen. Figure 7 provides an overview of respondents' ratings of scenario believability. This result suggests that, among the populations we surveyed, there is widespread if potentially shallow awareness of behavioral targeting capabilities and the potential for discrimination, intentional or otherwise.

6 Limitations

Our study, like most similar surveys, has several important limitations. First, while our sample included a broad variety of demographic groups, it was not a true probabilistic

Factor	OR	CI	*p*-value
T-Asian	0.99	[0.69, 1.42]	0.962
T-Black	0.84	[0.58, 1.20]	0.340
Behavior	1.34	[1.00, 1.79]	0.054
Human	1.05	[0.78, 1.41]	0.734
Advertiser	1.30	[0.97, 1.75]	0.080
Age of respondent	0.97	[0.95, 0.98]	<0.001*
HS+	0.75	[0.49, 1.14]	0.177
BS+	0.73	[0.48, 1.10]	0.131
R/E-Asian	1.96	[1.03, 3.73]	0.041*
R/E-Black	1.66	[1.04, 2.67]	0.035*
R/E-Hispanic or Latino	1.31	[0.76, 2.28]	0.330
SSI	2.40	[1.73, 3.32]	<0.001*

Table 14: Regression results for end-user responsibility (n=851), where OR > 1 is associated with more responsibility. See Table 10 caption for more detailed explanation.

Factor	OR	CI	*p*-value
T-Asian	0.87	[0.55, 1.37]	0.535
T-Black	1.04	[0.65, 1.65]	0.885
Behavior	0.42	[0.28, 0.62]	<0.001*
Human	0.97	[0.67, 1.42]	0.885
Advertiser	0.81	[0.56, 1.19]	0.284
Age of respondent	1.00	[0.98, 1.01]	0.635
HS+	0.82	[0.46, 1.46]	0.495
BS+	0.61	[0.35, 1.06]	0.078
R/E-Asian	5.55	[1.30, 23.59]	0.020*
R/E-Black	1.58	[0.80, 3.14]	0.189
R/E-Hispanic or Latino	1.70	[0.73, 3.94]	0.216
SSI	0.80	[0.53, 1.21]	0.293

Table 15: Regression results for ethical behavior by the ad network (n=891), where OR > 1 is associated with stronger disagreement that the ad network behaved ethically. See Table 10 caption for more detailed explanation.

sample. While we believe our conclusions can to some extent generalize, Turkers and web panel participants are generally more active internet users than average. People with less technical knowledge might find our scenarios less believable or feel differently about what constitutes a severe problem.

Second, our surveys dealt with the highly sensitive topic of discrimination, especially racial discrimination. Social desirability bias may cause respondents to report higher-than-realistic severity of discrimination scenarios, particularly with respect to historically disadvantaged groups.

Third, the ad eco-system is complex and complicated. There are many different entities involved in the publishing of an ad. In this survey, we took some of the involved entities and incorporated them into simplified scenarios. Despite simplification and pre-testing via cognitive interviews, it is possible some respondents did not understand important subtleties of these scenarios, affecting their responses. However, the fact that respondents tended to most blame whichever entity was implicated by the scenario (Section 5.3.4 suggests that respondents understood the scenarios to at least some degree.

More generally, all self-report surveys are susceptible to respondents who hurry through, answer haphazardly, or do not think deeply about the questions. In this particular survey, we were concerned that the scenarios might be too complex for some participants to understand, or that participants who did not believe the discriminatory scenario might not answer meaningfully. To minimize these effects, we kept the survey short and used cognitive interviews to ensure that our questions and answer choices could be easily understood. We explicitly measured believability and found that the majority of participants did find our scenario plausible. In addition, our major results proved consistent across two pilots and our main survey. As a result, we are reasonably confident that respondents were able to provide thoughtful answers to our questions.

Fourth, only some of our variables, the factors Target, Mechanism, Entity, and Decider, were experimentally randomized. Thus, our causal claims only extend to them. For the other variables, the covariates Age, Ethnicity, Education, and Sample Provider, we can only makes claims of correlation.

Fifth, despite basing our conclusions on three rounds of data collection, false positives remain possible. Pilot 2 produced a series of hypotheses about what variables would constitute a useful, parsimonious logistic regression model that we could apply across the nine questions we asked. The main study applied this model to new data, and investigated how these variables were correlated with each outcome. (For Decider, based on Pilot 2, we anticipated a coefficient statistically indistinguishable from zero.) The main regressions also controlled for the data coming from SSI, about which we had no explicit hypotheses motivated by Pilot 2.

In our main study we consider each variable-question combination as an independent hypothesis; we do not aggregate across questions or variables. Intuitively, these variables and questions are distinguishable from one another and different explanations may apply to each. As such, we do not correct for multiple hypothesis testing. Ultimately the question of when to aggregate and adjust *p*-values or not to comes down judgements about the individual hypotheses being interesting each on their own [20] and the goals of the study [7].

Factor	OR	CI	*p*-value
T-Asian	0.94	[0.60, 1.45]	0.765
T-Black	1.02	[0.66, 1.59]	0.925
Behavior	0.46	[0.32, 0.67]	<0.001*
Human	0.91	[0.63, 1.30]	0.602
Advertiser	1.42	[0.99, 2.03]	0.058
Age of respondent	1.00	[0.98, 1.01]	0.832
HS+	0.97	[0.57, 1.65]	0.912
BS+	0.76	[0.45, 1.27]	0.293
R/E-Asian	2.76	[0.95, 7.99]	0.062
R/E-Black	1.07	[0.59, 1.95]	0.818
R/E-Hispanic or Latino	1.94	[0.84, 4.48]	0.120
SSI	0.83	[0.56, 1.24]	0.365

Table 16: Regression results for ethical behavior by the advertiser (n=891), where OR > 1 is associated with stronger disagreement that the advertiser behaved ethically. See Table 10 caption for more detailed explanation.

Factor	OR	CI	*p*-value
T-Asian	0.98	[0.65, 1.47]	0.925
T-Black	1.16	[0.77, 1.76]	0.470
Behavior	0.45	[0.32, 0.64]	<0.001*
Human	0.91	[0.65, 1.27]	0.569
Advertiser	0.80	[0.57, 1.12]	0.198
Age of respondent	0.99	[0.98, 1.01]	0.419
HS+	1.31	[0.80, 2.12]	0.279
BS+	0.94	[0.59, 1.50]	0.808
R/E-Asian	3.25	[1.12, 9.38]	0.029*
R/E-Black	1.14	[0.65, 2.02]	0.641
R/E-Hispanic or Latino	1.29	[0.64, 2.60]	0.475
SSI	1.04	[0.71, 1.52]	0.841

Table 17: Regression results for ethical behavior by the local news site (n=891), where OR > 1 is associated with stronger disagreement that the site behaved ethically. See Table 10 caption for more detailed explanation.

7 Discussion and Conclusion

Below, we present a summary of our findings, discussion on the respondents' understanding, implications for governance and policy guidelines for OBA, and suggestions for future work.

7.1 Summary of Findings

Overall, we find that for most questions we examined, people's perceptions of discriminatory ad-targeting scenarios depend on how the discrimination occurred. As might be expected, respondents rated scenarios in which the discrimination occurred based on how users behaved, with no explicit intent to discriminate based on demographic characteristics, to be significantly less problematic than scenarios with explicit racial targeting. Respondents also assigned more blame to the ad network, advertiser, and host website, and rated these entities' behavior as less ethical, in the behavioral scenarios.

Respondents also found scenarios in which minorities (in our scenarios, people of black or Asian race) benefited from such ad-targeting discrimination less problematic than scenarios in which the majority benefited. Relatedly, we also find that black respondents are more likely to view discriminatory scenarios as a more severe problem. We hypothesize that these ratings are influenced by discriminatory history in the U.S., where we recruited our respondents.

We find that whether the ad network or advertiser is explicitly mentioned in the scenario as causing the discrimination influences the accrual of responsibility to those entities; however, to our surprise, the named entity did not influence respondents' ratings of the severity of the scenarios, or of whether any entity had behaved ethically. Overall, the median ethics rating for both the ad network and the advertiser was neutral. We suspect this may relate part to many respondents not entirely understanding some subtleties of the online ad ecosystem. Nevertheless, these results suggest that it is not necessarily helpful for entities to "pass the blame" to other players, as the mechanism of discrimination seems more important. We were also surprised to find that whether a person or an algorithm was responsible for selecting how and whom to target made no difference in respondents' ratings of the severity of the scenario, suggesting that "an algorithm did it" will not be a viable excuse.

Finally, we find that the majority (88%) of respondents believed our scenario, suggesting a wariness or even awareness of these issues, at least among heavily-internet-using Turkers and SSI panel members.

7.2 Governance and Policy Implications

A number of organizations, including the FTC, the EFF, and industry groups such as the American Advertising Federation, provide guidelines and recommendations for the ethical use of targeted advertising [1, 18, 25]. Of these recommendations, only the EFF policy document mentions discrimination as a potential, unethical consequence. Our results, as well as prior research that has brought to light instances of discrimination (e.g., [13, 37]), highlight the importance of discrimination as an ad-targeting consideration. We find that 43% of respondents rated our discriminatory advertising scenarios a significant or moderate problem. More specifically, in the more prob-

Factor	OR	CI	*p*-value
T-Asian	0.91	[0.65, 1.27]	0.574
T-Black	1.05	[0.75, 1.47]	0.789
Behavior	1.15	[0.87, 1.51]	0.321
Advertiser	0.94	[0.71, 1.24]	0.656
Human	0.91	[0.69, 1.20]	0.509
Age of respondent	1.00	[0.99, 1.01]	0.566
HS+	0.96	[0.64, 1.44]	0.845
BS+	0.68	[0.46, 1.00]	0.051
R/E-Asian	1.90	[0.97, 3.74]	0.063
R/E-Black	1.57	[0.97, 2.53]	0.067
R/E-Hispanic or Latino	1.30	[0.76, 2.24]	0.340
SSI	1.03	[0.76, 1.40]	0.839

Table 18: Regression results for ethical behavior by the end user (n=891), where OR > 1 is associated with stronger disagreement that the end user behaved ethically. See Table 10 caption for more detailed explanation.

lematic demographic scenario, 53% did so; even in the less problematic behavioral scenario, when discrimination happened as a result of targeting based on users' web browsing history, 34.2% did so. Thus, we propose that guidelines, especially those issued by government agencies, should include explicit language about discrimination to address this topic of common concern.

Our findings suggest that while respondents distinguish behavioral from demographic targeting, they are not especially concerned with whether an algorithm was involved in the outcome. This suggests that responses that focus on the algorithmic nature of the ad ecosystem may not be helpful for addressing public concerns.

Finally, our findings represent a broad cross-section of users' opinions, but they do not represent a normative guideline for what *should* be appropriate. Many kinds of discrimination that may seem acceptable to the general public today may in fact be illegal, immoral, or unjust. Activists and advocates who are concerned about online discrimination can use our work as a starting point to better understand where more education, persuasion, and lobbying for new regulations may be most needed for furthering their agenda.

7.3 Future Work

Overall, our work addresses only a small portion of the critical topic of online algorithmic discrimination. Our results highlight an important distinction between users' perceptions of scenarios involving explicitly racial vs. implicitly racial, online-behavior-based discrimination. However, we explored only web-history-based targeting, and thus, future work may seek to explore whether users react similarly to other types of behaviors, or whether certain online behaviors are more sensitive.

Similarly, future work is needed to explore reactions to discrimination based on factors other than race. Our first pilot results suggested that users did not feel as strongly about topics such as pre-existing health conditions, at least in our advertising scenario, this should be explored in further detail in a wider range of scenarios.

Relatedly, we only explored user perceptions of scenarios involving advertising discrimination, and only in the context of a potentially desirable ad (for a job). It would be interesting to explore whether reactions remain the same when the ad in question is potentially undesirable, for example related to bail bonds or drug-abuse treatment. Related work [17, 35] has also shown evidence of discrimination in the search results shown to different users; questions about discrimination in pricing, insurance, and other services also remain open. Thus, future work could focus on exploring and comparing user reactions to discriminatory results in a variety of settings.

Finally, the concrete regression models, with particular coefficient values, as described in Section 5.3, were not tested for predictive power against independent test data. Such validation may make interesting future work for those interested in accurately predicting people's responses to cases of discrimination.

8 Acknowledgments

We thank Sean Kross for statistical guidance, the respondents in our surveys, and the anonymous reviewers of our submission. We gratefully acknowledge funding support from the National Science Foundation (Grant 1514509), and from a UMIACS contract under the partnership between the University of Maryland and DoD. The opinions in this paper are those of the authors and do not necessarily reflect the opinions of any funding sponsor or the United States Government.

References

[1] Online behavioral tracking and targeting concerns and solutions: Legislative primer. https://www.eff.org/files/onlineprivacylegprimersept09.pdf, 2009.

[2] U.S. digital ad spending to surpass TV this year: Digital will represent 37% of U.S. total media ad spending. *eMarketer* (2016). https://www.emarketer.com/Article/US-Digital-Ad-Spending-Surpass-TV-this-Year/1014469.

[3] Agarwal, L., Shrivastava, N., Jaiswal, S., and Panjwani, S. Do not embarrass. In *SOUPS* (2013).

[4] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*, 6 (1974), 716–723.

[5] Balebako, R., Leon, P., Shay, R., Ur, B., Wang, Y., and Cranor, L. Measuring the effectiveness of privacy tools for limiting behavioral advertising. In *Web 2.0 Security and Privacy Workshop* (2012).

[6] Barford, P., Canadi, I., Krushevskaja, D., Ma, Q., and Muthukrishnan, S. Adscape: Harvesting and analyzing online display ads. In *WWW* (2014).

[7] Bendera, R., and Lange, S. Adjusting for multiple testing–When and how? *The Journal of Clinical Epidemiology 54*, 4 (Apr. 2001), 343–349.

[8] Bergemann, D., and Bonatti, A. Targeting in advertising markets: implications for offline versus online media. *The RAND Journal of Economics 42*, 3 (2011).

[9] Carpenter, J. Google's algorithm shows prestigious job ads to men, but not to women. Here's why that should worry you. . *The Washington Post* (2015). https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/.

[10] Carrascosa, J. M., Mikians, J., Cuevas, R., Erramilli, V., and Laoutaris, N. I always feel like somebody's watching me. Measuring online behavioural advertising. In *CoNEXT* (2015).

[11] Coen, R., Paul, E., Vanegas, P., Lange, A., and Hans, G. A user-centered perspective on algorithmic personalization. Master's thesis, University of California, Berkeley, 2016.

[12] Datta, A., Tschantz, M., and Datta, A. Discrimination and opacity in online behavioral advertising. http://possibility.cylab.cmu.edu/adfisher/. Accessed June 2017.

[13] Datta, A., Tschantz, M. C., and Datta, A. Automated experiments on ad privacy settings. *PoPETS* (2015).

[14] Debatin, B., Lovejoy, J. P., Horn, A.-K., and Hughes, B. N. Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication 15*, 1 (2009).

[15] Efron, B., and Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science* (1986), 54–75.

[16] Evans, D. S. The online advertising industry: Economics, evolution, and privacy. *The Journal of Economic Perspectives 23*, 3 (2009).

[17] FairSearch. Can search discrimination by a monopolist violate U.S. antitrust laws? http://www.fairsearch.org/wp-content/uploads/2011/07/Can-Search-Discrimination-by-a-Monopolist-Violate-U.S.-Antitrust-Laws1.pdf, 2011.

[18] FTC. Self-regulatory principles, for online behavioral advertising. https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-staff-report-self-regulatory-principles-online-behavioral-advertising/p085400behavadreport.pdf, 2009.

[19] Goldfarb, A., and Tucker, C. E. Online advertising, behavioral targeting, and privacy. *Communications of the ACM 54*, 5 (2011).

[20] Goodman, S. N. Multiple comparisons, explained. *American Journal of Epidemiology 147*, 9 (May 1998), 807–812.

[21] Guha, S., Cheng, B., and Francis, P. Challenges in measuring online advertising systems. In *IMC* (2010).

[22] Guynn, J. Google Photos labeled black people 'gorillas'. *USA Today* (2015).

[23] Hosmer, D. W., and Lemeshow, S. *Applied logistic regression*. 2000.

[24] Huff, C., and Tingley, D. "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research and Politics* (2015).

[25] Institute for Advertising Ethics. Principles and practices for advertising ethics. https://www.aaf.org/_PDF/AAF%20Website%20Content/513_Ethics/IAE_Principles_Practices.pdf, 2011.

[26] Ipeirotis, P. G. Demographics of Mechanical Turk. *SSRN* (2010).

[27] Johnson, L. U.S. digital advertising will make $83 billion this year, says EMarketer. *Adweek* (2017). http://www.adweek.com/digital/u-s-digital-advertising-will-make-83-billion-this-year-says-emarketer.

[28] Lecuyer, M., Ducoffe, G., Lan, F., Papancea, A., Petsios, T., Spahn, R., Chaintreau, A., and Geambasu, R. XRay: Enhancing the Web's Transparency

with Differential Correlation. In *USENIX Security* (2014).

[29] Lecuyer, M., Spahn, R., Spiliopolous, Y., Chaintreau, A., Geambasu, R., and Hsu, D. Sunlight: Fine-grained targeting detection at scale with statistical confidence. In *CCS* (2015).

[30] Leon, P., Ur, B., Wang, Y., Sleeper, M., Balebako, R., Shay, R., Bauer, L., Christodorescu, M., and Cranor, L. What matters to users? Factors that affect users' willingness to share information with online advertisers. In *SOUPS* (2013).

[31] Levinson, N. The wiener (root mean square) error criterion in filter design and prediction. *Studies in Applied Mathematics 25*, 1-4 (1946).

[32] Liu, B., Sheth, A., Weinsberg, U., Chandrashekar, J., and Govindan, R. AdReveal: Improving transparency into online targeted advertising. In *HotNets* (2013).

[33] Malheiros, M., Jennett, C., Patel, S., Brostoff, S., and Sasse, M. A. Too close for comfort: A study of the effectiveness and acceptability of rich-media personalized advertising. In *CHI* (2012).

[34] McDonald, A. M., and Cranor, L. F. Americans' attitudes about internet behavioral advertising practices. In *WPES* (2010).

[35] Mikians, J., Gyarmati, L., Erramilli, V., and Laoutaris, N. Detecting price and search discrimination on the internet. In *SOUPS* (2012).

[36] Silverman, D. IAB internet advertising revenue report. *Interactive Advertising Bureau. New York* (2010).

[37] Sweeney, L. Discrimination in online ad delivery. *Queue 11*, 3 (2013).

[38] Tschantz, M., Egelman, S., Choi, J., Weaver, N., and Friedland, G. The accuracy of the demographic inferences shown on Google's Ad Settings. Tech. Rep. TR-16-003, International Computer Science Institute, 2016.

[39] Turow, J., King, J., Hoofnagle, C. J., Bleakley, A., and Hennessy, M. Americans reject tailored advertising and three activities that enable it. In *SSRN* (2009), vol. 1478214.

[40] Ur, B., Leon, P. G., Cranor, L. F., Shay, R., and Wang, Y. Smart, useful, scary, creepy. In *SOUPS* (2012).

[41] U.S. Census Bureau. American community survey 5-year estimates. http://www.census.gov/programs-surveys/acs/news/data-releases/2015/release.html, 2015.

[42] U.S. Equal Employment Opportunity Commission. Prohibited employment policiespractices. https://www1.eeoc.gov//laws/practices/index.cfm?renderforprint=1. Accessed June 2017.

[43] Vagias, W. M. Likert-type scale response anchors. http://www.peru.edu/oira/wp-content/uploads/sites/65/2016/09/Likert-Scale-Examples.pdf, 2006.

[44] Valentino-Davies, J., Singer-Vine, J., and Soltani, A. Websites Vary Prices, Deals Based on Users' Information. *The Wall Street Journal* (2012).

[45] Warshaw, J., Taft, N., and Woodruff, A. Intuitions, Analytics, and Killing Ants - Inference Literacy of High School-educated Adults in the US. In *SOUPS* (2016).

[46] Willis, G. B. *A Tool for Improving Questionnaire Design*. Sage Publications, 2005.

[47] Wills, C. E., and Tatar, C. Understanding what they do with what they know. In *WPES* (2012).

A Survey Questions

Q1-4: How much responsibility does *entity* have for the fact that their ads are seen much more frequently by people who are *target race* than individuals of other races?

- Not at all responsible
- Somewhat responsible
- Mostly responsible
- Completely responsible
- Don't know

This question would be asked four times in a random order, each time with a new entity. Either Systemy (the advertiser), Bezo Media (the ad network), the individual visiting the website, or the the local news website.

Q5: Do you think it's a problem that Systemy job ads are seen much more frequently by people who are *target race* than individuals of other races?

- Not at all a problem
- Minor problem
- Moderate problem
- Serious Problem
- Don't know

Q6-9: Please tell us how much you agree or disagree with the following statements: *entity* behaved ethically in this situation

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

This question would be again be asked four times in a random order, each time with a new entity. Either Systemy (the advertiser), Bezo Media (the ad network), the individual visiting the website, or the the local news website.

Q10: Do you think the scenario we described could happen in real life?

- Definitely could happen
- Probably could happen
- Neutral
- Probably could not happen
- Definitely could not happen

Q11:Please specify your age. [drop-down menu of ages 18-100 or over]

Q12: Please specify the gender with which you most closely identify.

- Male
- Female
- Other

Q13: Please specify the highest degree or level of school you have completed.

- Some high school credit, no diploma or equivalent
- High school graduate, diploma or the equivalent (for example: GED)
- Some college credit, no degree
- Trade/technical/vocational training
- Associate degree
- Bachelor's degree
- Master's degree
- Professional degree
- Doctorate degree

Q14: Please specify your ethnicity.

- Hispanic or Latino
- Black or African American
- White
- American Indian or Alaska Native
- Asian
- Other