ABSTRACT

Title of Thesis: REACTIONS REGARDING ONLINE DISCRIMINATION AND AD PROFILES

Angelisa Carol Plane, Masters of Computer Science, 2017

Thesis Directed By: Dr. Michelle Mazurek, Assistant Professor

Department of Computer Science

The accessibility and amount of information obtained by online companies has grown over the past decade. This growth has led to the ability of companies to target a desired population to show certain content, products and other services. The two studies conducted for this thesis examine different aspects of online targeting and users' reactions using advertising as the primary tool. One of the main goals for the studies was to develop policy recommendations and guide policymakers into making ethical decisions. But to do this effectively some of the primary elements that we need to know is what the user understands, what they care about and why that concerns them. Keeping that in mind, we conducted three surveys that made up the first study which examined scenarios around discriminatory ads. For each scenario, we asked the user about their perception when it came to the level of problem and ethical behavior. For

the second study, we conducted interviews that had participants look at the profiles that Google and Facebook have created about them based on their online activity. We were able to ask questions in regard to their comfort level, their understanding of why certain interests might be shown to them, and their general understanding of how the profiling works.

These two studies were analyzed independently of each other, but the results and possible implications of each were combined to make recommendations to businesses and policy makers. From the first study, we found 43% of participants were moderately or very concerned by the scenarios, even when discrimination took place as result of online behavioral targeting. From the second study, we found several themes emerge from the interviews including the idea that more inaccurate inferences made make them feel more uncomfortable than accurate inferences. That sentiment was expressed by 64% of the participants.

REACTIONS REGARDING ONLINE DISCRIMINATION AND AD PROFILES

by

Angelisa Carol Plane

Thesis submitted to the Faculty of the Graduate School of the

University of Maryland, College Park, in partial fulfillment

of the requirements for the degree of

Masters of Computer
Science

2017

Advisory Committee:

Professor Michelle Mazurek, Chair

Professor Jon Froehlich

Professor Katie Shilton

Dr. Michal Tschantz

Acknowledgements

Firstly, I would like to thank my advisor, Dr. Michelle Mazurek, for her incredible support, guidance, and patience through this entire process. Her support and teaching made this possible and for that I am forever grateful.

I would also like to thank Elissa Redmiles for her support and help through this process.

Finally, I would like to thank my parents, Dr. Jandelyn and Michael for supporting me through all 19 years of my education.

Table of Contents

Acknowledgements ii
Table of Contents iii
List of Tables v
List of Figures vi
List of Abbreviations vii
Chapter 1: Introduction and Motivation 1
Chapter 2: Background and Related Work 4
  Human Perception 5
  Tools for Measuring Online Tracking and Advertising 8
  Discrimination in Targeted Advertising 9
  Existing Policy on Online Tracking 10
Chapter 3: Study One: Survey regarding Online Discrimination 12
  Pilot 1 12
    Scenarios 12
    Cognitive Interviews 13
    Respondents 14
    Results 15
  Pilot 2 17
    Respondents 18
    Analysis and Results 19
  Main Study 20
    Respondents 21
    Severity of Problem 21
    Degree of Responsibility 24
    Ethical Behavior 28
    Believability 32
  Limitations 33
  Discussion and Conclusion 34
    Summary of Findings 34
    Governance and Policy Implications 36
Chapter 4: Study Two: Interviews regarding Reactions to Online Ad Profiles 38
  Methodology 38
    Introduction to Profile Sites Used 39
    Recruitment 41
    Procedure 42
    Analysis 43
    Limitations 44
  Results 44
    Participants 44
    Reactions to inaccuracy 46
    Expressed reasons for discomfort 47
    Potential Benefits 49
  Limitations, Discussion, and Implications 50

Limitations 50
Discussion 50
Implications.. 51
Chapter 5: Conclusion and Future Work 53
Conclusion 53
Future Work 53
Appendices.. 56
Appendix A: Study 1: Survey and Question Text 56
Appendix B: Study 2: Interview Protocol.. 57
Bibliography 62

List of Tables

Table 1: Scenarios for First Pilot Study. Each participant viewed one explanation, with one targeted group filled in (...) as receiving more of the targeted ads. 13
Table 2: Cognitive Interview Demographics 14
Table 3: Variables included in the scenarios. 17
Table 4: Scenarios for Second Pilot Study. Each participant viewed one explanation, with one targeted group filled in (...) as receiving more of the targeted ads. 18
Table 5: Factors used in the regression models for problem responsibility, ethics, and believability 20
Table 6: Sample demographics. The combined column is the demographics of the total sample including both the MTurk and SSI respondents. 21
Table 7: Regression results for problem severity (n=853). 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 problem severity Likert. CI is the 95% confidence interval for the odds ratio. Statistically significant factors p< 0.05) are denoted with a *. T- stands for the race of the targeted group while R/E stands for race or ethnicity of the respondent. 23
Table 8: Regression results for advertiser responsibility (n=840), where OR>1 is associated with more responsibility. See Table 7 caption for more detailed explanation. 25
Table 9: Regression results for host responsibility (n=878), where OR > 1 associated with more responsibility. See Table 7 caption for more detailed explanation.... 26
Table 10: Regression results for ad network responsibility (n=809), where OR > 1 is associated with more responsibility. See Table 7 caption for more detailed explanation. 26
Table 11: Regression results for user responsibility (n=895), where OR > 1 is associated with more responsibility. See Table 7 caption for more detailed explanation. 27
Table 12: Regression results for ethical behavior by the advertiser (n=874), where OR > 1 is associated with stronger agreement that the advertiser behaved ethically. See Table 7 caption for more detailed explanation. 30
Table 13: Regression results for ethical behavior by the host (n=857), where OR > 1 is associated with stronger agreement that the advertiser behaved ethically. See Table 7 caption for more detailed explanation. 30
Table 14: Regression results for ethical behavior by the ad network (n=868), where OR > 1 is associated with stronger agreement that the advertiser behaved ethically. See Table 7 caption for more detailed explanation. 31
Table 15: Regression results for ethical behavior by the end user (n=867), where OR > 1 is associated with stronger agreement that the advertiser behaved ethically. See Table 7 caption for more detailed explanation. 31
Table 16: Interview participants' demographics. The columns represent the Participant ID number (coded by the interview date order), the gender of the participant, age race and education level. 45

List of Figures

Figure 1: Target Problem for Pilot 1 16
Figure 2: Responses for problem severity, broken down into behavior and demographic conditions. 22
Figure 3: Responsibility scores, per entity, broken down by the behavioral and demographic conditions. 25
Figure 4: Agreement that each entity was behaving ethically, broken down by the behavioral and demographic conditions. 32
Figure 5:Responses for scenario believability, broken down into behavior and demographic conditions. 33
Figure 6: The author's interests profile as presented by Google 40
Figure 7: The author's interests profile as presented by Facebook. 41

List of Abbreviations

Online Behavioral Advertising: OBA

Chapter 1: Introduction and Motivation

With the rise of the internet, an increasing number of companies are tracking their users activity. This tracking information can be as simple as when they entered the site, how long they were there and what, if any, links they clicked. But the company can also track other information about the consumer. This information can also lead to inferences about the consumer, such as age, race, gender, as well as behavioral traits, leaving the company with a profile of that user. How this profile is used differs for each company but the most obvious way to the user that this profile is used is through advertisements.

Online advertisers spent more than $59 billion in 2015. This was a 20% increase over the previous year [1], and digital advertising now constitutes approximately 37% of the United States' media spending [2]. This growth can be attributed to several factors, such as the increased number of digital users as well as the ability to target users [3]. Targeted advertisements can have a lot of benefits to the user (such as seeing more interesting or relevant ads), but they have raised some serious concerns, including threats to consumer privacy and the potential for discrimination. One factor to consider when trying to unravel the targeting of users is how the average user perceives and understands what information is gathered about them and how it is used.

Today, a consumer will most likely interact with several targeted advertisements daily. They could be shown that advertisement because of an email that they received, something they bought online, what websites they visited and what advertisements they clicked on [4]. When browsing on social media, ads are integrated into their experience. Users do have the option to view the profiles compiled about them on many sites, such as Google and Facebook, but it is unclear how many users take advantage of this option.

There has been an abundance of research analyzing these topics, most of which can be divided into four main areas: human perception of targeting, existing tools that explore targeting, recorded instances of discrimination and the policies that have been put in place or recommended. In chapter 2, this thesis will provide a detailed literature review of existing work in these four areas. One term that has emerged from the literature and the advertising market is online behavioral advertising (OBA).

In this thesis, we expand upon existing work by exploring end user attitudes towards inferencing and discrimination in two new ways. In the first study (Ch. 3), we consider different discriminatory scenarios and primarily gauge the participant's level of concern through surveys. In the second study (Ch. 4), we interview participants with the goal to gauge their reactions to profiles that companies have created about them, focusing on Facebook and Google.

Researchers have been able to identify thousands of unique advertisements and link them to certain profile characteristics (such as race and gender[5]–[7]). However, the research has not been able to identify the process by which these ads are placed. Research has also looked at how people react to targeted ads, but has not explicitly considered discrimination. This issue is addressed in the first study, seen in chapter 3.

The first study is designed to understand what people's perception is when presented with an advertisement that is discriminatory towards a certain race. To do this, we analyzed several different scenarios. Several variables were changed in the scenarios (the entity-making decisions, the race of the targeted individual, and if the decider was human or an algorithm). The survey asked how much of a problem their given scenario was, the responsibility level of each entity involved and how believable the scenario was. The respondents consistently rated that when behavior was used to determine who would see the ad being placed it was less problematic than when demographics was used. However, there did not seem to be any difference in the level of problem when it came to the different entities involved in placing the advertisement.

For the second study, we examined the (previously under-explored) question of how end users react to their real inferenced data, as reported by Google and Facebook. By contrast, prior work often focuses on hypothetical scenarios[8]. In this study, however, we can observe and have a conversation about the end user's reaction to their live data, which was often the first time that it was seen.

The first study gives insight into how people react to the idea of discriminatory advertisements. While the second study gives insight into the end user's own reaction to the inferenced data about them. Taken together, we can learn overall about people's reaction to how their information is tracked over time and the kinds of inferences that can be made from their collected data. Together, these insights will be able to inform policy, give recommendation for educating consumers and give recommendations to business that might influence their targeting behavior.

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 [9] [10].

Knowledge about people's attitudes could also help advocates for algorithmic fairness to understand how to focus their public awareness efforts. Finally, data about consumer attitudes could 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.

This thesis is broken down into four more chapters. First, the related work to this project. The next two chapters are the studies performed. To close, the last chapter will go into future work, lessons learned and broad conclusions to be taken from the two studies.

Chapter 2: Background and Related Work

There are four main areas of existing work that have contributed to this research: (1) human understanding and perception of online tracking, (2) existing tools that have been built to inspect different aspects of online advertising and their findings, (3) existing policy and policy recommendations, and (4) the discrimination found online and recommendations to help prevent this in the future. This section will present an overview of the existing work that falls into those four categories and influenced the creation, design, and analysis of the studies mentioned in the next two chapters.

Human Perception

Since one of the most common ways for a user to see how their inferenced data is being used is through ads, most related work in this area is focused on how the end user understands OBA and the level of impact that OBA has on their lives. This subsection will give an overview of some of the existing studies that have looked at human's reaction to, perception of, and understanding of OBA and similar practices. The studies mentioned below are organized by related topics and their impact on this thesis.

One such study was done in 2013 by Agarwal et al. [11]. In their study, they interviewed 53 individuals to gather their reactions to third-party tracking situations. During this interview, the participants were shown a video of the process by which OBA and third-party tracking takes place (in about 9 minutes). The participants expressed concerns about embarrassing or intimate advertisements shown to them. They also expressed concern about the amount and type of information that advertisers were able to gain about them. In Chapter 4, we expand on this and look at people's perception of real-life predictions made about them.

Ur et al.[4] in 2012 found through 48 semi-structured interviews that there seemed to a substantial mismatch between what the average user understands about OBA and the approaches that are taken to inform the user about OBA. These misinterpretations were as simple as misunderstanding the purpose of an icon to the entities involved in the process of advertising. Additionally, they found that the users felt that targeted advertising can be beneficial to them but also privacy invasive. We, however, build on this by having the participant look at their own profile to better understand their attitude and understanding of targeted inferencing.

Another study concerned people's reactions to existing user plug-ins, namely ad-blockers. Leon et al. [12] interviewed 45 individuals about their understanding of and attitudes toward OBA, and they had the participants interact with existing internet privacy tools (such as opt-out tools). The researchers found serious flaws in all nine systems that were tested. The issues ranged from the user interface to ineffective communication, which negatively affected the user's ability to properly use the tool. They concluded that lack of knowledge about the OBA infrastructure prevents most users from using the tool properly. Regarding both studies, we wish to contribute to the gaps of understanding users have and give recommendations to help fill those gaps.

A similar study done in 2016 examined people's perception of OBA and how it influences their lives. Coen et al. [13] conducted a nearly 800-person survey of how people viewed the use of different aspects of the individual through provided, correctly inferred, and incorrectly inferred attributes of a person in ad, search, and pricing results. Certain aspects, such as race and household income, were viewed more negatively than others. The authors made policy recommendations as to how and what information is used. The survey in this thesis, discussed in Chapter 3, runs along similar lines but expands to look at the different entities involved in the advertising pipeline but expands upon it to look at the discrimination.

Another interview study conducted by Malheiros et al. [8] examined different levels of personalization and the reactions that people had to them. They interviewed 30 participants and showed them a travel website that had content ranging from generic to the participant’s photo or name. The researchers found the participants were indeed more likely to notice the ads that were unique to them, but also viewed these ads as more uncomfortable and inappropriate.

In contrast, Grossklags and Acquisti found that given the right incentive, people were more willing to share their information with companies. In a two-part study, they looked at if a user would share their personal information for a set amount of money and also how much each participant was willing to spend to keep their information private. The researchers found the participants, on average, had a much higher preference for money (even 25 cents) than for data privacy. This relates to the second study of this thesis because it sheds light on the mindset that most users have when it comes to data privacy.

Warshaw et al.’s [14] interviews found that high school educated individuals did not believe companies could make in-depth analyses about them. They also found there were two main participant subgroups. The first subgroup believed that most of the targeting was based on stereotyping; whereas the other group subgroup believed the targeted advertisements were based on straightforward intuitions. When looking at this population, there seems to be limited understanding of how inferencing works. We build on this by (a) measuring reactions when the effects of inferencing are made clear (study 1) and (b) examining reactions to learning more about inferencing for themselves personally (study 2).

Finally, Warshaw et al.[15] did another interview study that had participants look at hyper-personal attributes about them. They found that the participant was very sensitive about these traits and did not wish to share them with the researcher. However, the researchers found that most participants got over that feeling and did end up sharing, even if they weren't comfortable with doing so. These findings both influenced the structure and content of the interview study conducted for this paper's research. By recognizing the intimate nature that the profiles might be, we were mindful to let the participant have control over the computer and to only share what he/she were comfortable sharing.

Tools for Measuring Online Tracking and Advertising

Several "black box" tools have been created by researchers with the hope of better understanding what happens behind the scenes regarding online advertisements. Corresponding a user's input and the given advertisements has provided several insights into how and where information is used.

One such tool is XRay, created in 2014 by Lécuyer et al. [6]. This tool gives the user insight into how his or her personal data is used on the web. Given a user's account, this tool identifies which of the attributes are used to predict the shown outputs (in most cases, advertisements). The tool used similar, but not identical, accounts to make predictions about the output based on the unique input. By building on the idea of giving user insight into how their personal data was used, we designed the second study to show the attributes and interests that companies inferred about them.

Another tool is AdScape, also developed in 2014, also gave insight into what information is used to influence the advertisements that are shown to a user. Barford et al. [7] built a tool that scraped the web and gathered nearly 200,000 district ads. They found there were about 4,000 district advertising entities from a variety were responsible for those ads. Furthermore, through these ads, OBA is commonly used but not as widespread as often thought. On the other hand, Barford et al. did note that advertisements were more likely to vary based on a user's profile than website content. This tool, like XRay, influenced the creation of both parts of this thesis' studies.

A framework tool put forward by Carrascosa et al. [16] found that OBA is prevalent in advertisements online. By training online personas that were like simplified human personas, the researchers could identify the targeted advertisements and those attributes of the highest value to advertisers. Another finding was that sensitive characteristics, such as health and religion, have been used as attributes in deciding the advertisement shown to the user (even though government regulation bans this). This work influenced the first study by using potentially sensitive attributes used in the different scenarios.

Another set of guidelines put forward by Guha et al. [17] in 2010 looked at how to measure OBA. The researchers examined a snapshot of time and took a long-term look at the prevalence of OBA. Their study brought out the lack of transparency in the process by which targeted advertisements are produced. This lack of transparency has influenced the creation of the survey study. We included multiple entities in the different scenarios that were involved in the advertisement process and asked the participant's reaction to their level of responsibility.

Discrimination in Targeted Advertising

Targeting advertisements has been a commonplace practice for over 100 years[18]. Research has shown there are occurrences of targeting ads based on a certain aspect of a person, such as race and gender. This subsection will give an overview of the occurrences that have been recorded.

One such study by Sweeney [19] in 2013 found that the advertisements presented to the user after a search of a name depended on the predicted race of that name. Names that were predicted to be black were more likely to be shown an advertisement for finding an arrest record than names that were white. On average, black names were shown arrest records 25% more of the time than white names. Sweeney did note a few exceptions to this rule, such as the name Dustin.

AdFisher, a tool created by Datta et al. [20], looked at different advertisements and how they were assigned to different demographic groups. One interesting observation was that an ad for a high paying executive position was shown nearly six times as often to a male than a female. When considering how this happened, it became clear there was little accountability and little known about how the online advertising infrastructure works. This result directly feeds into the first study, where we looked at whom the public thought responsible.

Existing Policy on Online Tracking

Lawmakers and policy advisors have made some headway into how the legal side of OBA should be handled. Most contemporary work shows that the user needs to be better educated and informed about how different organizations execute OBA. Another issue is the lack of regulation and consistency.

One recommendation put forward by Mayfield et al. in 2015 hopes to bring more options to consumers (Internet users) by allowing them to opt in or out. They argued that the "choice and notice" option is not a good choice because all responsibility lies with the consumer. Consumers are supposed to understand their choices and make decisions based on their knowledge. The authors pointed out that the average user does not have the knowledge necessary to make informed choices, which allows many corporations to take and use their information without regard. In Chapters 3 and 4, we can see that people do not have a good understanding of the inner working of the advertising infrastructure.

In 2009, the Federal Trade Commission (FTC) overhauled its principles on OBA. These principles cover how first and third-party tracking should (or should not) take place. The FTC also introduced a much stronger self-regulatory initiative to protect consumer privacy and interests. They also released a website about how the consumer can protect him or herself on the Internet.

Chapter 3: Study One: Survey regarding Online Discrimination

There were three main components that went into this first study: two pilots and a confirmatory study. The surveys used each dealt with a scenario that had a type of discrimination. Through the studies, we were able to narrow the focus of from a broad look at many different types of discrimination to a more concise set of scenarios. This chapter will explore the methodology and results of each of these studies.

Pilot 1

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.

Scenarios

As described in the previous section, 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 64 total scenarios, consisting of combinations eight groups of people who were the *target* of discriminatory ads (e.g., saw the job ad more frequently), and eight *explanations* for the targeting decision. These explanations were drawn in part from suggested explanations posited by the authors of an ad-discrimination measurement study, and were intended span a range of both real-life plausibility and discriminatory intent [20]. The targets and explanations used in Pilot-1 are listed in Table 1.

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:

- 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 First Pilot Study. Each participant viewed one explanation, with one targeted group filled in (...) as receiving more of the targeted ads.

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.

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 pre-tested the wording of our explanations and subsequent questions using {\it cognitive interviews}, a standard technique for evaluating the intelligibility and effectiveness of survey questions by asking respondents to think aloud while answering the survey questions~\cite{cogInterview1}. We conducted eight in-person cognitive interviews with respondents from a variety of demographic groups (Table 2). Because of these interviews, we made the scenario descriptions more narrative, clarified the wording of some questions, and added the question about believability.

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

Table 2: Cognitive Interview Demographics

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 considering prior work suggesting that people with less educational attainment have important misconceptions about targeted advertising~\cite{Warshaw:2016wz}. 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, with slightly more educated individuals. Between 15 and 16 respondents were assigned to each of the 72 scenarios.

Results

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

First, we considered the issue of who was targeted in the scenario. That is, which group of people benefited or was short-changed by the discriminatory advertising. We found that the scenarios that targeted race were more controversial (elicited a wider range of ratings regarding whether the scenario was problematic, on a four-point Likert scale) than the other targets (e.g. political affiliation, health condition) that we considered (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.

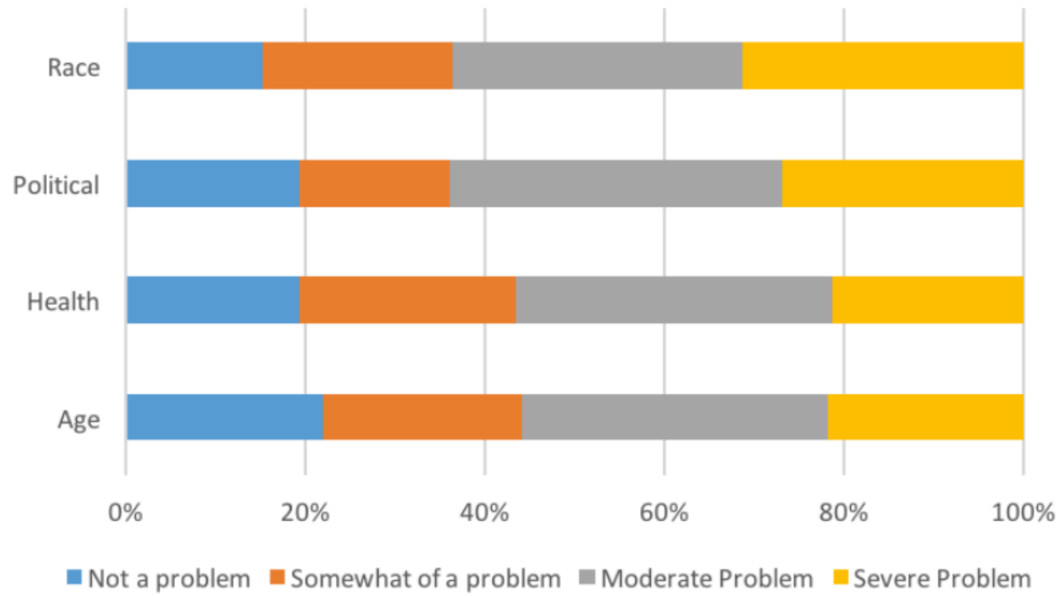

Figure 1: Target Problem for Pilot 1

Second, we considered respondents' responses regarding the severity of the various scenarios. The most noticeable pattern was that scenarios that targeting based on behavior (e.g. browsing history), rather than their explicit demographics, were generally rated less problematic.

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 strong evidence for this (MWU test resulted in $p=0.095$), but decided to include it in our subsequent studies in hopes of confirming (or not) its lack of importance.

Pilot 2

Based on the results from Pilot 1, we designed a follow-up survey to explicitly test a few concrete variables related to targeting explanations. We decided to contrast 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 added a factor locating the decision making either at Systemy (the company placing the ad) or Bezo (the ad network). The final set of 24 scenarios (three targets *x* demographic vs. behavioral *x* human vs. algorithmic *x* two entities) is detailed in Table 3.

Target	**Mechanisms**	**Decider**	**Entities**
White	Behavior	Human	Advertiser
Asian	Demographics	Algorithm	Ad Network
Black			

Table 3: Variables included in the scenarios.

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 people of other races.*

The explanations shown to the respondents can be found in Table 4.

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 4: Scenarios for Second Pilot Study. Each participant viewed one explanation, with one targeted group filled in (...) as receiving more of the targeted ads.

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

Respondents

The goal of Pilot 2 was to create training data that we could use to test a variety of potential regression models, without worrying about erosion of statistical confidence due to multiple testing. For this purpose, we considered it sufficient to test a smaller, somewhat less diverse-and also less expensive-sample. We deployed our four- to five-minute survey to 192 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 [21][22]. 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. To avoid duplicate respondents, each participant's unique MTurk identification number was recorded and duplicate ids were prevented from completing the survey again.

Perhaps surprisingly, we noted a higher rate of thoughtful responses to our free-response question in the MTurk sample than in the SSI sample from Pilot 1.

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 the target and explanation, as well as demographic factors, correlate with respondents' reactions to the presented scenario. For the degree of responsibility and problem questions, we generated an initial model including as covariates the targets and scenarios from Table 3; participant demographic factors 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) [24]. Multiple models were very close in AIC value; we selected a final model that included the three variables of interest (mechanism, decider, entity) and was near-

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

minimal AIC for each of the five questions. The factors used by the final model are shown in Table~\ref{tab:RegressionExplained}.

Factor	Description	Baseline
Age	Of participant. Continuous.	n/a
Ethnicity	Of participant. White, Black, Hispanic or Latino, Asian, or Other	White
Education	Of participant. High school diploma or less, Some college (HS+), Bachelor's Degree and up (BS+)	High school or less
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
Entity	Entity making the decision: Either the ad network or the advertiser.	Ad network
Decider	Whether the targeting decision was made by an algorithm or a human.	Algorithm
Sample Provider	Amazon's Mechanical Turk and SSI	MTurk

Table 5: Factors used in the regression models for problem responsibility, ethics, and believability.

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

Main Study

Finally, we conducted a confirmatory study to test the regression model developed during Pilot 2. We deployed the same survey as in Pilot 2. To promote both high data quality and broad generalizability in our results, 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 5).

Respondents

We collected responses from 534 MTurk respondents and 390 SSI respondents, for a total of 924. Demographics for the two samples are shown in Table 6 with U.S. Census data for comparison [25]. By collecting this large representative sample through two different sample providers, MTurk and SSI, we will be able to draw generalizable conclusions pertaining to perceptions people have about OBA.

Metric	SSI	MTurk	Total	Census
Male	42.3%	49.9%	47.3 %	48.2%
Female	57.4%	50.1%	52.6%	51.8%
Caucasian	63.3%	82.7%	75.9%	64%
Hispanic	12.3%	6.4%	7.0%	16%
Asian	5.6%	4.8%	5.2%	5.4%
African American	17.2%	4.1%	10.2%	12%
Other	1.5%	1.9%	1.8%	2.6%
up to H.S.	31.5%	11.8%	18.7%	41.3%
Some college	35.9%	33.3%	34.2%	31.0%
B.S. or above	32.6%	54.9%	47.1%	27.7%
18–29 years	20.5%	27.0%	23.9%	20.9%
30–49 years	41.2%	55.6%	49.3%	34.7%
50–64 years	31.0%	15.3%	23.0%	26.0%
65+ years	6.7%	2.2%	3.7%	18.4%

Table 6: Sample demographics. The combined column is the demographics of the total sample including both the MTurk and SSI respondents.

The 20 respondents who reported their race as 'other' were excluded from the dataset, because the small sample frequently prevented the regression model from converging.

Severity of Problem

Respondents were asked, on a four-point scale from “not a problem'” to “a serious problem”, to rate how problematic the found the discrimination scenario with which they were presented. The ordering and phrasing of the scale was based on Clemson's Likert scale [26]. Overall, respondents gave a median rating of 'minor problem' (2) to scenarios in which the discriminatory OBA 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) were a minor problem; while they gave a median rating of 'moderate problem' (3) to discriminatory OBA scenarios that occurred due to direct demographic targeting. Figure 2 provides an overview of the scores. Additionally, respondents gave a median rating of 'minor problem' (2) to both the scenarios in which a human decided to target the advertisements and those in which an algorithm decided on the targeting.

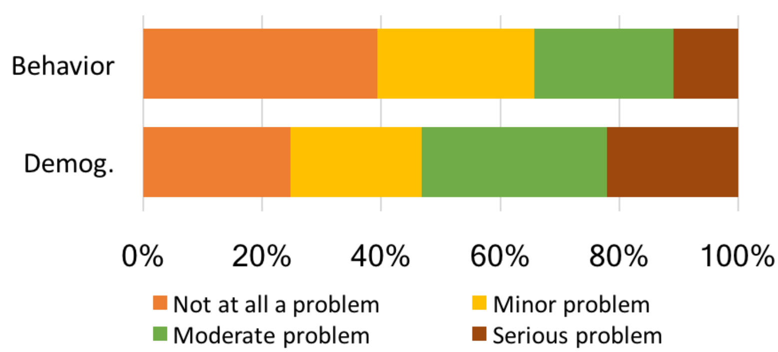

Figure 2: Responses for problem severity, broken down into behavior and demographic conditions.

In order to gain more insight into the factors that influence respondents' perceptions of OBA, we conducted a regression analysis (results shown in Table 7). Based on our 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. their demographics). Behavior based ad targeting was 48\%

as likely as demographic-based targeting to increase respondents' rating of the severity of the scenario. 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.

Factor	OR	CI	*p*-value
SSI	1.58	[1.14, 2.19]	0.006*
T-Asian	0.55	[0.39, 0.78]	<0.001*
T-Black	0.58	[0.41, 0.82]	0.002*
Behavior	0.48	[0.36, 0.63]	<0.001*
Advertiser	0.93	[0.71, 1.23]	0.623
Human	1.05	[0.8, 1.38]	0.718
Age of Participant	0.99	[0.98, 1]	0.022*
HS+	1.62	[1.08, 2.43]	0.019*
BS+	1.41	[0.96, 2.08]	0.081
R/E-Asian	1.24	[0.66, 2.33]	0.506
R/E-Black	2.39	[1.38, 4.15]	0.002*
R/E-Hispanic or Latino	1.40	[0.78, 2.53]	0.259

*Table 7: Regression results for problem severity (n=853). 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 problem severity Likert. CI is the 95% confidence interval for the odds ratio. Statistically significant factors p< 0.05) are denoted with a *. T- stands for the race of the targeted group while R/E stands for race or ethnicity of the respondent.*

Respondents also found targeting black and Asian individuals for more job ads less problematic (58% and 55% as likely to increase severity rating, respectively) than targeting white individuals. Figure 2 illustrates the problem severity scores for certain subsets of our sample.

On the other hand, 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. Who was 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 7} shows that older respondents are associated with lower severity ratings; for example, a 10-year age gap is associated with only 82% ($0.98^{10} = 0.82$) likelihood of increased severity. Black respondents were 2.39X as likely as baseline white respondents to rate the problem as more severe. Results for education level were mixed, so we do not interpret this result. Finally, respondents recruited through SSI were 2.58X more likely to increase one step in severity, even when controlling for age and ethnicity.

Degree of Responsibility

We next consider the responsibility level respondents assign to different entities involved in the OBA scenario: the user, the ad network (Bezo Media), the advertiser (Systemy), and the local news website on which the advertisement was displayed. Respondents provided their responsibility ratings on a four-point scale from not at all responsible, somewhat responsible, mostly responsible, to completely responsible.

64% of respondents rated the user as 'not at all responsible' (1) for the outcome of the OBA scenario (median Likert score = 1). Respondents also did not attribute a high level of responsibility to the local news network: the median responsibility score for the local news network was 2, with 41% of respondents rated the local news website 'not at all responsible' (1). On the other hand, only 15% and 17%, respectively, of respondents rated the ad network and the advertiser 'not at all responsible' (1); with the median score for the ad network a 'moderate problem' (3) and the median rating of responsibility for the advertiser a 'minor problem' (2). Respondents' ratings of the responsibility of each entity is shown in Figure 3.

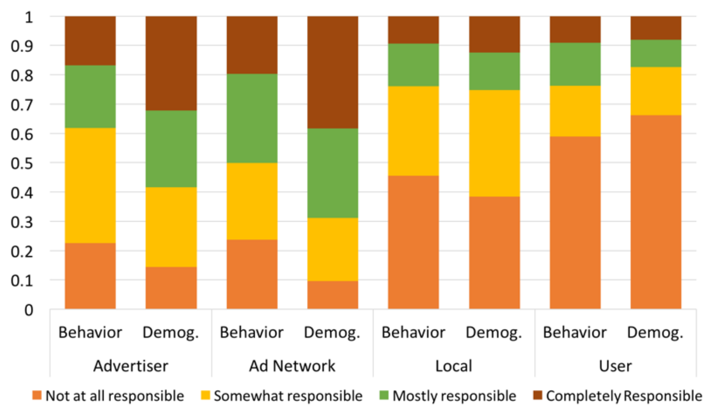

Figure 3: Responsibility scores, per entity, broken down by the behavioral and demographic conditions.

We also conducted a regression analysis to determine what factors influenced respondents' ratings of responsibility for each of these entities. Tables 8-11 illustrate the results of the regressions for each entity.

Factor	OR	CI	*p*-value
SSI	1.03	[0.71, 1.49]	0.869
T-Asian	1.56	[1.05, 2.33]	0.028*
T-Black	1.08	[0.74, 1.57]	0.699
Behavior	0.54	[0.39, 0.75]	<0.001*
Advertiser	2.15	[1.55, 2.98]	<0.001*
Human	0.75	[0.55, 1.04]	0.082
Age of Participant	1.00	[0.99, 1.01]	0.851
HS+	0.80	[0.5, 1.27]	0.342
BS+	1.02	[0.64, 1.63]	0.925
R/E-Asian	2.05	[0.84, 4.99]	0.114
R/E-Black	1.51	[0.82, 2.78]	0.189
R/E-Hispanic or Latino	1.14	[0.6, 2.17]	0.68

Table 8: Regression results for advertiser responsibility (n=840), where OR>1 is associated with more responsibility. See Table 7 caption for more detailed explanation.

Factor	OR	CI	*p*-value
SSI	1.96	[1.46, 2.64]	<0.001*
T-Asian	0.87	[0.64, 1.18]	0.365
T-Black	1.02	[0.75, 1.39]	0.899
Behavior	0.64	[0.5, 0.83]	<0.001*
Advertiser	1.03	[0.8, 1.32]	0.836
Human	1.02	[0.8, 1.32]	0.851
Age of Participant	0.98	[0.97, 0.99]	<0.001*
HS+	0.74	[0.51, 1.09]	0.124
BS+	0.80	[0.56, 1.16]	0.246
R/E-Asian	1.46	[0.79, 2.67]	0.225
R/E-Black	1.12	[0.72, 1.76]	0.611
R/E-Hispanic or Latino	0.99	[0.59, 1.66]	0.972

Table 9: Regression results for host responsibility (n=878), where OR > 1 associated with more responsibility. See Table 7 caption for more detailed explanation.

Factor	OR	CI	*p*-value
SSI	0.92	[0.62, 1.38]	0.693
T-Asian	1.69	[1.09, 2.61]	0.02*
T-Black	1.20	[0.79, 1.81]	0.386
Behavior	0.31	[0.21, 0.45]	<0.001*
Advertiser	0.40	[0.28, 0.58]	<0.001*
Human	1.21	[0.85, 1.72]	0.283
Age of Participant	0.98	[0.96, 0.99]	0.001*
HS+	0.92	[0.56, 1.52]	0.745
BS+	1.36	[0.82, 2.23]	0.232
R/E-Asian	1.13	[0.49, 2.58]	0.778
R/E-Black	3.10	[1.37, 7.05]	0.007*
R/E-Hispanic or Latino	1.45	[0.67, 3.13]	0.345

Table 10: Regression results for ad network responsibility (n=809), where OR > 1 is associated with more responsibility. See Table 7 caption for more detailed explanation.

Factor	OR	CI	*p*-value
SSI	2.55	[1.89, 3.45]	<0.001*
T-Asian	0.93	[0.67, 1.28]	0.643
T-Black	0.90	[0.65, 1.24]	0.519
Behavior	1.38	[1.06, 1.8]	0.018*
Advertiser	1.21	[0.93, 1.58]	0.157
Human	1.13	[0.86, 1.47]	0.375
Age of Participant	0.96	[0.95, 0.98]	<0.001*
HS+	0.87	[0.59, 1.28]	0.48
BS+	0.89	[0.61, 1.31]	0.563
R/E-Asian	1.56	[0.87, 2.81]	0.134
R/E-Black	1.70	[1.1, 2.64]	0.017*
R/E-Hispanic or Latino	1.30	[0.78, 2.17]	0.318

Table 11: Regression results for user responsibility (n=895), where OR > 1 is associated with more responsibility. See Table 7 caption for more detailed explanation.

For all entities, the way in 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 69% more likely to rate the ad network as responsible for the OBA when demographic targeting rather than behavioral targeting is used. This effect reverses when respondents determine the user's level of responsibility: respondents assign greater responsibility to the user in the behavioral case. While this makes some sense--the behavioral case is linked to the user's web browsing behavior--the discriminatory targeting can also be seen as a function of many people's behavior, rather than one specific end user who sees an ad.

As might be expected, responsibility aligns with the details provided in the scenarios seen by the respondents': the advertiser gets more responsibility when the scenario they were shown implicated 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 each case, 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.

Interestingly, unlike problem severity, the targeted racial group and the race of the participant appear to have little impact on responsibility assignment in most cases. We note that all 20 respondents who identified their race as ``other'' in the survey assigned high responsibility to the ad network, which prevented the regression model from properly converging. As a result, we removed those 20 people from the regression shown in Table 10.

Ethical Behavior

Next, we consider respondents' responses 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 disagree to strongly agree. Across all scenarios, 76% of respondents agreed or strongly agreed that the user behaved ethically (median Likert score = 2 (agree)). 58% of respondents also reported that the local news network behaved ethnically (median=2), we note that these ratings follow a trend similar to that observed in the responsibility ratings. Contrary to the prior trend observed with responsibility ratings, however, 49% of respondents agreed or strongly agreed that the advertiser behaved ethically (median=3 (neutral)) and 40.5% agreed or strongly agreed that the ad network behaved ethically (median=3).

The regression analyses for ethical behavior are shown in Tables 12-15. Consistent with the findings from previous questions, the mechanism of targeting is significant for all four entities; in every case behavior-based targeting is significantly correlated with a higher perception of ethical behavior than the demographic-based targeting. This is illustrated in Figure 4. Human vs. algorithmic decision making continues to show no significant effect. As with responsibility, there is a predictable connection to the entity making the decision in the scenario: the advertiser is viewed as less ethical when it is named in the scenario than when the ad network is named, and vice versa. The ad network and the advertiser are also perceived as behaving more ethically when Asian or black people see more job ads than when white people are favored.

Factor	OR	CI	*p*-value
SSI	2.55	[1.89, 3.45]	<0.001*
T-Asian	0.93	[0.67, 1.28]	0.643
T-Black	0.90	[0.65, 1.24]	0.519
Behavior	1.38	[1.06, 1.8]	0.018*
Advertiser	1.21	[0.93, 1.58]	0.157
Human	1.13	[0.86, 1.47]	0.375
Age of Participant	0.96	[0.95, 0.98]	<0.001*
HS+	0.87	[0.59, 1.28]	0.48
BS+	0.89	[0.61, 1.31]	0.563
R/E-Asian	1.56	[0.87, 2.81]	0.134
R/E-Black	1.70	[1.1, 2.64]	0.017*
R/E-Hispanic or Latino	1.30	[0.78, 2.17]	0.318

Table 12: Regression results for ethical behavior by the advertiser (n=874), where OR > 1 is associated with stronger agreement that the advertiser behaved ethically. See Table 7 caption for more detailed explanation.

Factor	OR	CI	*p*-value
SSI	0.62	[0.28, 1.41]	0.254
T-Asian	1.73	[0.69, 4.33]	0.239
T-Black	1.32	[0.56, 3.11]	0.528
Behavior	5.94	[2.23, 15.78]	<0.001*
Advertiser	0.58	[0.28, 1.23]	0.159
Human	0.91	[0.44, 1.91]	0.81
Age of Participant	1.01	[0.98, 1.04]	0.706
HS+	1.66	[0.62, 4.41]	0.313
BS+	1.55	[0.6, 4.03]	0.369
R/E-Asian	0.32	[0.08, 1.21]	0.092
R/E-Black	0.61	[0.21, 1.75]	0.357
R/E-Hispanic or Latino	1.19	[0.26, 5.42]	0.826

Table 13: Regression results for ethical behavior by the host (n=857), where OR > 1 is associated with stronger agreement that the advertiser behaved ethically. See Table 7 caption for more detailed explanation.

Factor	OR	CI	*p*-value
SSI	1.09	[0.66, 1.79]	0.748
T-Asian	1.90	[1.13, 3.19]	0.016*
T-Black	2.27	[1.32, 3.9]	0.003*
Behavior	8.73	[4.67, 16.31]	<0.001*
Advertiser	2.33	[1.46, 3.7]	<0.001*
Human	0.75	[0.48, 1.17]	0.201
Age of Participant	1.00	[0.98, 1.02]	0.796
HS+	1.01	[0.55, 1.86]	0.962
BS+	1.39	[0.75, 2.58]	0.296
R/E-Asian	0.39	[0.15, 0.98]	0.046*
R/E-Black	0.48	[0.25, 0.93]	0.029*
R/E-Hispanic or Latino	1.18	[0.47, 2.96]	0.723

Table 14: Regression results for ethical behavior by the ad network (n=868), where OR > 1 is associated with stronger agreement that the advertiser behaved ethically. See Table 7 caption for more detailed explanation.

Factor	OR	CI	*p*-value
SSI	0.09	[0.02, 0.38]	<0.001*
T-Asian	2.70	[0.58, 12.51]	0.203
T-Black	0.90	[0.25, 3.2]	0.873
Behavior	2.11	[0.66, 6.69]	0.207
Advertiser	0.40	[0.12, 1.34]	0.137
Human	6.96	[1.42, 34.06]	0.017*
Age of Participant	1.06	[1.01, 1.11]	0.018*
HS+	1.41	[0.29, 6.92]	0.674
BS+	0.84	[0.21, 3.32]	0.806
R/E-Asian	0.54	[0.06, 5.13]	0.589
R/E-Black	0.51	[0.14, 1.89]	0.313
R/E-Hispanic or Latino	2.06	[0.23, 18.32]	0.516

Table 15: Regression results for ethical behavior by the end user (n=867), where OR > 1 is associated with stronger agreement that the advertiser behaved ethically. See Table 7 caption for more detailed explanation.

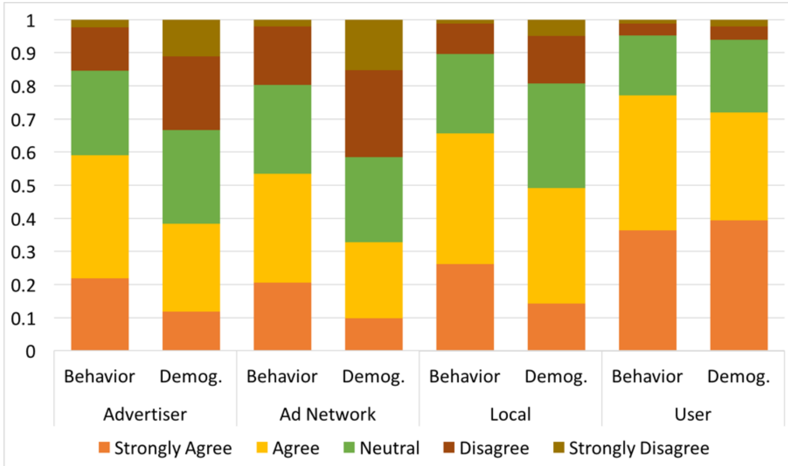

Figure 4: Agreement that each entity was behaving ethically, broken down by the behavioral and demographic conditions.

There are some mixed results on the effects of respondent demographics: respondents who are black are less likely (42%) to provide high ethical scores for the advertiser, and respondents who are black (48%) and respondents who are Asian (39%) provide lower ethical scores than white respondents regarding the ad network.

Similar to the prior scenarios, older respondents are 1.06X more likely than younger respondents to believe the end user who viewed the ad acted ethically. Oddly, respondents recruited by SSI are significantly 9% less likely to believe the end user acted ethically; we have no immediate explanation for this phenomenon.

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. 88.4% of respondents reported that the scenario 'definitely' or 'probably' could happen, thus we feel confident in the validity of their responses. Figure provides an overview of respondents' ratings of scenario believability.

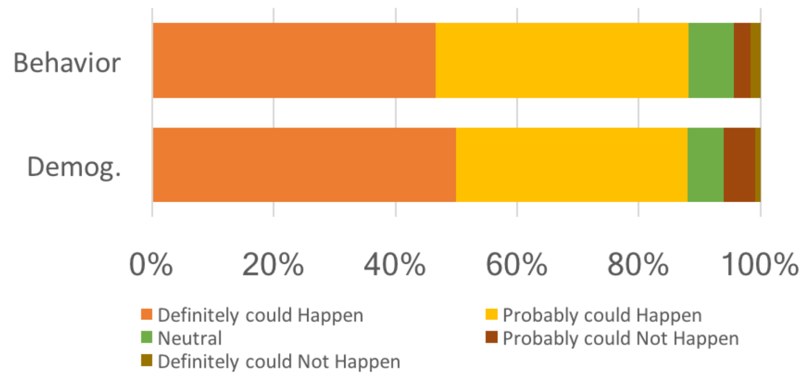

Figure 5:Responses for scenario believability, broken down into behavior and demographic conditions.

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 sample. While we believe our conclusions can generalize somewhat, they do reflect the fact that 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.

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.

Discussion and Conclusion

Below, we present a summary of our findings, implications for governance and policy guidelines for OBA, and suggestions for future work.

Summary of Findings

Overall, we find that people's perceptions of discriminatory OBA scenarios depend on how the discrimination occurred and who was discriminated against. Perhaps unsurprisingly, 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 found scenarios in which minorities (in our scenarios people of black or Asian race) were benefited by OBA discrimination less problematic and more ethical than scenarios in which the majority was benefited.

Relatedly, we also find that black respondents are more likely to view discriminatory scenarios as a problem and as unethical. We hypothesize that these ratings are influenced by the relatively wide acceptance of “affirmative action” - a U.S. educational and workplace policy that gives preference to racial minorities[27] - in the U.S., where we recruited our survey respondents.

Surprisingly, we find that the entity causing the discrimination (e.g. the ad network vs. the advertiser) did not influence respondent's ratings of the severity of the scenarios. This suggests that it is not helpful for entities to ``pass the blame'' as it is the mechanism and beneficiaries of discrimination with which users are truly concerned. Finally, we were also surprised to find that whether it was a person or an algorithm responsible for selecting how and whom to target made no difference in respondents ratings of the severity of the scenario.

Overall, we find that respondents did not hold the user responsible for the outcome of these scenarios, and the majority did not hold the local news site on which the ads were placed responsible, either. Respondents did hold the ad network and the advertiser responsible, although this placement of responsibility did not translate into a perception of unethical behavior (the median `ethics rating' for both the ad network and the advertiser was neutral).

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.

Governance and Policy Implications

While 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 OBA [28]–[30]. Of these recommendations, only the EFF policy document mentions discrimination as potential, unethical consequence of OBA. Our results, as well as the findings of X and Y who brought to light the prevalence of OBA discrimination, highlight the importance of discrimination as an OBA consideration. We find that 43\% of respondents rated our discriminatory OBA scenarios a significant or moderate problem. Indeed, respondents were concerned even when the discrimination happened as a result of targeting based on users' web browsing history (34.2%). The high percentage of respondents (88%) who were confident that our proposed scenarios could occur, further bolsters the argument that users care about discrimination in OBA. Thus, we propose that guidelines, especially those issued by government bodies such as the FTC should include explicit language about discrimination.

Further, our findings suggest that respondents are most concerned with the outcome of the scenario, not who was responsible. This suggests that responses such as Google's, when they were confronted about a higher number of job ads shown to men over women[31] are not productive for improving public perception. Thus, the websites hosting advertisements, the advertising networks (if separate from the hosts) and the clients wishing to advertise should work together to avoid discriminatory outcomes. Consequently, it may be beneficial to develop a single set of guidelines for ethical behavior in OBA, explicitly encouraging cooperation to comply with these guidelines.

Chapter 4: Study Two: Interviews regarding Reactions to Online Ad Profiles

The second study involved interviewing participants to gain insight into their reactions and understanding of Facebook and Google's interest profiles about them as a user. By conducting 15 semi-structured interviews, themes such as discomfort of incorrectly inferred data and mixed responses to level of accuracy reported by user emerged.

In this chapter, we will introduce the profiles used in the interviews, describe the methodology, and themes that surfaced during the interviews. We will then close with a discussion of the results and implications.

Methodology

In this subsection, we will explore semi-structured interviews conducted in the MC2 space between the April and June of 2017. These interviews were scheduled and performed on an ongoing basis until new themes stopped emerging, which happened at about 15 interviews. This number is in sync with the literature about qualitative interviews, which recommends about 12-20 interviews to generalize results [32]. Since we are in the middle of that recommended range, we feel as though the themes pulled from the interviews can represent a broad group of individuals.

All aspects of this research were approved by the University of Maryland's research review board (IRB).

Introduction to Profile Sites Used

Two of the most common websites[2], Facebook and Google, allow their users to view a profile that is inferred about them based on their activities. In this section, we will explore those profiles by looking at the interface, text given to the user and how to get to these sites.

Google

To navigate to Google's 'Ad Personalization' page, there are two main ways: by searching: google ad preference or by navigating through the setting page and clicking on 'Ads Settings' and on 'Manage Ads Settings.[3]' For the purposes of the interviews, the site was up on the computer and the participant just had to login. After login, they were directed to the ad profile page.

Once on the page, the top banner reads: "Make the ads you see more useful to you: Control the information Google uses to show you ads." There are several informational pulldown menus that give the user more information such as what information Google give to their partners they responded: "Google does not give our partners information you provide us that personally identifies you, such as your name, email, or billing information, unless you ask us to. We never sell your personal information.[4]"

Further down the page the collection of interests inferred about you are presented, as seen in Figure 6.

[2] http://www.alexa.com/topsites/countries/US

[3] This is the path that is taken as of June 2017, but this is subject to change at any time. This might also slightly differ depending on the version Google is using.

[4] https://www.google.com/settings/u/0/ads/authenticated

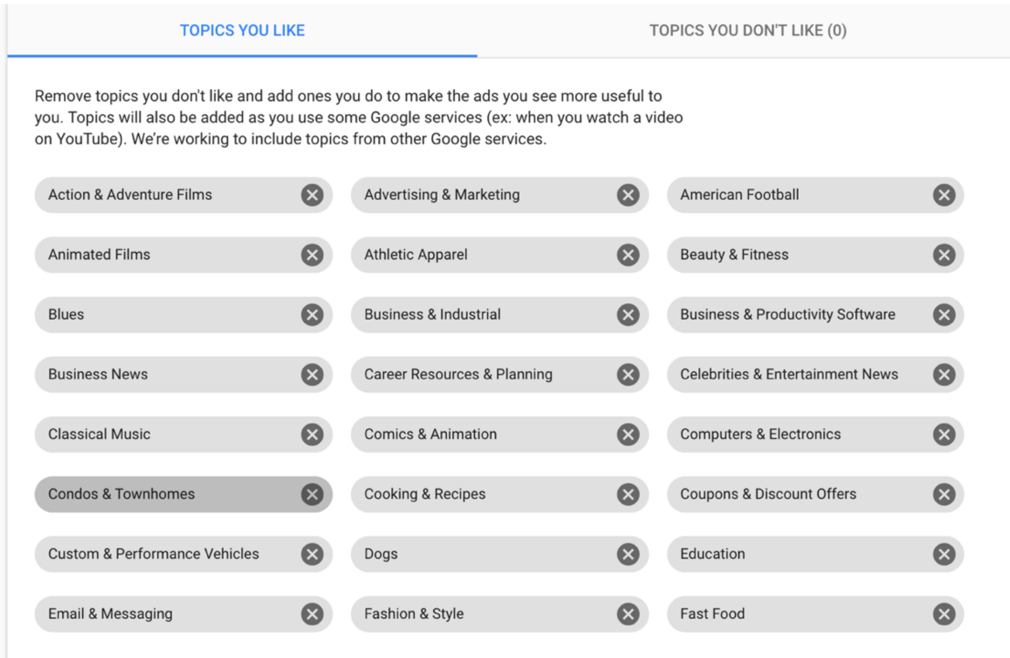

Figure 6: The author's interests profile as presented by Google

On Google, there is no indication on this page of where these topics come from, only how they are used in the Google ecosystem and the partnering sites.

Facebook

Unlike Google, the only way to navigate to Facebook's profile is through the Facebook homepage. First, the user must go to the dropdown menu at the top of the page, indicated by a down arrow and click on settings. From there, on the left-hand side of the screen there will be an option such as 'Ads', 'Adverts', or 'Advertisements' (these differ user to user). The participant in interviews was directed to the profile site after they logged in.

Once one that page, the banner reads: "Your Advert Preferences: Learn what influences the adverts you see and take control of your adverts experience" and also offers a link to learn about Facebook adverts. The first option then, further down the page is a red heart with 'Your interests,' as seen in figure 7.

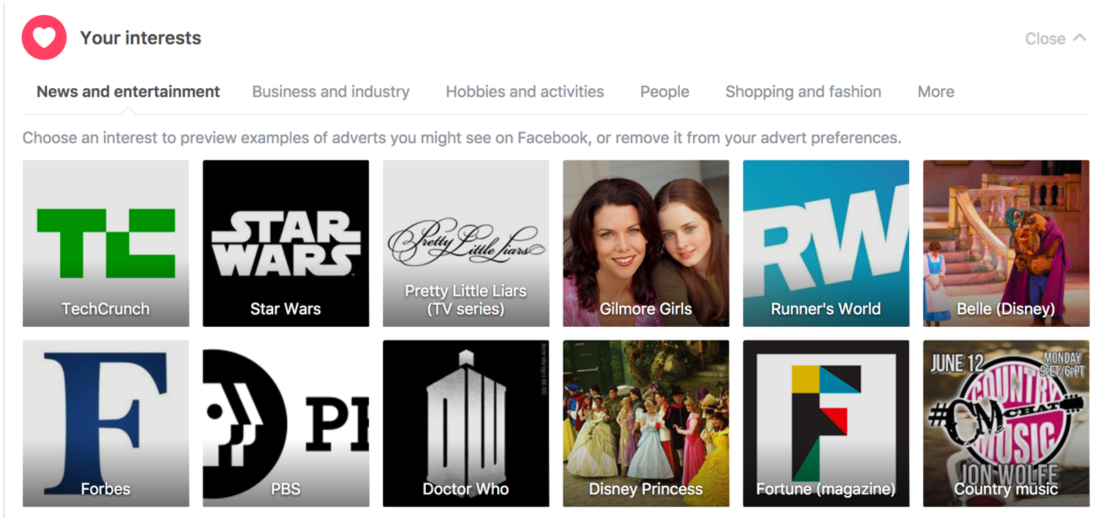

Figure 7: The author's interests profile as presented by Facebook.

Recruitment

To bring participants into the lab, there were two main ways of contacting them. The primary way was through several Craigslist postings to the Maryland, Washington D.C, Baltimore and Northern Virginia areas. A supplemental method that was used was the listserv for the College Park area through NextDoor, which is a community mailing list. Each posting, flyer and email contained the compensation ($20) and length of time of interview (45-60 minutes).

Each potential participant was asked to complete a demographic survey where their age, gender, race and education was asked hosted on Qualtrics. They were also asked if they had active Facebook and Google accounts and if they would be willing to log into the accounts during the interview. At that point, they were also informed that they would have control over the computer and it would be facing them the entire time so that the participant could only share what they wished to share. Finally, they also provided their email so they could be contacted to schedule the interview.

Procedure

Since the interview setup was semi-structured, the actual length of the interviews ranged from 40 to 57 minutes. Each of these interviews used the same base protocol, but the question wording, any relevant follow ups, and arrangement of the questions was adapted for each situation. By using this semi-structured methodology, we could better gauge the participant's comfort level and reaction to different situations in a less rigid environment.

The structure of each interview was in four parts. First, a few brief introductory questions about their general social media use. For example, "How often do you go onto Facebook?" and "Are you logged into Facebook and Google all the time?." This section of the interview was to gain a baseline and familiarity with the participant.
The second and third section of the interview is when the participant logged into their Facebook and Google account and went to the ad profile. The order of the sites was alternated between participants to keep the overall data from being slanted in one direction or the other on the order that they saw the profiles. For each of these sites we had the participant read through the different inferenced interests. They were asked to share any or all of them with the interviewer as well as any initial reactions. After that portion was done, they were asked questions such as "which of these are actually interests of yours" and "which of these are not interests of yours" and asked them to explain how they believe this came to be part of their interest profile. This was repeated several times and a conversation related to each interest emerged as they explained how they believe that interest was inferred about them. Additionally, we asked about their general level of comfort of the profile as well as their level of comfort when it came to

certain inferred interests. To close each part we also asked them how this profile might influence their experience with each site and what, if any, benefits and drawbacks they saw from the two sites having this profile about them.

The closing part of the interview asked questions such as "Are there any situations where other people might see a targeted ad towards you where this would make you feel uncomfortable?" and asked them to compare the two profiles so gain insight into what elements they liked and which they did not like.

Analysis

To analyze these interviews, we used an iterative open coding process. As the interviews were being performed, the researcher transcribed the interviews and created a starter code book. After five of the interviews were done and transcribed, the researchers met to review the transcripts and the code book. The two researchers then coded the first two interviews independently then compared the assigned codes. The code book was updated with additional codes and details, and the first two interviews were recoded to match the new code book. This was repeated 2 more times until there was a high level of consistency between the two coders. The rest of the interviews were then transcribed and, using Dedoose.com, were coded. The rest of the interviews were then coded independently and after all of the coding was done, Krippendorf's alpha[33] was calculated, resulting in a value of 0.657. After this was calculated, the two coders met to review and discuss the codes that were not agreed upon until there was a 100% agreement.

Limitations

As with any interview study, there are limitations to the generalization and application of the results. One of the primary limitations is because we asked individuals to recall their internet habits and how they might have influenced the interests listed. The technical experience of the participant might have also influenced their responses. To help compensate for this we recruited a diverse group of individuals across age, race, education and gender. Finally, we were talking about some potentially sensitive topics that were personal to the participant. This might have influenced the responses that we received if the they downplayed or did not wish to share. We tried to minimize this by allowing the participant to have control of the computer and share only what they wished to share.

Results

In this section, we will discuss the participants recruited and the themes that arose from the interviews performed.

Participants

Through the different recruiting methods (Craigslist and NextDoor listserv) 132 individuals filled out the demographic questionnaire. Of those, we selected 32 individuals to interview on a rolling basis. We wished to gain a demographically diverse set of individuals. To do so we selected individuals that had certain demographic attributes (age, race, gender, education) that did not have previously represented. Of the 32 invited, 25 filled out the scheduling doodle poll, and 15 of those attended their assigned interview appointments.

As you can see in Table 16, the participants are from a diverse background. We interviewed eight females and seven males. The age range was from 19 to 61 years old. There were six white participants, five black participants, two Asian, one Hispanic and one participant who identified as 'Other.' Through this diverse set of individuals, we believe that we have achieved a snapshot of a demographically diverse group of individuals.

Participant ID	Gender	Age	Race	Education
P1	F	61 yrs	Black	High School
P2	F	23 yrs	White	Some College
P3	M	19 yrs	White	Some College
P4	M	52 yrs	Black	Trade/Technical School
P5	F	29 yrs	Asian	B.S.
P6	M	43 yrs	Other	Associates Degree
P7	F	25 yrs	White	B.S.
P8	F	35 yrs	Black	Associates Degree
P9	M	31 yrs	White	Masters
P10	F	28 yrs	Black	B.S.
P11	F	46 yrs	White	Trade/Technical School
P12	F	28 yrs	White	B.S.
P13	M	22 yrs	Black	B.S.
P14	M	32 yrs	Asian	Some College
P15	M	22 yrs	Hispanic/Latino	B.S.

Table 16: Interview participants' demographics. The columns represent the Participant ID number (coded by the interview date order), the gender of the participant, age race and education level.

We also noticed that the participants had a variety of the amount of social media use, from going on once a week or so to checking every hour. We also noticed that most participants report that they are logged in most of the time on their devices. While interviewing the participants twelve of the fifteen participants made more than one attempt to login to at least one of the social media accounts. Six of those twelve had to change the password while in the interview.

We also found that 13 of the 15 (87%) participants had active accounts on other forms of social media ranging from Google+ to Snapchat. Two participants noted that they have accounts on other forms of social media (namely Google+) but are not actively posting or monitoring them.

Reactions to inaccuracy

The level of accuracy, when self-rated ranged from 20% to 90% for both Facebook and Google, with the mean being about 70%. The participants identified many reasons why their profile was accurate or not to them.

Interests vs Searching Habits

When talking with the participants there seemed to be a divide between their 'interests' and what they actively searched for. For example, one participant (P8) had "Lighting and Home Repair" as an interest listed on her Google profile. She commented that this was most likely because she had been searching for how to install a light above her kitchen table about six months before. She noted that these searchers were only done out of necessity and she does not plan on doing any more in the future. She did not identify this as an interest of hers but could understand how it could have been identified as one.

Additionally, four of the participants explicitly separated personal interest from searching interest without being prompted. Four more (making a total of 53%), when prompted, also separated the 'real world' interest from the online interests. One example of this was seen in P13. One of the interests listed was "Deals and Couponing" and he noted that he does use google to search for good coupons and savings but he felt as though "this isn't really an interest" of his.

Multiple User

Another interesting situation happened with P2 when she mentioned basketball on her profile. She was initially surprised to see that listed but then said “Oh, that must have happened when my husband used my computer. He must have searched for a player or a score or something like that.” She had no problem with that being part of her profile. The fact that multiple people with distinct interests used the same account was only mentioned once during the fifteen interviews, but it has a larger potential to lead to inaccurate interests.

Stereotyping

One participant (P15) believed that the only reason that Latino music was part of his Facebook profile was because he identified as Hispanic on Facebook. He said: “I have never listened to that genre of music and I don’t think I was ever tagged in a posting relating to that.” This stereotyping, though a singleton case, could also lead to inaccurate inferences.

Surprising accuracy

Eleven of the fifteen uttered or expressed some level of surprise at some point through this process. Eight of the eleven agreed with the interest they were surprised about while three of them disagreed with the interest inferred about them.

Expressed reasons for discomfort

Along with accuracy we looked at the level of comfort that the participants had about the elements predicted about them.

Inaccuracy leads to more discomfort

One major theme that emerged from this was that most participants, nine of fourteen, (one expressed either comfort or discomfort) felt more discomfort when it came to interests that were not accurate. They felt as though, as P6 put it, “that doesn’t accurately represent me, so that sorta makes me feel uncomfortable.”

Level of detail

Of the participants that expressed an accuracy level higher than 70% (seven of the fifteen), five expressed little to no discomfort in the profile’s accuracy while the other two felt very uncomfortable with the level of details in the profile, particularly the Facebook one. Both expressed concern about other companies either using or influencing the results of the profile. For example, P15 noted that he only listed to classical music on Spotify and did not mention, like or get tagged in any posts that were related to classical, from what he can recall, and yet Classical Music was the first interest listed in his Facebook profile.

Embarrassing ads being shown

The closing portion of the interview opened with the question: “Are there any situations where other people might see these targeted ads towards you where this would make you feel uncomfortable?” P11’s response was: “Nobody was around, but something happened, I don't know maybe a month or so ago where I was on there, and I saw something about needing to lose, it was a ginormous amount of weight. It was an obnoxious ad about being fat and losing weight. And I actually called a friend. I said, ‘I can't believe this is happening. The nerve of them’” in response to a Facebook ad that she saw. Four other participants did state that there might be a situation where there might be an ad that would make them feel uncomfortable, but none of the others gave an example. The remaining ten did not indicate that there was situation for them that would make them feel that way.

Potential Benefits

For each profile the participant was prompted with: “Do you see any drawbacks or benefits from (Facebook/Google) having this profile about you?” Eleven of the participants noted advertisements as the first benefit mentioned. Several (4) mentioned that this profile helped weed out the ads that weren’t relevant to them and only show products or services that they weren’t interested in. None of the participants mentioned a drawback first and ten of them only mentioned one when prompted again. Six of the participants noted that if the profile was wrong then they would be shown content that wasn’t relevant to them which three of the six thought would be annoying. Another drawback that was mentioned by five participants was the limiting of content shown to them, both in search results and on their new feed. Four of the five also noted that this could be a benefit if it something that they are looking for, but could make it “harder to find what I actually want if they make assumptions about what they think I want” (P14).

Limitations, Discussion, and Implications

In this section, we will discuss the limitations and implications from the results found. Additionally, we will discuss the main themes and possible reasons that those were prominent with the participants.

Limitations

Many of the themes mentioned above started to emerge after the first five or so interviews and became clearer with more participants.

Something to note here is the fact that a participant didn't mention or comment on a feeling, concern, or any other topic does not mean that he/she does not feel that way. This could happen because they might not have thought it important, might not have been thought about in that way or the did not wish to share that with the interviewer.

Discussion

One surprising theme that emerged was the idea that participants felt more uncomfortable with something that is incorrectly assumed or inferred about them.

Another interesting result was the level of reported accuracy of the profiles. The large range of next to nothing being correct to all but an item or two being correct was also surprising. There are several possible explanations for why this happened. First, the reported 'interests' of the participant might line up with their internet use but not their real-world life. Second, they did not wish to report that the interests were accurate to the interviewer. Third, their interests might have changed over time and the profile shows some older interests that are no longer relevant. Fourth, there could be multiple users on the account that were not mentioned in the interview that might influence the results of the algorithm. Lastly, lack of use or inconsistent use might lead to incorrect assumptions about the user. Of these options, it would seem like the first one listed is the most likely. Based on the discussion in the interview, most people had a hard time separating their internet life from their physical world life. For some, there might not be any difference but for others, there might be a large divide between the interests expressed in each.

Implications

Returning to how interests are effected by searches, there seems to be a disconnect that should be remedied. One of ways to do this could be to reframe the way that the different companies frame the 'interests' that are inferred. Instead of framing them in such a way that the user would assume that they are indeed things that would be of interest to them they could frame it in a way that either explains how these topics could be useful to the user.

Another aspect that became apparent is the fact that not everything in the profiles are correct. How the algorithms collect, analyze and use the information will and should continually evolve. This evolution should pay particular attention to the topics/interests that the profile got wrong and how that happened. By doing this it would seem a level of discomfort would decrease among the users.

In closing, this is an ever-evolving field as algorithms become better at predicting interests and people either become more familiar and more open with the internet or more familiar and more cautious about what they share on the internet.

Chapter 5: Conclusion and Future Work

This final chapter will summarize the results of the study, lessons learned from each of the studies, and future work that can be done to expand upon both studies in the future.

Conclusion

Each of the studies offered both surprising and expected answers to the proposed hypothesis.

In the first study, people viewed discrimination when it happened due to behavior rather than demographics less problematic. Surprising, though, there wasn't any significant difference between how the different entities were viewed when it came to responsibility.

For the second study, some general themes arose from each of the parts of the interview. Overall it seems like people are either very concerned about the online tracking or not at all concerned (more people in the latter group).

Future Work

As stated in chapter 3, our results highlight an important distinction between users' perceptions of scenarios involving racial vs. online behavior based discrimination. Our research explored only web history based targeting, and thus, future work may seek to explore whether users are more accepting of advertisement targeting, or OBA discrimination,, based on all types of online behaviors or whether there are acceptable vs. unacceptable behaviors, on which to target ads. A similar exploration may also be prudent for exploring user reactions to the use of different demographics. While our pilot results indicated that users do not feel as strongly about discrimination based on other factors such as pre-existing health conditions, there is room for more fine-grained exploration. Additionally, we only explored user perceptions of scenarios involving advertising discrimination. Related work [34][16] has also shown evidence of discrimination in the search results that are shown to different users.

Thus, future work may wish to explore and compare user reactions to discriminatory advertising vs. search results. Finally, prior work[33][34] has shown mixed results regarding whether users will act on their privacy preferences. To better understand the depth of users concerns about OBA, it may be prudent to conduct behavioral-

economics based experiments, to determine whether they would change their behavior or buying patterns based on discriminatory OBA practices

In the second study, we explored people's reactions to different ad profiles created about them on Facebook and Google. There are several ways to expand upon this including a diary study, examination of different companies' ad profiles and also some of the studies mentioned above would combine the research questions.

The diary study could help identify a longer exposure to targeted advertisements and people's reactions over time. Through this study, we could gain insight into if, or how fast, the ad profiles change and if there is a direct correlation between the advertisements shown to the user and the interests that are part of the profile or added to the profile.

Another way to expand upon this study would be to have the user look at other companies' ad profiles that are available to them, such as twitter. There are also third party plugins that will allow the user to view the information gathered. By exposing them to a larger variety of companies' profiles we will be able to expand and confirm the themes that were observed in this study.

Lastly, the behavioral economics study mentioned above would also fit here because it would be able to test how the user's level of discomfort would translate to a monetary fee.

Appendices

Appendix A: Study 1: Survey and Question Text

Q 1-4: 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

Q 6-9: 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 Sys- temy (the advertiser), Bezo Media (the ad network), the individual visiting the website, or the the local news web- site.

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

Q 11-14 Age, Gender, Education Level and Ethnicity demographics collected

Appendix B: Study 2: Interview Protocol

**Questions very similar, if not identical, to those listed below will be asked of participants. Some follow-up questions, not listed, may be asked dependent on the conversation with the participants. Not all participants may be asked all questions.*

Introduction: (*about 5 min)*

Hello. My name is [INSERT NAME]. Today we will be conducting a study that looks at your reactions to online advertisements and the different companies have inferred about you.

First, let's quickly go over how this study is going to work. This study will be broken up into two parts. The first, looking at how you interact with social media and advertisements that you see. The second part we will look at what Google and Facebook have inferred about you based on your online behavior. During the second part of the study you will have complete control over the computer and will only share what you feel comfortable with. I expect that this study will take about 45-50 minutes.

If at any point you become uncomfortable during the study, please let me know.

Do you have any questions at this point?

Give subject the consent form
I have this consent form here. Please take a moment to read over it and please let me know if you have any questions. I'll give you two copies – one is for you to keep, and the other is for you to sign and return. [POINT OUT THE PLACES THE SUBJECT NEEDS TO SIGN, POINT OUT SECTION WHERE THEY CAN CHOOSE WHETHER THEY ARE OK BEING VIDEO RECORDED]

Part 1: (*about 5 min)*

Like I mentioned earlier, during the first part of the study we are just looking at your general use of the internet and social media.

First, what type of social media do you use? (*give Facebook, snapchat as examples if needed)*

What do you use google for?

Are you signed into Facebook and Google all the time? Or do you sign in each time you use it?

How often do you say you are on these sites (*every hour, several times a day ect.)*

While using social media or the internet in general, do you notice any ads?
If yes: Do you find that the ads are relevant to you? In general, how do you feel about these ads?
If no: Do you use an ad blocker?

Have participant sign into Facebook and Google and proceed to ad preferences.

Part 2: (*about 35 min)*

Now you are going to look at the profiles that Facebook and Google have put together about you based on your online activity.

The computer will be facing you the entire time and you only have to share what you feel comfortable with. Any questions so far?

Ok let's start with (*interviews will alternate starting with Google and Facebook and then do the other one*).

Facebook: (*about 15 min)*

Alright, please turn to the Facebook tab that you logged into earlier.

Along the top of the page you will see "Your Advert Preferences" and right below that a heart with "Your Interests."

In that section there are different tabs that range from news and entertainment to Hobbies and activities to Travel, Places, and Events. Please take a moment to look through [*choose one at random]* and please think out loud about what you see and how you feel about it.

Pause as they read through the different areas.

Were you aware of this feature offered by Facebook?

If Yes: have you viewed these before?

If No: Were you surprised by this?

Can you share with me the topics that you see?

How accurately does this topic show your interests? (*follow up with Facebook activity vs everyday activities*)
If inaccurate: What is inaccurate.

Why did you choose that answer?

How do you feel about the fact that Facebook lists this as an interest of yours?

How comfortable are you with Facebook having this as one of your areas of interest?

How do you believe that was chosen by Facebook as something that you like? (*Discuss for the topics that were brought up.)*

Now, in that same or different category, would you mind sharing something that you disagree with or were surprised to see? … Why where do you disagree/surprised to see that? (*Repeat as necessary if the participant has more topics that they are willing to share.)*

Do the ads that you were shown on Facebook seem to follow these topics? If so, can you think of an example? (Go to home page)

Do you see any benefit from this?

Are there any other comments that you have at this time?

Google: (*about 15 min)*
Alright, please turn to the Google tab that you logged into earlier.

First, at the top of the page, do you have 'Ads Personalization' turned on or off.

If off, look and talk about profile since there will be no topics listed.

A little way down the page there is a list of topics that google has inferred about you. Please take a moment to read through some of those and if you wish, please share any initial thoughts or reactions that you have to these.

Pause as they read through the different areas.

Were you aware of this feature offered by Google?
If Yes: have you viewed these before?
If No: Were you surprised by this?

Can you share with me the topics that you see?

How accurately does this topic show your interests? (*follow up with Google activity vs everyday activities*)
If inaccurate: What is inaccurate.

Why did you choose that answer?

How do you feel about the fact that Google lists this as an interest of yours?

How comfortable are you with Google having this as one of your areas of interest?

How do you believe that was chosen by Google as something that you like? (*Discuss for the topics that were brought up.)*

Now, in that same or different category, would you mind sharing something that you disagree with or were surprised to see? … Why where do you disagree/surprised to see that? (*Repeat as necessary if the participant has more topics that they are willing to share.)*

Do the ads that you were shown on Google seem to follow these topics? If so, can you think of an example? (Go to home page)

Do you see any benefit from this?

Are there any other comments that you have at this time?

Closing: (*about 10 min)*

Are there any situations where other people might see these targeted ads towards you where this would make you feel uncomfortable?

Overall, how comfortable do you feel with Facebook and Google inferring this information about you?

How do you think Facebook and Google use these preferences?

Now that you have seen these different areas that Google and Facebook predicted about you, do you have any other reactions or comments?

Between the two sites, which do you feel is easier to understand? And why?

Between the two sites, which do you feel is, in general, more accurate to your preferences?

Between the two sites, which do you feel more comfortable with?

What aspects of each site did you like? Which aspects did you dislike?

Would you be willing to complete a couple minute online survey in about a week that will follow up with you about the advertisements that you have seen recently?

Bibliography

[1] pwc and iab, "IAB internet adverting revenue report," *IAB*, vol. 1, no. 1, pp. 1–30.

[2] "US Digital Ad Spending to Surpass TV this Year: Digital will represent 37% of US total media ad spending." .

[3] N. Hollis, "Ten years of learning on how online advertising builds brands," *Journal of Advertising Research*, vol. 45, no. 2. pp. 255–268, 2005.

[4] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang, "Smart, useful, scary, creepy," in *the Eighth Symposium*, 2012, p. 1.

[5] B. Liu, A. Sheth, and U. Weinsberg, "AdReveal: improving transparency into online targeted advertising," *... Hot Top. ...*, pp. 1–7, 2013.

[6] M. Lécuyer *et al.*, "{XR}ay: Increasing the Web's Transparency with Differential Correlation," in *Proceedings of the USENIX Security Symposium*, 2014.

[7] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan, "Adscape: Harvesting and Analyzing Online Display Ads," *arXiv.org*, Jul. 2014.

[8] M. Malheiros, C. Jennett, S. Patel, S. Brostoff, and M. A. Sasse, "Too Close for Comfort: A Study of the Effectiveness and Acceptability of Rich-media Personalized Advertising," in *Proceedings of the SIGCHI Conference on*

Human Factors in Computing Systems, 2012, pp. 579–588.

[9] J. Guynn, “Google {P}hotos labeled black people `gorillas’,” *{USA} Today*, 2015.

[10] J. Valentino-Davies, J. Singer-Vine, and A. Soltani, “Websites Vary Prices, Deals Based on Users’ Information,” *Wall Str. J.*, 2012.

[11] L. Agarwal, N. Shrivastava, S. Jaiswal, and S. Panjwani, “Do not embarrass,” in *the Ninth Symposium*, 2013, p. 1.

[12] P. Leon, B. Ur, R. Shay, Y. Wang, R. Balebako, and L. Cranor, “Why Johnny can’t opt out,” in *the 2012 ACM annual conference*, 2012, pp. 589–598.

[13] R. Coen, E. Paul, P. Vanegas, A. Lange, and G. S. Hans, “A user-centered perspective on algorithmic personalization,” University of California, Berkeley, 2016.

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

[15] J. Warshaw, T. Matthews, S. Whittaker, C. Kau, M. Bengualid, and B. a Smith, “Can an Algorithm Know the ‘ Real You ’? Understanding People ’ s Reactions to Hyper-personal Analytics Systems,” *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, pp. 797–806, 2015.

[16] J. M. Carrascosa, J. Mikians, R. Cuevas, V. Erramilli, and N. Laoutaris, “I

Always Feel Like Somebody's Watching Me. Measuring Online Behavioural Advertising," *arXiv.org*, Nov. 2014.

[17] S. Guha, B. Cheng, and P. Francis, "Challenges in measuring online advertising systems," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, 2010, pp. 81–87.

[18] J. P. Pierce, "Smoking Initiation by Adolescent Girls, 1944 Through 1988," *JAMA*, vol. 271, no. 8, p. 608, 1994.

[19] L. Sweeney, "Discrimination in Online Ad Delivery," *arXiv.org*, Jan. 2013.

[20] A. Datta, M. C. Tschantz, and A. Datta, "Discrimination and Opacity in Online Behavioral Advertising." .

[21] C. Huff and D. Tingley, "'Who are these people?' Evaluating the demographic characteristics and political preferences of MTurk survey respondents," *Res. Polit.*, 2015.

[22] P. G. Ipeirotis, "Demographics of Mechanical Turk," *SSRN*, 2010.

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

[24] H. Akaike, "A new look at the statistical model identification," in *IEEE Transactions on Automatic Control*, 1974.

[25] "American Community Survey 5-Year Estimates." 2015.

[26] "Likert-Type Scale Response Anchors." .

[27] F. J. Crosby and D. I. Cordova, “Words Worth of Wisdom: Toward an Understanding of Affirmative Action,” *J. Soc. Issues*, vol. 52, no. 4, pp. 33–49, 1996.

[28] “Self-Regulatory Principles, For Online Behavioral Advertising.” 2009.

[29] “Online Behavioral Tracking and Targeting Concerns and Solutions Legislative Primer.” 2009.

[30] “Principles and Practices for Advertising Ethics.” .

[31] “Google?s algorithm shows prestigious job ads to men, but not to women. Here?s why that should worry you. .” .

[32] G. Guest, A. Bunce, and L. Johnson, “How many interviews are enough?,” *Field methods*, vol. 18, no. 1, p. 59, 2006.

[33] K. L. Gwet, “On The Krippendorff’s Alpha Coefficient,” *Manuscr. Submitt. Publ.*, vol. 2, no. 1971, pp. 1–16, 2011.

[34] “CAN SEARCH DISCRIMINATION BY A MONOPOLIST VIOLATE U.S. ANTITRUST LAWS?” .

[35] J. Grossklags and A. Acquisti, “When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information.,” in *WEIS*, 2007.

[36] J. Y. Tsai, S. Egelman, L. Cranor, and A. Acquisti, “The effect of online

privacy information on purchasing behavior: An experimental study," *Inf. Syst. Res.*, vol. 22, no. 2, pp. 254–268, 2011.