"I Have a Narrow Thought Process": Constraints on Explanations Connecting Inferences and Self-Perceptions

Emilee Rader
Michigan State University
emilee@msu.edu

Samantha Hautea
Michigan State University
hauteasa@msu.edu

Anjali Munasinghe
Michigan State University
amunasinghe1620@gmail.com

Abstract

Most people are unfamiliar with the kinds of inferences that platforms like Facebook and Google can automatically associate with them, despite the existence of interfaces designed to provide transparency to end users. We conducted a study to investigate people's reactions upon being exposed to these inferences, to learn if and how they perceived the inferences to be connected to themselves. Through qualitative analysis, we found that the evidence participants used to relate the inferences with their self-perceptions was bounded by what they remembered about their own past behaviors in connection with the platform. Inferences that participants felt were implausible given their own behavior were rationalized as being related to family members, outdated, or could fit anyone with similar demographic characteristics. Participants also identified some inferences they believed had no connection with themselves whatsoever. We discuss implications for how participants' reasoning might lead to expectations about what kinds of inferences are possible, and what this means for people's ability to make informed privacy decisions regarding consent and disclosure.

1 Introduction

The current model for governing digital data collection and use, *notice and choice*, entails providing access to complex terms of use or privacy policies. These documents describe how platforms—systems consisting of networked hardware and software that connect people with information and with each other—will collect, use, and share the data they collect about people. The notice and choice model assumes that if proper transparency is provided, then people will choose not to use platforms that have data collection and use practices they don't like [28]. This model reinforces the idea that digital privacy—control over how data about oneself is collected and used—can be effectively managed by individuals who make informed choices that are aligned with their preferences.

But, in reality, people struggle to manage their privacy. Widespread data collection and use of machine learning technologies combined with the rise of data brokers makes it possible for platforms to generate inferences about users, which consist of new information that is automatically derived from data the platforms collect or obtain [2]. For example, one person's browsing history aggregated with data collected from millions of other people can be used to derive inferences about anyone who uses the platform, revealing sensitive personal characteristics people might prefer not to disclose [4, 7, 14, 17]. Most people don't have a very good understanding of what kinds of inferences can be derived in this way [16, 23, 24, 32] or how those inferences can be used to direct their attention and influence their choices and opportunities.

New data privacy laws like the E.U. General Data Protection Regulation (GDPR)[1] and the U.S. California Consumer Privacy Act (CCPA)[2] have mandated better transparency about data collection and use. Partially in response to this, platforms have begun showing more information to users about what they are doing with the data they collect about them. Both Facebook and Google provide a web page where people with an account on the platform may view their "Ad Settings factors" (Google)[3] or "Ad Preferences categories" (Facebook)[4]. Despite language that connotes some level of user ownership and control—the inferences are described as "your ads" by Google, and "your information" by Facebook—these pages display information that has been automatically generated by the platform, not specified by the user. Interfaces

---

[1] https://gdpr-info.eu
[2] https://oag.ca.gov/privacy/ccpa
[3] https://adssettings.google.com/
[4] https://www.facebook.com/ads/preferences/

like these provide some visibility into the kinds of information it is possible for platforms to infer about users, and as such are potentially an important way for people to develop an understanding of their inference-generating capabilities.

We conducted a survey and interview study to investigate participants' reactions to the inferences Facebook and Google have made about them, and what their explanations about how and why the inferences were assigned reveal about their understanding of the platforms' inferencing capabilities and motivations. The way participants explain the connection (or lack of connection) they perceive between themselves and the inferences can reveal their guesses, knowledge, and insights about where the inferences come from.

We found that participants' understanding of the inferences was based primarily on their own memories about their past online behavior, their self-perceptions of their interests, and their perceptions of the interests of family members. Inferences that could not be explained as related to these things were assumed to be mistakes. These findings show that the framing people use for understanding inferences is limited to things they are directly knowledgeable about. This framing does not support envisioning inferences from the platforms' perspective. Therefore, inferences derived from data people cannot anticipate or for purposes they are unfamiliar with would be extremely difficult for them to truly consent to. This paper contributes novel findings to the privacy literature about the constraints on people's knowledge and understanding of what platforms can infer about them, which have implications for their ability to make informed informed privacy decisions regarding consent and disclosure.

2 Related Work

2.1 User Awareness of Online Tracking and Inferences
Tracking of user browsing behavior is a phenomenon that has become increasingly complex. At least one study has empirically shown that the scope of information tracked and the variety of trackers on websites have greatly increased over the past two decades, though studies on such tracking behavior came much later [21]. Media controversies surrounding large-scale tracking and the sharing of data have made some of the tensions between users and companies more public, but may not necessarily equip users to conceptualize more broadly about these systems [12].

One type of user tracking that has been studied extensively is online behavioral advertising (OBA), which presents users with advertisements targeted to them based on their tracked behavior (see [3] and [30] for a more comprehensive overview). User studies of reactions to targeted advertising show that users perceive useful and beneficial properties of OBA, even as there are aspects of it that they find uncomfortable. The delivery of more personalized, relevant content is considered useful, while discomfort can stem from perceptions of excessive intrusiveness, sometimes described as 'scary' or 'creepy' [5, 29]. Users who 'feel watched' in this way may be intuitively expecting systems to conform with social norms of not conducting unsolicited observation of others [22]. What these user perceptions have in common is an understanding of algorithmic inferences as driven by social entities and human logic, which may not be true for complex systems that make inferences across patterns beyond the intuitive capacity of the human mind to connect. Indeed, though websites may provide various disclosures about information collection (whether legally required or industry self-regulated), these do not appear to be particularly effective in providing users with meaningful notice that it is taking place [18, 34].

2.2 Awareness and Inference Literacy
The increasing complexity of online tracking, and the implications this tracking holds for privacy decision-making, has led some scholars to call for the need to recognize 'inference literacy,' or the "beliefs and misconceptions users hold about companies inferencing methods and capabilities" [31]. Users perceive that their online activities are being tracked or followed, though industry-standard icons, taglines, and explicit disclosures are often misinterpreted [19, 29]. Uncertainty about the usage of behavioral data is a subject of concern for users [25], and individuals articulate clear preferences for the type of information and categories they are comfortable with having associated with them based on what they believe others can do with that data [11, 20].

Feedback from information collection systems can greatly influence what users are able to conceptualize them doing [8, 10, 24]. Some companies have provided their users with access to their online behavioral profile. Experiments have been used to gain a more technical understanding of how some of the infrastructures underlying these data collection and inference activities respond to user behavior [6], but studies have also explored user reactions to these systems. A study that exposed users to their site-generated behavioral profile reported that participants found these difficult to comprehend, and focused solely on identifying user concerns with data collection [25]. Several tools have been created to increase user awareness about tracking and inferences, some of which employ a soft paternalism or 'nudging' approach to guide users to different decision-making processes by using underlying behavioral biases [1]. Other approaches have focused on creating novel ways of visualizing behavioral data, exemplified in one study where browsing information was collected from a plugin in order to show users what can be inferred from certain types of data [32]. A key theme that emerges from these studies is that user awareness is greater when they are able to recognize themselves in the data collected about them.

2.3 Research Goals
This study focuses on users' reactions to and explanations of real-world inferences assigned to them, in order to investigate how they make sense of the inferences and relate them

to their lives and self-perceptions. This differs from previous research on user understanding of underlying tracking mechanisms [29, 35], and platforms' reasons for tracking and tailoring ads and how that relates to privacy concern regarding tracking [9]. It also differs from Weinshel et al. [32] which generated its own inferences rather than investigating reactions to platforms' actual inferences. While Eslami et al. [11] found that that users "justify" algorithmic decisions by looking for connections between themselves and inferences, we take this idea further by identifying patterns in the explanations and evidence present the reactions our study elicited from participants to the inferences that had been assigned to them. These reactions, and the connections users make between the inferences and their self-perceptions, can help researchers and designers understand what makes inferences seem plausible and therefore what kinds of inferences people might expect when consenting to a platform.

3 Method

3.1 Approach

Our study focused on eliciting participants' reactions to the actual inferences made about them by Facebook and Google. We focused on these platforms because earlier research about awareness of inferences showed that participants' perceptions of the likelihood of different types of inferences varied by platform. For example, more participants who were shown a hypothetical scenario involving clicking on links in the Facebook News Feed believed inferences about their social relationships were likely than participants who viewed a scenario involving clicking on links in Google search results [23].

Many studies have used hypothetical scenarios to investigate inference-related awareness, concern, and privacy preferences; however, people can react differently to real evidence of platforms' data practices contextualized in their own lives and experiences, than to general descriptions or hypotheticals [22]. We began by conducting a survey in which we asked respondents to answer questions about the specific inferences made about them by either Facebook or Google. We then invited a subset of the survey respondents to participate in a follow-up interview where we showed them a report that presented the inferences they had answered questions about and a summary of their responses to the questions. We used the reports as a form of breaching experiment [13] that provided visibility into the platforms' inferencing capabilities, which most of our participants were previously not aware of. The interview focused on participants' overall reactions to the inferences, and their explanations about why they thought those inferences had been assigned to them and how they had been generated.

3.2 Survey

The survey took place in late March and early April 2019. Respondents were recruited using a subject pool consisting

	Facebook	*Google*
Man	12	16
Woman	31	35
Other	1	0
Total	44	51
No college	1	4
Some college	7	3
Four-year college degree	15	19
Some postgraduate	1	4
Postgraduate degree	20	21
Income less than $34,999	15	13
Income $35,000 to $74,999	13	19
Income $75,000 to $149,999	11	17
Income $150,000 or more	5	2
Aware of inferences page: Yes	11	5
Aware of inferences page: Unsure	4	5
Aware of inferences page: No.	29	41
Num. inferences: Mean (SD)	12 (9.86)	67.6 (35.1)
Num. of inferences: Median	4	66
Num. of inferences: Max	33	131

Table 1: Characteristics of the survey respondents.

of members of the community surrounding Michigan State University. The first page of the survey informed potential respondents before they consented that they would be asked to log in to either Facebook or Google and navigate to a web page which they would be asked to download and then submit to the survey so that the survey could generate questions for them based on the content on that page. Eligible participants were at least 18 years old (*Min*=22, *Mean*=38, *Max*=71), and regular users of either Google or Facebook. Students at the university were not eligible to participate. People who consented to the survey and passed the screening questions were randomly assigned to answer questions about either their Facebook or Google inferences.

The survey asked questions about respondents' use of Facebook or Google and their demographics, and then it provided instructions for navigating to their inferences web page, saving the web page as a file, and uploading it to our server via the survey interface. The files submitted by respondents were parsed to identify the inference categories and then deleted. The survey then generated three questions about each inference, which respondents answered on a 7-point Likert scale (Strongly disagree–Strongly agree). "[inference]" in each question was replaced by the text parsed out of the file each respondent submitted via the survey:

- *sensible*: It makes sense that [inference] is associated with me.
- *relevant*: [inference] is relevant to who I am as a person.
- *accurate*: [inference] is an accurate description of my everyday activities.

	ID	Gender	Age	Aware?	Number of Inferences	Mean Accuracy
Facebook Participants	P01	Woman	62	No	20	5.05
	P02	Woman	28	No	26	5.44
	P10	Man	44	Unsure	25	4.48
	P11	Man	63	No	24	5.62
	P13	Woman	28	No	33	4.39
	P15	Woman	44	Yes	14	4.00
	P16	Woman	34	No	25	4.84
	P20	Man	38	Unsure	26	4.42
Google Participants	P03	Man	38	No	59	4.05
	P04	Woman	32	No	98	4.76
	P05	Man	29	Yes	63	2.32
	P07	Man	71	No	62	4.44
	P08	Woman	59	Unsure	83	4.49
	P09	Woman	63	No	40	4.78
	P12	Man	31	No	48	3.77
	P14	Man	30	No	94	4.68
	P17	Woman	48	No	65	3.54
	P18	Woman	30	No	104	3.87
	P19	Man	38	No	44	3.05
	P21	Woman	40	No	46	4.00
	P22	Woman	43	No	109	3.07

Table 2: Characteristics of the interview participants. 'Aware' indicates whether the participant had previously seen the platform's inferences web page. 'Mean Accuracy' is the average of each participant's responses to the *accuracy* survey question about each inference. Higher means greater perceived accuracy on a scale of 1–7.

For respondents with a very large number of inferences assigned to them, our code randomly selected 85 to ask about, because our pilot testing indicated that any more than this would result in an excessively long survey. The last question in the survey asked if respondents would be interested in participating in a follow-up interview, and 57% said yes and provided their contact information. On average, it took 32.5 minutes (*SD*=91) to complete the survey. Ninety-five respondents completed the survey (44 Facebook, 51 Google; 28 men, 66 women, 1 did not disclose gender). Eighty-seven percent of respondents were white, 7% were Asian, and the remaining respondents reported multiple ethnicities. Further demographics are presented in Table 1. Respondents who completed the survey received a $5 USD Amazon gift card. The survey questions are included in Appendix A.

3.3 Interviews

The interviews took place in April 2019, shortly after the survey data collection ended. From the 57 survey respondents who said they were interested in being interviewed, we randomly chose participants to invite, prioritizing gender and platform balance. After the first five interviews had been conducted, we realized that the nature of the Facebook inferences made them more difficult for participants to react to (see more
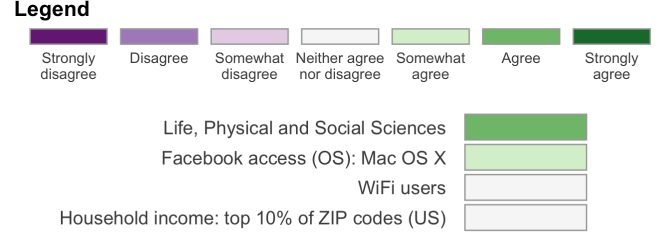

Figure 1: Example excerpt from a Facebook inferences report generated from data about one of the authors.

about this in the Findings section). At that point we decided to conduct more of the remaining interviews with Google participants than Facebook participants. We conducted 22 interviews, but excluded P06 before analysis due to a high proportion of low quality, off-topic answers to questions. Of the 21 remaining interviews, 9 were with men and 12 with women, and 8 focused on Facebook inferences while 13 focused on Google inferences. Interview participants ranged in age from 28 to 71 (*Mean*=43). Most participants were currently or formerly employed as knowledge workers (e.g., staff or instructors at the university; teachers; working in local government; retired). See Table 2 for further information about the interview participants.

Each interview began by showing the participant a report that we generated which included the inferences they had answered questions about in the survey, and the average of the ratings of the three questions asked about each inference, rounded to the nearest whole number and then color-coded according to the position of the average on the 7-point disagree-agree Likert scale. The report was designed to be a simple artifact that enabled us to elicit participants' reactions to the inferences, independent of the jargon and branding on the platforms' own pages. The last page of the report also showed inferences that were unique to each participant among the survey respondents who had been assigned the same platform, and inferences that all survey respondents asked about that platform had been assigned. See Figure 1 for an excerpt of a report generated using the inferences of one of the authors. The interview protocol and example Facebook and Google inference reports are included in Appendix C.

The semi-structured interview protocol included questions designed to investigate participants' beliefs about why and how the inferences had been assigned to them, and about what the inferences are used for by Facebook and Google. Care was taken to elicit participants' thoughts and reactions in their own words and to ask non-leading follow-up questions, so that the interviewers did not impart a sense that technical correctness or incorrectness was an important framework for thinking about the inferences or prime their answers in other ways. We also avoided using the word "privacy" in the interviews unless the participant mentioned it first, to avoid influencing their reactions to the inferences. Interviews lasted 30-40 minutes,

and participants received a $25 USD Amazon gift card as a thank you for participating.

3.4 Analysis

The interviews were digitally recorded, transcribed, and identifying information was removed from the transcripts. We conducted iterative qualitative analysis of the transcripts that progressed over several rounds [26]. In the first round, two coders conducted inductive open coding that identified descriptive themes in the data. These themes included beliefs about participants' interests, characteristics, past activities, goals, etc. that were related to any connection they perceived between themselves and the inference they were talking about, and varying levels of understanding about how the inferences might have been selected. After the first round, the coders summarized all of the codes by participant, and the research team engaged in a several-day immersive interpretation session with the goal of identifying higher level themes. Afterward, one member of the research team engaged in a second round of coding focusing on characterizing types of reactions to the inferences (e.g., plausible, implausible, no connection; see the Findings section for details), and then a second member of the team focused on the reactions in order to group them by the types of evidence and examples participants provided in their explanations about how and why those inferences had been assigned to them. Our findings primarily focus on the higher-level themes and evidence identified in the second round of coding.

3.5 Limitations and Ethical Considerations

We used a convenience sample that was more highly educated, white, and generally had a higher income level than the general population of the United States. A sample with a different demographic composition would likely have been assigned different kinds of inferences. In addition, people who saw sensitive or uncomfortable inferences when they viewed their inference web page before saving it may have chosen not to complete the survey, and so people with characteristics that may make them more likely to be assigned sensitive inferences may not have participated. Our study was also U.S.-focused, and people in other countries and from other cultures might have different inferences, and different reactions to the inferences, than we observed in our sample.

In the interviews, we asked participants to react to and think about inferences that most were not aware of before the study. This means that participants' reactions were what first came to mind as they processed and thought about the inferences, which most of them had not spent much time doing before. These initial reactions are based on their existing mental models—the knowledge, beliefs and understandings of cause and effect related to data collection and use—that were present before the interviews. This means that our method was able to identify the kinds of thinking and reactions that

might occur when people encounter information for the first time about the kinds of inferences a platform makes about users, at the time when they are choosing whether to consent to its terms or not. But, we cannot know from these interviews how participants' thinking about the inferences could evolve over time, or how they might have reacted if they had been assigned different inferences or if the platform had revealed more/less about the inferences it makes. Also, our data do not allow us to conclusively enumerate all of the different types of inferences, data, and mechanisms for deriving inferences that participants were aware of and knowledgeable about, since we just asked them about the specific inferences assigned to them by a single platform.

This research was approved by our institution's IRB. Consent was obtained separately for the survey and follow-up interviews. The inferences files were parsed and only the labels for the inferences were stored in our database, associated with a random participant identifier string. No other information was gathered from the files. Previous research about Facebook's inferences has found that only a very small proportion (1.66%) are potentially sensitive, although the same study provided evidence that 73% of EU Facebook users in their sample had been assigned at least one sensitive inference [4]. Through piloting, we determined that participants would be very unlikely to feel that any of the inferences in the files were something they would be uncomfortable discussing in an interview. Inferences about survey respondents who did not complete the survey were deleted from our database. Also, both platforms provide controls that allow the user to "turn off" (Google) or "remove" (Facebook) inferences from the visible list. However, our testing showed that these deactivated inferences were present in participants' files. We did not store deactivated inferences in our database, nor did we ask questions about them in the survey.

4 Findings

4.1 Inferences Differed Across Platforms

The Facebook and Google inferences differed in number and in character for our survey respondents. Facebook had assigned fewer inferences to each respondent in our survey sample (*Mean*=12, *Max*=33) than Google had (*Mean*=66, *Max*=131). In addition, there were more unique inferences assigned by Google across all of our survey respondents than by Facebook (Google: 561; Facebook: 110), indicating that Google's inferences might appear to be more highly personalized to end users based on the information that is provided in the interface. However, Facebook survey respondents reported that they felt the inferences were more accurate, on average, than Google respondents (Facebook: *Mean*=4.64, *SD*=1.88; Google: *Mean*=3.84, *SD*=2.05).

The most common inferences assigned to Facebook respondents in our sample had to do with how respondents accessed Facebook. For example, "WiFi users" and "Mobile network or

device users" were two of the most common Facebook inferences in our data, both assigned to 42 of 44 Facebook survey respondents. All Facebook interview participants found it hard to react to inferences like these, because they seemed to be about obvious facts that the participants easily recognized as being related to themselves and therefore did not dispute. (Where participants' reactions were about a specific inference, we have included the inference in italics immediately before the quote.)

> *[Facebook access (browser): Chrome; Birthday in August]* I use Chrome. That's when I was born. –P01, Woman, Facebook

> *[WiFi users; Gmail users]* I use wifi, I use Gmail. –P10, Man, Facebook

In contrast, the most common inference among Google survey respondents was "Parenting" (46 out of 51 Google respondents) followed by "Home & Garden" (43) and "Shopping" (41). Interview participants asked about their Google inferences typically had a lot more to say about them, because the Google inferences seemed more descriptive and personally relevant. For example:

> *[Vehicle Shopping; Autos & Vehicles]* Yeah, I've been doing that recently for sure. I'm kind of in the market for a different vehicle. I mean, yeah I guess autos and vehicles probably went along with vehicle shopping... actually I was just, before I came here I was Googling how to replace a headlight in one of our cars. So yeah, I use that a lot. –P12, Man, Google

Google interview participants reacted in a similar manner as the Facebook respondents did to the inferences that seemed to be facts about themselves, e.g., "I am male, I'm over 65. No problems with that." (P07, Man, Google) and "Mobile phones, I'm sure that's because of Sprint." (P09, Woman, Google). However, the majority of Google inferences felt more personal to interview participants, and potentially descriptive of their interests. There were many Google inferences for genres of music, travel destinations, hobbies, sports, types of news (e.g., "American Football", 33 participants; "Gourmet & Specialty Foods", 31; "TV Talk Shows", 30). See Appendix B for examples of the most and least common inferences among our survey respondents.

Sixty-six percent of Facebook survey respondents and 80% of Google respondents had not seen their inferences page before. Only 9% of Facebook respondents and 10% of Google respondents said they had seen the page; the remaining respondents from both platforms were unsure. In addition, only 2 of the 21 interview participants were among those who indicated they'd seen their inferences page before the survey.

4.2 Inference Reactions and Explanations

Our analysis of the interview transcripts identified characteristics of participants' reactions to the inferences assigned to them that revealed three broad themes. The themes differed in

Inference:	Information that a platform has automatically assigned to a person that is derived from data the platform has collected or obtained.
Reaction:	A person's sense of the extent to which an inference seems to apply to them, focused on the relationship between the inference and themselves.
Explanation:	Interpretations that involve causal relations, about why and how an inference was assigned.
Plausible:	The inference was believable, and participants provided specific evidence or explanations supporting why it made sense to them.
Implausible:	The inference was initially not believable, but participants subsequently provided an explanation justifying why it might be related to them in some way.
No Connection:	The inference did not make sense, and participants were certain that it did not apply to them.

Table 3: Summary of the conceptual framework used to understand participants' reactions to the inferences.

terms of the types of evidence participants described to support their thinking about how the inferences connected with their self-perceptions, and in terms of the degree of connection they believed was present. We focus below on describing these themes, and characterizing participants' explanations of how they thought the inferences were connected to themselves. These explanations are important, because they illustrate participants' understanding of how the platforms' inferences relate to them and their lives. In general, generating explanations helps people to develop and evolve their understanding of how and why things work the way they do [15, 33]. As such, characterizing participants' reactions and explanations is important for identifying how people might anticipate what kinds of inferences are possible in these and other platforms and systems. See Table 3 for a summary of definitions of important concepts used in this section.

4.2.1 'Plausible' Reactions Were Grounded

The most common reaction to the inferences from both Facebook and Google participants was certainty that the inference was related to themselves and their interests. We refer to this type of reaction as a *plausible* reaction, because the inference was believable to them. The example below illustrates what a plausible reaction looked like:

> *[Parenting]* I think just because it's a main part of my life. It makes sense... it's I guess obvious that it would be on there. I feel like a good majority of stuff that I'm doing online outside of work is stuff for the kids. Like Googling summer camps and trying to figure out activities and I think of all this stuff that I'm randomly looking up online as kid related... –P21, Woman, Google

Participants had this type of reaction when inferences seemed to them to be grounded in what they could remember about their past actions and perceptions of their present activities, interests, and demographics. They felt like these inferences made sense to them, and they could produce specific evidence or explanations for why this was, without hesitation. Fifty-five percent of the reactions to inferences that were discussed by participants in the interviews (176 of 321) were *plausible*, and every participant had at least one plausible reaction to an inference; the average number of plausible reactions per participant was 8[5].

In a large proportion of the plausible reactions, the participant gave evidence of their past activities on the platform to support their feeling that the inference made sense for them. This evidence looked mostly like general recollections about the kinds of topics they search for (Google) or about typical content of posts they had made or articles they had read (Facebook). Nineteen participants had plausible reactions similar to these:

> *[TV Comedies]* How can I explain... I do view, they're my favorite genre of television. And [this inference] describes, I guess, my everyday activities probably more so than other things. I guess probably, I do cast searches on occasion, things like that. –P05, Man, Google

> *[Parents (All)]* Because I have three kids, well, a lot of my posts are about my kids. –P13, Woman, Facebook

Other plausible reactions, described by 11 participants, seemed related to memories of specific recent past searches. For example, in reaction to the inferences "Flooring" and "Lamps & Lighting", one participant described how she and her family had just been shopping for lamps online and had their floors replaced as part of remodeling their home (P22, Woman, Google). Another participant talked about how she was in the process of selling her house and was doing some searching online about appraisals ("Property Inspections & Appraisals"; P08, Woman, Google).

A third very common form of plausible reaction, mentioned by 14 participants, talked about the participant's general interests or how they spent a lot of their time as an explanation for why the inference made sense to them, but without mentioning any specific online activities. For example, P17 related one of her inferences to her volunteer activities, but did not mention online searching or web browsing:

> *[Dogs]* Because ninety-nine point nine percent of anything I do online outside of work is probably dog related. I volunteer at an animal shelter and I volunteer for a rescue. And I foster for them. –P17, Woman, Google

---

[5]Note that some participants commented on multiple inferences at once, and these were all coded together as one reaction. All participants only discussed a subset of the inferences present in their reports. The counts, averages and percentages reported in this section are only intended to convey relative prevalence within our sample.

There was little difference in the proportion and type of plausible reactions between Facebook and Google participants, with one exception: Facebook participants were more likely to mention inferences as related to their Facebook Friends or engagement with content on Facebook, whereas Google participants were more likely to mention search history. The two examples below illustrate this difference:

> *[Close Friends of Women with a Birthday in 7-30 days]* That's based on my friends list in Facebook –P13, Woman, Facebook

> *[Cooking & Recipes]* Yeah, all of my recipes come from the Internet, so same thing. I'm googling food a lot. –P22, Woman, Google

Another pattern we noticed in the data was that participants were more clear in their plausible reactions about describing the types of online activity they believed the inferences were based on, and therefore the types of data they understood the platforms to have access to, than any specifics about technical mechanisms that gave the platforms access to these data. When specifically probed about how they thought the platforms were able to assign specific inferences to them, participants' explanations were vague and high level, conveying notions of access and visibility but not detailed speculation about transmission or collection. Below are a few examples of this.

- *P13, Facebook:* Based on the content that I engage with and the pages that I follow
- *P16, Facebook:* Facebook has some way of knowing
- *P03, Google:* It'll look at everything from your email history to your search history
- *P22, Google:* Google sees it in my searches

As these findings show, participants had a *plausible* reaction to an inference when they were certain that it was relevant to them, because they could explain its' relevance using evidence they felt was true about themselves. Explanations about what caused plausible-seeming inferences to be assigned were constrained by participants' ability to identify potential reasons for this, related to the person's self-perceptions and platform use.

4.2.2 'Implausible' Reactions Were Rationalized

The second theme that emerged from our coding of participants' reactions to the inferences was uncertainty about why the inferences had been assigned, followed quickly by an explanation rationalizing a possible source for the connection between the inference and themselves. We refer to this type of reaction as an *implausible* reaction, because these inferences initially were difficult for participants to believe or relate to. But, after further consideration, the participant identified a reason that the inference may have been assigned. For example, below is a participant's reaction to an inference assigned to her by Facebook:

> *[Lived in United States (Formerly Expats - United States)]* Well, I still currently live here.

So those were pretty inaccurate to me... I'm sure they're collecting information on where I'm at. I do travel a lot for work so I can see how they would get that. –P02, Woman, Facebook

This participant initially rejected the idea that the inference was related to her, but then backtracked, explaining that perhaps her recent location history had something to do with why this inference had been assigned. It was as though, for these inferences, no immediate reason it would have been assigned to the participant came to mind. But, then a realization occurred that something about their past history, behavior online, or relationship with someone might be interpreted as supporting the inference, even if it seemed like a stretch to them. The explanations participants generated as part of implausible reactions to inferences were more speculative, and less convincing to themselves, than the explanations they had given for the inferences for which they'd had plausible reactions. Twenty-eight percent of the reactions we coded (89 of 321) were like this, and 19 out of 21 participants had an implausible reaction to at least one of the inferences that had been assigned to them. On average, participants had 5 implausible reactions to inferences.

The implausible reactions for which clear exceptions did not come to mind for participants formed a spectrum of perceived separation between the participant's self-perception and the inference. The most related inferences were those participants felt used to be true about them, but were not true any longer, and six participants reacted in this way:

> *[Groupon]* Yeah, when I first got into that, I used it for golf stuff, but I haven't done that in quite some time. It's like these things never go away. –P07, Man, Google

The next level of separation, slightly more distant than an outdated inference, had to do with inferences that the participant felt were related to their close family members or friends, but not themselves. They believed these inferences had to do with the specific interests of those people, or with searching they had done for things they themselves were not actually interested in, but people they were close to were. Seven participants had a similar reaction to an inference:

> *[Coffee & Tea]* I don't drink coffee. My husband does though. So again maybe, you know? –P18, Woman, Google

> *[Fitness]* I don't do a lot with fitness online, so I'm not sure why they came up. Although I do go to soccer.com a lot and do stuff a lot related to soccer because of my children, maybe that's why that's there? –P22, Woman, Google

The third and most distant characterization of the relatedness of inferences that seemed implausible is a little bit more far-fetched, in that the participants imagined that the inference might just be assigned to everyone in a certain demographic. P03 provided an example of this, in reaction to the inference "Parenting" (8 participants reacted similarly):

> *[Parenting]* I don't have any kids and so I was like, why in the world did parenting come up... I mean they can determine your age bracket pretty quickly and if they feel like, okay that should be something that's relevant to majority of people in that age bracket, we're going to add it or something... –P03, Man, Google

Participants' *implausible* reactions occurred when they were unsure about whether an inference applied to them or not, and then justified the presence of a connection with weak, speculative evidence. This evidence was not some recent fact or belief about themselves; rather, it relied on things they were knowledgeable about like their family members' behavior and interests to contextualize the inferences. The constraints on explaining a connection with an implausible-seeming inference involved having an awareness of one's past self, as well as of other people one is close to, or similar to in the case of demographic-based explanations for these inferences.

4.2.3 'No Connection' Reactions Were Unrecognizable

The third theme among the reactions our participants had to the inferences was that they seemed certain some inferences did not apply to them. We call these *no connection* reactions, because participants could not identify a reason why these inferences had been assigned to them. The two examples below, from a Facebook participant and a Google participant, illustrate this type of reaction:

> *[Multicultural Affinity: African American (US)]* Well, I'm not African American... maybe it's something that got put in, maybe it's an unintentional error... or if it is based on an algorithm, then somebody's got some flaws in their system. –P10, Man, Facebook

> *[Rothy's]* I looked at Rothy's because an ad popped up. I think it's shoes. If I remember right it's shoes, and I looked, and I can't buy these. These are way out of my price range. Why are they targeting me? I understand targeting for Walmart. I don't understand targeting me for Rothy's because I am not going to pay 200 dollars for a pair of shoes. –P08, Woman, Google

Participants felt these inferences were unrecognizable, because they had no connection with their own past behavior or self-perceptions, and because they could not imagine other ways to justify why those inferences might have been assigned. Participants could not see themselves in them, and sometimes expressed surprise or disbelief as part of their reaction to them. In addition, participants could not produce evidence or an explanation, as they could in their plausible and implausible reactions, supporting that inference's relationship to themselves. Instead, the evidence and reasons they provided were about why these inferences were *not* connected to them. This type of reaction appeared least often in our data: 56 out of the 321 reactions (17%) that we coded were like

this, and 19 out of 21 participants reacted in this way to an inference at least once (*Mean*=3).

There were several ways in which participants described the lack of connection between themselves and these inferences. Their descriptions illustrate the kinds of evidence and reasoning they were using to understand what the platforms were basing the inferences on. Most often, participants differentiated their activities and interests from their understanding of the meaning of the inference. For example, there were a few inferences for which participants provided a strong refutation in terms that conveyed they never, or hardly ever, do anything like what the particular inference implied about them:

> *[Golf Equipment]* I don't know why that's on there... golf equipment? I don't golf. That's weird. –P04, Woman, Google

> *[Convenience Stores]* Except that I never go [to convenience stores], unless I have to use the restroom. I never go into the store part to buy something. –P09, Woman, Google

Participants also reacted to some of these inferences not by refuting specific activities that the inference implied, but by refuting the idea that they'd be interested in anything related to the inferences at all. Below, a Facebook and a Google participant both describe how their interests do not align with certain inferences:

> *[US politics (moderate)]* I'm just not interested in politics at all, really. Once in a while I'll post something that I find funny that I know a little bit about, but as far as getting down to the truth of politics and stuff, I just don't care. –P16, Woman, Facebook

> *[Apparel]* I mean, I'm not really interested in fashion or clothing all that much, other than as a utilitarian pursuit, something that you need. I'm not particularly concerned about brands or anything like that. –P19, Man, Google

There were some inferences that participants were so confused by that it was difficult to react to them. For example, this was one participant's reaction to such an inference:

> *[Combat Sports]* And combat sports, I have no clue. –P18, Woman, Google

In their *no connection* reactions, participants were unable to establish a causal relation between themselves—their past behavior, interests, relationships, demographics, etc.—and the inferences. Without evidence supporting an interpretation of the inference as plausible, it was difficult for participants to reason about why these inferences were associated with them.

4.3 "Household Income" and Evidence of Aggregation

One particular type of inference made by Facebook provided a unique glimpse into participants' beliefs about how information that might not be readily ascertained from their web use behaviors could be taken into account (or not) when assigning inferences to platform users. Facebook assigned inferences about "Household income: top X% of ZIP codes" to 7 interview participants (and 13 out of 44 Facebook survey respondents). The 'X%' was either 10%, 10%-25%, or 25%-50%. Five interview respondents chose to talk about this inference during the interview. Four participants had *implausible* reactions, and one had a *no connection* reaction. For example, here is the implausible reaction that P13 had to this inference:

> *[Household income: top 10%-25% of ZIP codes (US)]* The household income thing... because that's not something that's public on your Facebook profile. So I'm not really sure where they got that information. That's kind of weird... and we just recently bought a house. So I'm wondering if they somehow got our income information from our mortgage, from our bank. But yeah, that's kind of weird. I mean, I understand why they would want that. But they're digging pretty deep to get that information. It's not readily available on my Facebook. It's like where did you get that? –P13, Woman, Facebook

This participant essentially said that the inference is implausible because Facebook should not be able to figure out her household income from the information that the platform had about her. We (the research team) interpreted this inference type to be indicative not of information about the participants' income specifically, but about the aggregate income level in the zip code in which the participants were located when they connected to Facebook most often. However, in addition to P13, the other four respondents with this inference also seemed to have made the same assumption that it was about their specific household income rather than about incomes in general in their location. Only P11 acknowledged the comparative aspect of this inference, while at the same time relating it to his own income level:

> *[Household income: top 10% of ZIP codes (US)]* I'm not in the top 10% although I live in a community that probably is in the top 10%... I live in [city] so I would consider that would be probably a high income zip, right? So I live there but I don't earn the top 10%, does that make sense? –P11, Man, Facebook

The reactions to the "Household income" inference are an interesting special case among all of the reactions in our data, because of the way this inference was presumably derived. It appears at face value, to an outside observer, to be about the zip code in which the participant lives, and not necessarily about their specific income situation. Our participants did not think this inference applied to them, but this was because they reacted to the inference as if it was about them and not about the community in which they live. The evidence they relied on to establish a connection (or not) between themselves and the inference involved their own beliefs about how wealthy they are compared to others. This interpretation of the inference was likely evoked by "top percent" framing. Therefore, the constraints on explaining this particular inference involve

participants' assumptions that all of the information in their list of inferences is about them personally, as an individual human being.

4.4 Acceptance and Discomfort Coexist

After discussing the specific inferences in their reports, participants were asked about what they thought the purpose of the inferences was. All 21 participants stated that they believed that the purpose of the inferences was primarily for advertising. This is not surprising, since both platforms use the word "ad" in their presentation of the inferences to users. While some participants drew on their personal experiences and knowledge when relating the inferences to the ads they saw online, the majority of participants appeared to be mainly describing a vague awareness gained through hearsay or news coverage. Some participants talked about seeing advertisements that were directly relevant to their search history and online activity as they were browsing, but did not offer overly technical explanations. For example:

> Google's like probably getting our search histories or whatever and then maybe selling them to companies and that's how like the ads pop up on the side of your screen, like that these may interest you. –P18, Woman, Google

Eleven participants explicitly stated their belief that inferences are part of the platforms' business model, and assumed a profit motivation for the companies. For example, P21 (Woman, Google) observed that since they don't pay to use Google's services, it seems reasonable that the platform must have an alternative way of earning revenue to support its operations. And P11 said:

> I'm assuming they're monetizing it in some way. They're probably selling it to companies so that they can target ads to me, the consumer, to everybody else as consumers. – P11, Man, Facebook

This is in line with previous research that has found a general awareness of targeted advertising, even if users are not necessarily able to articulate the mechanism behind it [5, 29, 35].

While the majority of participants (17 out of 21) appeared to accept the idea of information collected about them being used for advertising, 20 participants also expressed feelings about the inferences which indicated a level discomfort with some of them. Acceptance of the business model and discomfort with specific inferences were not mutually exclusive, and reactions of discomfort and acceptance were often even expressed in the same sentence. P07 referred to the "evil empire" and asserted that he did not like being the target for ads, but also did not take take great offense to receiving targeted ads. Some participants explicitly referenced the idea of the "world" (P01) or "most people" (P22) during their interviews, suggesting that data collection is something that's going on at a grand scale over which they felt they had little control. This seemed to play a role in their willingness to accept information being collected about them, but it was also a source of discomfort. For example:

> [It makes me uncomfortable] that they can get so many specific things without me realizing that people have it. But, I kind of know that. I wish they didn't know anything. I would rather be anonymous, but I know that's not our world. –P01, Woman, Facebook

> I think most people feel the same way that I do. I mean, there's a level of acceptance that it happens, but then the more you think about it, it kind of starts to disturb you a little bit more because when you see lists like this, then some of the things that they make connections for, it takes you a while in your head to get to how they got to that. And it's a little bit, I don't know, disconcerting or something. It's just a little bit uncomfortable. –P22, Woman, Google

Highly specific inferences which participants could connect to non-typical daily activities also provoked speculation and reactions of discomfort. Participants' comments about these inferences suggested the discomfort came from being made explicitly aware of the extent and detail to which their actions are being tracked by platforms. This also served as an opportunity for participants to concretely self-reflect on their online activities to an extent they appeared to not have previously done before. Although they knew information was being collected, and sometimes even had experiences seeing targeted ads that allowed them to deduce thair their shopping habits were being tracked in some way, it was still different to see the information aggregated in one place.

> On the one hand, I understand it, but at the same time it kind of makes you think like, wow, they really can actually key in on very specific things... everything else was so broad and then that one [inference] really stuck out because it was the one thing out of everything that was very, very specific to a span of time of search history... I mean, it's not like it's so personal that I'm freaking out about it. It's just that it brought to my attention how much data is floating around out there that Google or Facebook or whoever is able to capture and is aware of about you that you might've even long since forgotten about. –P03, Man, Google

> [It made me feel] a little bit in the spotlight. A little bit like yeah, they know what I'm doing. A little uncomfortable I guess... I'm not surprised, but you just don't think about what you're leaving, the tracks that you're leaving. You don't think about it when you're doing it. –P08, Woman, Google

Some participants appeared to be uneasy about the breadth and variety of inferences associated with them, indicating a platform's ability to know things about them to an extent nobody else does. This feeling appeared to be related to inferences that participants considered to be accurate or relevant to themselves, suggesting that accurate information can feel

uncomfortable when a great amount of it is visibly collected in one place:

> As I went through this I was like, oh my God look at all the information they've gathered about me... Google knows more about me than anybody else. It's scary. –P17, Woman, Google

> No matter like where you... They track your movements or they know what you've searched for or what you're interested in. So I mean probably my husband doesn't even know that, right? So I find it a little creepy. –P15, Woman, Facebook

In contrast to this, some participants expressed little discomfort with the existing inferences. One reason commonly cited was that the categories appeared to be harmless, or sufficiently generic that it could not reflect badly on them or compromise their personal information. For example, P08 (Woman, Google) talked about not caring what information the platforms have collected about her, because "I don't think that it's necessarily done to hurt people."

5 Discussion

Our findings focused on participants' reactions to the inferences Facebook and Google had made about them, and the evidence they used in their explanations for how and why the inferences were related to them. Generating explanations is one way that people increase understanding of phenomena they encounter, by helping them to predict and make sense of future events and situations [15]. Participants' explanations for the inferences they were assigned provided visibility into the knowledge they used to try to make sense of the inferences, and the new understandings that resulted [33].

5.1 Three Ways of Framing Inferences

Our findings show that the understandings participants demonstrated of the inferences assigned to them are constrained by their own awareness of the information the inferences might be generated from, at the time they are considering the inferences. P14 actually had a moment of insight related to this during his interview:

> I have a narrow thought process. For example, when we talked about sports. The idea that I'd be thinking about sports I play and not the sports I might watch or might have watched. –P14, Man, Google

This insight happened when the participant was asked by the interviewer to compare the information in the inferences with the kinds of information that a person would normally enter when filling out an online profile. In response, he described his thinking when he had reacted to the sports that were listed among his inferences (e.g., "Basketball," "Hockey," "Golf"). He indicated that his understanding of the meaning of these inferences was initially narrower—framed by his recollection of sports he plays—than what the inference might actually represent to the platform.

There were three broad framings which participants used to reason about their connection to the inferences assigned to them. The inferences were framed as 1) being related to past and present online behavior and interests, 2) stemming from their relationships with others they are close to, and 3) inaccurate, and therefore not useful. The extent to which participants recollected or speculated about evidence consistent with one of these framings determined whether they had a *plausible*, *implausible*, or *no connection* reaction to an inference.

5.1.1 Past vs. Future

The most prevalent framing, that inferences are related to online behavior and interests, demonstrates an assumption by our participants that inferences are descriptive of their pasts. But, advertisers use inferences to exert control over people's future attention and actions through the ads that are targeted to them [36]. This means that the goal of inferences is more about labeling people so that they can be targeted, rather than creating a representation that agrees with how people would describe themselves. If people assume that inferences are descriptive rather than predictive, that assumption leads to explanations and understandings that fail to anticipate the true purpose of the inferences and could lead to people deeming inferences as inaccurate or inapplicable that could actually be effective for targeting by the platforms and institutions that use them.

One aspect of the platforms' user interfaces that may reinforce this impression, ironically, is the explanations of the inferences provided by the platforms themselves. Both Google's and Facebook's help webpages describing how ads are targeted focus on the data the inferences are based on, e.g. "Adding a product to a shopping cart or making a purchase" (Facebook)[6] or "Previous search activity" (Google)[7]. One way to convey the predictive nature of inferences to end users could be to instead use language that describes the inferences as predictions or guesses. It might also be possible for platforms to provide an accuracy score or indicator for each inference to indicate how well each user's behavior since the inference was assigned aligned with the prediction. This might convey a different framing that emphasizes an orientation toward the future rather than the past.

5.1.2 Individual vs. Aggregate

The second framing, that inferences are related in some way to people's close relationships, shows that our participants were constrained in how they reasoned about about aggregation as it related to how inferences were assigned. Our participants said that they sometimes did things online that reflected others' interests and not their own, and felt that this could result in the interests of people they were close to being be mis-attributed to them. However, this framing is still based on the unit of analysis being the individual person, and thus their belief

---

[6] https://www.facebook.com/ads/about/?entry_product=ad_preferences
[7] https://support.google.com/accounts/answer/1634057

that the mistake in generating the inference was related to mapping the interests of one person onto another.

There were a few instances where participants felt like interests were mapped by the platform onto broad groups of people according to "demographic" similarities—like "Parenting" being associated with participants that did not have children living at home with them or had no children (P03, P04, P08, P11, P12, P14, P19). But the demographic indicators they mentioned in rationalizing these inferences were the ones that were most salient to them, e.g., age, gender, and income level. This means that the framing of inferences as related to close relationships or similar others according to obvious demographic characteristics underestimates the ability of machine learning models to infer other features that don't fit into this framing. Thus, targeting is likely based on attributes that people cannot envision [2].

The key insight absent from participants' reasoning consistent this framing was that behavioral patterns of complete strangers can be used to infer and assign attributes to them. Facebook and Google both reinforce an understanding of inferences as belonging to individuals rather than being derived from aggregate patterns of characteristics by using language like "your activity" in the platforms' explanations. Providing some information about the proportion of other users who share inferences in common with an individual could help to convey a broader perspective of the similarities across users that the inferences are based on.

5.1.3 Accurate vs. Useful

The third framing, that inferences that are not accurate or explainable are not useful, indicates a presumption that the inferences should be accurate, or that the intention of the inferences is to describe users accurately. There were several examples of this in our data; the reaction to the "Rothy's" inference on page 8 is one. This framing assumes that platforms should not want to associate inaccurate inferences with users. However, it is possible that mis-targeting a few people could still be beneficial for platforms' goals in aggregate. Of course, we cannot know from this study whether the platform or advertisers think the inferences are either accurate or useful, but one can imagine that a particular inference doesn't need to be 100% accurate for all users it is assigned to in order to serve its intended purpose from the platform's perspective. This means that a framing that inferences should be accurate prevents people from speculating about situations in which inaccurate inferences might actually be useful or even intentional. One way that platforms could attempt to convey the notion that even inaccurate inferences might be beneficial from their perspective is to include information about the overall profitability (or other platform success metric) of various inferences that a user has been assigned, along with the accuracy rating mentioned above. This could serve to make the platform's stake in the inferences more transparent.

5.2 Implications for Privacy Decisions about Consent

Users' implicit assumptions and guesses about the data collection and inference generation they are consenting to may vary based on their existing knowledge and understanding of cause and effect related to the consent decision they need to make. For example, people cannot envision inferences that may be related to their potential future interests, or derived via aggregation with data from people they don't know, or that are inaccurate from their perspective but still useful for the platform. And yet, notice and choice demands that users provide one-time, up front consent to incomplete descriptions of platforms' data practices. New privacy legislation mandates improved transparency (while retaining the notice and choice model), but our findings show that the framing by which people explain and understand the inferences does not support them in envisioning the inferences from the platforms' perspective.

Without framing the inferences from the perspective of advertisers and platforms and imagining how they might put the inferences to use, people are unlikely to be able to give informed consent to many of the inferences they are assigned. Knowing the inferences are used for advertising is not enough for consent, because in our study participants were broadly aware of that and still did not achieve that understanding. Hence the discomfort when they began to realize the amount and specificity of the information the platform must have amassed about them. Based on this work, we argue that the understandings people develop about inferences through the 'Ad Settings/Preferences' model of transparency are unlikely to help them realize that what they are really consenting to is allowing the platform to make whatever inferences it wants to about their future, to target them based on the behavior of people they don't even know exist, and to profit from inaccurate assumptions about them.

Our study shows that there are types of inferences which are straightforward for people to understand and anticipate: inferences that seem *plausible* because of their relationship to easily recalled or past actions. Platforms which place a priority on obtaining true consent should therefore restrict the inferences they make about users to those which fit the constraint of being explainable by users, given their normal use of the platform. It may be possible to expand user understanding of inferences through framing them differently in the ways suggested in this paper. Future research is needed to explore the effectiveness of lightweight interventions such as these for expanding the range of inferences that seem plausible to users.

Acknowledgments

References

[1] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, and et al. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Comput. Surv.*, 50(3), 2017.

[2] Athanasios Andreou, Márcio Silva, Fabrício Benevenuto, Oana Goga, Patrick Loiseau, and Alan Mislove. Measuring the Facebook Advertising Ecosystem. In *Network and Distributed Systems Security Symposium (NDSS)*, 2019.

[3] Sophie C. Boerman, Sanne Kruikemeier, and Frederik J. Zuiderveen Borgesius. Online Behavioral Advertising: A Literature Review and Research Agenda. *Journal of Advertising*, 46(3):363–376, 2017.

[4] José González Cabañas, Ángel Cuevas, and Rubén Cuevas. Unveiling and Quantifying Facebook Exploitation of Sensitive Personal Data for Advertising Purposes. In *Proceedings of the 27th USENIX Security Symposium*, pages 479–495, 2018.

[5] Farah Chanchary and Sonia Chiasson. User Perceptions of Sharing, Advertising, and Tracking. In *Proceedings of the Eleventh Symposium On Usable Privacy and Security (SOUPS)*, pages 53–67, 2015.

[6] Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies*, pages 92 –112, 2015.

[7] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting Depression via Social Media. In *Proceedings of the 7th International Conference on Web and Social Media*, 2013.

[8] Michael A. DeVito, Jeremy Birnholtz, Jeffery T. Hancock, Megan French, and Sunny Liu. How People Form Folk Theories of Social Media Feeds and What It Means for How We Study Self-Presentation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018.

[9] Claire Dolin, Ben Weinshel, Shawn Shan, Chang Min Hahn, Euirim Choi, Michelle L. Mazurek, and Blase Ur. Unpacking perceptions of data-driven inferences underlying online targeting and personalization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018.

[10] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. First I "like" It, Then I Hide It: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2371–2382, 2016.

[11] Motahhare Eslami, Sneha R. Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. Communicating Algorithmic Process in Online Behavioral Advertising. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2018.

[12] Casey Fiesler and Blake Hallinan. "We Are the Product": Public Reactions to Online Data Sharing and Privacy Controversies in the Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2018.

[13] Harold Garfinkel. Studies of the Routine Grounds of Everyday Activities. *Social Problems*, 11(3):225–250, 1964.

[14] Jian Hu, Hua-Jun Zeng, Hua Li, Cheng Niu, and Zheng Chen. Demographic Prediction Based on User's Browsing Behavior. In *Proceedings of the 2007 World Wide Web Conference*, 2007.

[15] Frank C. Keil. Explanation and understanding. *Annu. Rev. Psychol.*, 57:227–254, 2006.

[16] Predrag Klasnja, Sunny Consolvo, Tanzeem Choudhury, Richard Beckwith, and Jeffrey Hightower. Exploring Privacy Concerns about Personal Sensing. In *Proceedings of the International Conference on Pervasive Computing*, pages 176–183, 2009.

[17] Michal Kosinski, David Stillwell, and Thore Graepel. Private Traits and Attributes are Predictable from Digital Records of Human Behavior. *PNAS*, 110(15):5802–5805, 2013.

[18] Oksana Kulyk, Annika Hilt, Nina Gerber, and Melanie Volkamer. This Website Uses Cookies: Users Perceptions and Reactions to the Cookie Disclaimer. In *Proceedings of the 3rd European Workshop on Usable Security*, 2018.

[19] Pedro Giovanni Leon, Justin Cranshaw, Lorrie Faith Cranor, Jim Graves, Manoj Hastak, Blase Ur, and Guzi Xu. What Do Online Behavioral Advertising Privacy Disclosures Communicate to Users? In *Proceedings of the ACM Workshop on Privacy in the Electronic Society*, pages 19–30, 2012.

[20] Pedro Giovanni Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. What

Matters to Users? Factors That Affect Users' Willingness to Share Information with Online Advertisers. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, 2013.

[21] Adam Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In *Proceedings of the 25th USENIX Security Symposium*, 2016.

[22] Chanda Phelan, Cliff Lampe, and Paul Resnick. It's Creepy, But It Doesn't Bother Me. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 5240–5251, 2016.

[23] Emilee Rader. Awareness of Behavioral Tracking and Information Privacy Concern in Facebook and Google. In *Proceedings of the Tenth Symposium On Usable Privacy and Security*, pages 51–67, 2014.

[24] Emilee J Rader and Janine Slaker. The Importance of Visibility for Folk Theories of Sensor Data. In *Proceedings of the Thirteenth Symposium on Usable Privacy and Security*, 2017.

[25] Ashwini Rao, Florian Schaub, and Norman M. Sadeh. What do they know about me? Contents and Concerns of Online Behavioral Profiles. In *Proceedings of the 6th ASE Conference on Privacy, Security, Risk and Trust*, 2014.

[26] Johnny Saldaña. *The Coding Manual for Qualitative Researchers*. Sage, 2015.

[27] H. Jeff Smith, Sandra J. Milberg, and Sandra J. Burke. Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *MIS Quarterly*, 20(2):167–196, 1996.

[28] Daniel J Solove. Privacy Self-Management and the Consent Dilemma. *126 Harvard Law Review*, pages 1880–1903, 2013.

[29] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, 2012.

[30] Kaan Varnali. Online Behavioral Advertising: An Integrative Review. *Journal of Marketing Communications*, 2019.

[31] Jeffrey Warshaw, Nina Taft, and Allison Woodruff. Intuitions, Analytics, and Killing Ants: Inference Literacy of High School-educated Adults in the US. *Proceedings of the Twelfth Symposium On Usable Privacy and Security*, 2016.

[32] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L. Mazurek, and Blase Ur. Oh, the Places You've Been! User Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, pages 149–166, 2019.

[33] Robert A. Wilson and Frank C. Keil. The Shadows and Shallows of Explanation. In Frank C. Keil and Robert A. Wilson, editors, *Explanation and Congnition*, pages 87–114. MIT Press, 2000.

[34] Yuxi Wu, Panya Gupta, Miranda Wei, Yasemin Acar, Sascha Fahl, and Blase Ur. Your Secrets Are Safe: How Browsers' Explanations Impact Misconceptions About Private Browsing Mode. In *Proceedings of the 2018 World Wide Web Conference*, pages 217–226, 2018.

[35] Yaxing Yao, Davide Lo Re, and Yang Wang. Folk Models of Online Behavioral Advertising. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1957–1969, 2017.

[36] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books, 2019.

Appendix

A Survey Questions

Participants were provided with a description of the survey and asked for consent before answering screening questions. If eligible, they were directed to the beginning of the survey. Respondents were informed about how long the survey would take, that they would be asked to log in to Facebook or Google and save and then upload a file, that there were attention check questions, and that they were free to withdraw from the study at any time.

A.1 Demographics and Screening

1. What gender do you identify as?
   - Man
   - Woman
   - Other (fill in the blank)
   - Prefer not to answer

2. What is the last grade or class you completed in school?
   - None, or grades 1-8
   - Some high school
   - High school graduate or GED certificate
   - Technical, trade, or vocational school AFTER high school
   - Some college, no 4-year degree
   - 4-year college degree
   - Some postgraduate or professional schooling, no postgraduate degree
   - Postgraduate or professional degree, including master's, doctorate, medical or law degree

3. What was your total household income before taxes during the past 12 months?
   - Less than $25,000
   - $25,000 to $34,999
   - $35,000 to $49,999
   - $50,000 to $74,999
   - $75,000 to $99,999
   - $100,000 to $149,999
   - $150,000 to $199,999
   - $200,000 or more

4. Which categories below best describe you? Select all that apply:
   - White
   - Hispanic, Latino or Spanish
   - Black or African American
   - Asian
   - American Indian or Alaska Native
   - Middle Eastern or North African
   - Native Hawaiian or Other Pacific Islander
   - Some Other Race, Ethnicity or Origin (please specify)

5. What is your current employment status?
   - Employed full time
   - Employed part time
   - Unemployed looking for work)
   - Unemployed not looking for work
   - Retired
   - Student
   - Disabled

6. Have you ever received formal training in computer science, software engineering, IT, computer networks, or a related technical field?
   - Yes
   - No

A.2 Instructions to save Ad Preferences (Facebook) or Ad Settings (Google) page

1. Click here to open [URL to Ad Preferences/Ad Settings] in a new window. Log in to Facebook/Google if you are prompted to do so.

2. Save that page as "HTML Only" (ctrl+S on Windows, Command+S on Mac, OR right click on the page –> Save As/Save Page As...).

3. Make sure you save the file somewhere it will be easy for you to find it. On the next screen you will submit the file you just downloaded to the survey, which will automatically generate questions for you based on the content of the file.

4. After you have saved the file, take a quick look at the information on the web page. Then, please close the browser tab or window you used to save the file and return here to complete the rest of this survey. If you are not using your own computer, don't forget to log out.

5. Have you ever seen your [Facebook Ad Preferences/Google Ad Settings] web page before?
   - Yes, I have seen it before.
   - No, I have not seen it before.
   - I'm not sure.

A.3 Questions about each Inference

After participants submitted their file, it was uploaded to our server where it was parsed to extract only the inference labels for "active" inferences. The inference labels were then sent back to the survey, where three questions were generated for each inference. We limited the total number of inferences asked about in the survey to 85, which were selected randomly from a respondent's inferences if they had more inferences than this. The three responses for each inference were then averaged to produce the reports used for interviews. See Appendix C for example reports.

1. *inference* is relevant to who I am as a person. [7-point Likert, Strongly agree–Strongly disagree]

2. *inference* is an accurate description of my everyday activities. [7-point Likert, Strongly agree–Strongly disagree]

3. It makes sense that *inference* is associated with me. [7-point Likert, Strongly agree–Strongly disagree]

A.4 Attention Check Questions

The survey included up to three attention check questions. Each page of questions about the inferences asked the three questions about five inferences, for a total of 15 questions per page. The attention check questions were inserted among these questions, on the third, eighth, and fifteenth page of questions about the inferences, if the participant had enough inferences to reach that point in the survey. Responses that failed to pass the checks were excluded.

1. To help us monitor the quality of our data, please select "Somewhat agree" from the choices below. [7-point Likert, Strongly agree–Strongly disagree]

2. To help us monitor the quality of our data, please select "Strongly disagree" from the choices below. [7-point Likert, Strongly agree–Strongly disagree]

3. Which website did you save a web page from earlier in this survey? [Select one: Pinterest, Facebook, Reddit, Twitter, Google]

A.5 Questions about Social Media and Online Behaviors

1. When was the last time you viewed the [Facebook Ad Preferences/Google Ad Settings] web page, not including for this survey?
   - Today
   - Yesterday
   - Within the past week
   - Within the past month
   - More than a month ago
   - I have never viewed the Facebook Ad Preferences web page before

2. Do you use an ad blocker when you browse the web?
   - Yes
   - No
   - I Don't Know

3. Have you ever had one of the following experiences? Select all that apply:
   - Fell victim to a phishing email message or other scam email

   - Received a notification from a company that your information was involved in a data breach
   - Had a virus on your computer or mobile device
   - Someone broke in or hacked your computer, mobile device, or account
   - Stranger used your credit card number without your knowledge or permission
   - Identity theft more extensive than use of your credit card number without permission
   - None of the above

4. How much negative information have you heard about [Facebook/Google] in the past several months?
   - A huge amount
   - A lot
   - A moderate amount
   - A little
   - None at all

5. Has how often you've used [Facebook/Google] increased, decreased or stayed about the same recently?
   - Increased a lot
   - Increased a little
   - Stayed about the same
   - Decreased a little
   - Decreased a lot

A.6 Privacy Questions

As part of the survey, we asked questions from an instrument developed by Smith, Milberg and Burke [27] to identify and measure concern about privacy practices. These questions were not analyzed for this paper. These questions were asked after the first question in the previous section.

Here are some statements about personal information. From the standpoint of personal privacy, please indicate the extent to which you, as an individual, agree or disagree with each statement. [7-point Likert, Strongly Agree–Strongly Disagree]

- It usually bothers me when companies ask me for personal information.
- When companies ask me for personal information, I sometimes think twice before providing it.
- It bothers me to give personal information to so many companies.
- I'm concerned that companies are collecting too much personal information about me.
- Companies should not use personal information for any purpose unless it has been authorized by the individuals who provided the information.
- When people give personal information to a company for some reason, the company should never use the information for any other reason.

- Companies should never sell the personal information they have about people to other companies.
- Companies should never share personal information with other companies unless it has been authorized by the individuals who provided the information.
- All the personal information companies have about people should be double-checked for accuracy, no matter how much this costs.
- Companies should take more steps to make sure that the personal information they have about people is accurate.
- Companies should have better procedures to correct errors in personal information.
- Companies should devote more time and effort to verifying the accuracy of the personal information they have about people.

B Descriptives about the Inferences

This appendix provides descriptive information about the Facebook and Google inferences assigned to study participants. We present figures to show how these inferences normally appear on the Facebook Ad Preferences and Google Ad Settings pages before they were parsed, summary statistics for inferences across platforms, and both the most common and most uncommon inferences in our dataset.

B.1 Inferences Web Pages

The figure below shows a screen capture from the Facebook "Ad Preferences" and Google "Ad Settings" of one of the authors. In June 2020, these pages could be found at:

- Google: https://adssettings.google.com/
- Facebook: https://www.facebook.com/ads/preferences/

(a) Facebook inferences (b) Google inferences
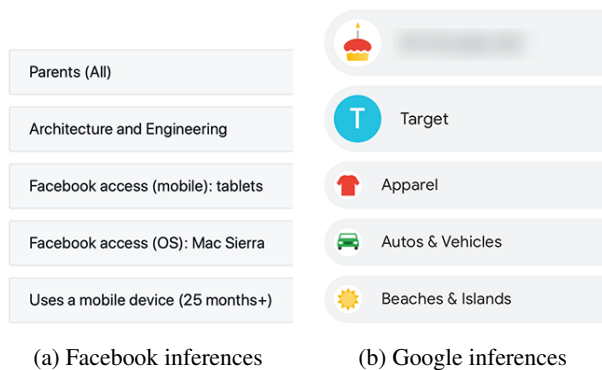
B.2 Average Inferences Per Survey Respondent

Summary statistics for inferences across Facebook and Google:

Service	Mean	SD	Min	Median	Max
Facebook	11.97727	9.86063	3	4	33
Google	67.60784	35.08451	1	66	131

B.3 Common and Uncommon Facebook Inferences

There were 110 unique Facebook inferences in the survey data. 41 were assigned to only one respondent. Here are the most common inferences assigned to Facebook survey respondents (n=44), ranked from 1-12 by number of participants.

Facebook Inference	*N*
Mobile network or device users	42
Potential mobile network or device change	42
Recent mobile network or device change	42
WiFi users	42
Facebook access (mobile): all mobile devices	20
Facebook access (mobile): smartphones and tablets	19
Facebook access (network type): 4G	14
Facebook access (network type): WiFi	13
Gmail users	12
Engaged Shoppers	11
Facebook access (mobile): Apple (iOS)devices	10
Likely engagement with US political content (liberal)	10

The table below shows 12 inferences randomly selected from the 41 Facebook inferences that were only assigned to one person in the survey data.

Facebook Inferences	*N*
Owns: Galaxy Tab 4	1
Arts, Entertainment, Sports and Media	1
Management	1
Newly engaged (1 year)	1
Owns: iPhone 7 Plus	1
Owns: ZTE	1
Anniversary within 30 days	1
Lived in UK (Formerly Expats-UK)	1
Business and Finance	1
Lived in United States (Formerly Expats-United States)	1
Business page admins	1
Returned from travels 1 week ago	1

B.4 Common and Uncommon Google Inferences

There were 561 unique Google inferences in the survey data. 294 were assigned to only one respondent. Here are the most common inferences assigned to Google survey respondents (n=51), ranked from 1-12 by number of participants.

Google Inference	N
Parenting	46
Home & Garden	43
Shopping	43
Business Services	41
Holidays & Seasonal Events	41
Performing Arts	39
Books & Literature	38
Family & Relationships	38
Visual Art & Design	38
Computers & Electronics	37
Cooking & Recipes	35
American Football	33

The table below shows 12 inferences randomly selected from the 294 Google inferences that were only assigned to one person in the survey data.

Google Inferences	N
Tile Games	1
Cincinnati	1
Food	1
Historical Sites & Buildings	1
Interior Design	1
GetYourGuide	1
Oceania	1
Perfumes & Fragrances	1
45+ years old	1
OfficeSupply.com	1
Morgans Hotel Group	1
Refrigerators & Freezers	1

B.5 Inference Ratings

The next table shows the mean and standard deviation of the ratings for each of the three questions (sensible, relevant, and

Platform	Mean	SD
Sensible:		
Facebook Inferences	5.00	1.78
Google Inferences	4.29	2.08
Relevant:		
Facebook Inferences	4.68	1.87
Google Inferences	4.16	2.03
Accurate:		
Facebook Inferences	4.64	1.88
Google Inferences	3.84	2.05

accurate) for both platforms.

The table below shows Pearson correlations between the inference ratings. The sensible, relevant, and accurate ratings of the inferences were highly correlated with each other. All correlations are statistically significant ($p < .001$).

	Facebook	Google
sensible, relevant	0.84	0.93
relevant, accurate	0.85	0.89
sensible, accurate	0.84	0.88

C Interview Protocol and Example Reports

Participants were interviewed about their reactions to the inferences they had been assigned by Facebook or Google using the semi-structured interview protocol on the next page. Each participant was given a report that presented the inferences they had answered questions about in the survey. Responses to the three questions about each inference were averaged, rounded to the nearest whole number, and then color-coded according to the 7-point Likert agreement scale used for the responses to the questions. Inferences in the Facebook reports were displayed in one grouping. Inferences in the Google reports were grouped into higher-level categories that were present in the labels of the inferences parsed from respondents' Ad Settings pages. Within each group, inferences were displayed to participants in order from Strongly Agree-Strongly Disagree. The reports included in this appendix were generated from the inferences and survey data of one of the authors.

Interview Guide and Protocol

General Background and Instructions

The goal of the interviews:

- The goal of the interviews is to collect data about people's beliefs about how the categories about them were assigned, and about their reactions to the categories, IN THEIR OWN WORDS.
- We want to know what they think and believe about the following kinds of things:
  - What the information is on the ad preferences page
  - How the information was chosen
  - How it feels to have this information associated with them
  - What the information IS being used for
  - What the information COULD be used for

About priming or leading the participant:

- It is important that the interviewer does not convey anything to the participant about what they know or believe about how the categories may be determined, or even what kinds of words to use to refer to what the platforms refer to as "categories".
- In other words, don't even call them "categories"; this may prime or bias the participants.
- This means that the interviewer must pay careful attention to the language each participant uses during the interview, and refer to the same concepts the participant talks about using the same kinds of words as the participant.
- If we indicate to them the kinds of vocabulary they should use, by using it ourselves, or if we ask leading questions, the data we collect will be about participants reactions to OUR beliefs, not about their own beliefs.
- The more we guide their responses, the more we will be collecting data about something they're only thinking about because we asked them to think about it. We want to know what they think about this, not what they think about what WE think about this.

Good ways to probe/follow up for more information:

- "Tell me more about [X]"
- "What do you mean by [X]"
- "How do you think [X] happens"
- "Where do you think [X] came from"
- "Can you give me a specific example of [X]"
- "What would [X] look like"

Things to listen for and follow up on in their answers:

The following are all common things that come up in interviews about privacy-related topics. These things are not useful data for this study if people just mention them and then quickly move on. But, if the interviewer can ask follow up questions and probe for their beliefs and thinking related to these things, the data will be more interesting and useful for identifying patterns and differences in people's beliefs.

So, for example, it isn't very useful to learn that people think having inferences assigned to them is creepy, or that they're concerned about it, without knowing more about why. Likewise, it isn't very useful to us to know that people think a certain kind of information is more sensitive than some other kind of information. We already know things like this from previous studies other people have done, and it doesn't help us figure out how people think systems are able to infer things about them or why certain inferences have been assigned to them.

But, what IS useful is learning about what makes it feel creepy, or what makes the information sensitive for them, and connecting that to their beliefs about where the inference categories come from and how they are assigned. So it is especially important to ask follow-up questions about topics like these:

- assumptions about why Facebook or Google assign categories, or how they will be used
- saying something is creepy or invasive

Interview Guide and Protocol

- commenting on what/who Facebook or Google thinks they are, as a person
- emotional reactions, like anger, irritation, amusement, pride, etc.
- talking about accuracy or mistakes related to the categories
- feeling surprised about any of the categories
- talking about what's missing from the list of categories

Tips for effective interviewing:

- *It is OK to wait for people to answer when you've asked them a question.*
  People may need to think for a minute before answering some of these questions, and if there is a pause in conversation while they do that it may feel a bit awkward. This is OK. The best way to give a participant space to answer is to remind yourself to PAUSE and let them think, even if the silence makes you uncomfortable. If you move on, and ask another question, they won't answer the first question! Count to 10 in your head if you have to.

- *Never interrupt the person you are interviewing!*
  This may mean that it feels like the person may be rambling into something that is off topic for the interview protocol. But, that doesn't mean the data won't be useful, and if you cut them off you will never know what they were going to say. People think out loud sometimes, and the process of talking about something is important for the process of thinking about it. Also, interrupting someone conveys to them that you weren't actually that interested in what they were saying, and which is absolutely the LAST thing we want interview participants to feel. We are VERY interested in what they have to say!

- *If you finish the interview feeling like " that was a nice conversation" you're not listening carefully enough.*
  Interviews should be tiring, because you should be trying to pay attention to **everything** the person is saying and thinking about how to follow up. This takes a lot of energy and focus. You shouldn't be thinking about other stuff going on in your life during the interview -- focus all of your attention on the participant, and asking good follow-up questions.

- *We really want the participants to actually think about the inferences during the interview.*
  Sometimes people's first response might be "I don't know" or "I have no idea" or "I've never thought about that before". This is because we're showing them information they really may not have thought about much before! It is important to follow up when they say that, don't just let it go! Some ways to follow up and get them talking about what they're thinking are: "Tell me more about that." or "Why do you think you haven't thought about it?" or "What is it that makes it hard to answer this question?" or "We're really interested in anything you can tell us about your thoughts about this."

Interview Guide and Protocol

Interview Protocol

At the beginning of the interview:

1. Introduce yourself and thank the participant for coming.

2. Tell them a brief overview of the study and why they're there. Something like:
   o We're going to be asking you some questions today about yourself and about your perceptions of the information available about you in popular web platforms like Facebook and Google.
   o Before we get started, I have a consent form here that I need you to look over and sign. It describes the study at a high level, and lets you know about what you will do, what your rights are as a participant, and about the Amazon gift card you'll receive at the end of the study.
   o I want to emphasize that what you say to us during the interview will be kept confidential, and you can stop participating at any time, just let us know.

3. Give them the consent form and a pen, and wait for them to sign it.

4. After they've signed the consent form, check it to make sure they have given consent to the audio recording before turning on the recorders. Ask them if they have any questions before starting.

5. Turn on the audio recorders once they've consented to both the interview AND the recording. DON'T FORGET TO TURN ON THE RECORDERS! Yes, both recorders.

6. Let them know when you've turned on the audio recorders.

1. Warm-Up Questions:

1. Tell me a little bit about yourself…
   o Want to probe for things like
     ▪ What they do for a living
     ▪ What their educational background is
     ▪ How much they use computers and/or mobile devices and the Internet
   o Goal of this question is to put the participant at ease (everybody always starts off a little nervous and unsure about what to expect) and learn background information that will help us to describe the participants in general.

2. You recently filled out a survey for us – that's why you're here! We're going to be talking about some information from that survey in a minute, but I wanted to ask you first about:
   o What do you remember off the top of your head about that [Facebook | Google] web page that the survey asked you to save?
   o Had you ever looked at that web page before you filled out the survey? When was the last time you went to that page? Tell me more about that time… Do you remember why you looked at it? What your impression at that time was? Was that the only time? What other times did you look at it? Etc.

2. Questions about the categories:

1. Give the report to the participant—but don't give them the page comparing their categories to other people's categories yet! Tell them:
   o Here's a report that we generated from the answers you filled out to the questions in the survey you did for us about the information about you on Facebook or Google. Take a few minutes to look this over, and choose two of the categories that we'll start off talking about. Pick whatever ones you like, we're interested in what stands out to you about this information. Let me know when you're done. (If they start talking during this, just go with the flow!)

Interview Guide and Protocol

2. Then ask them to talk about each of the things they picked—first one, then the other one – we want to make sure we end up with answers to each of these questions about each of the categories they choose to talk about:
   - **What** stood out to you about that? Why did you pick it to talk about?
   - **Why** do you think that information was in your list?
   - **How** do you think that information was chosen?

3. Then, ask them to mark or show you any other categories that stand out to them for any reason in the report. Ask them the **what**, **why** and **how** questions about each of these categories too.

4. Next, we want to know about whether any of the information violated their expectations in any way. They may have talked about some of these things (surprising, inaccurate, uncomfortable, missing, confusing) already, if so, you don't need to ask about them again. For each bullet point below that you end up talking with them about, ask the **why** and **how** questions.
   - Is there any information on here that is **surprising** to you? Why?
   - Is there anything you think might be on here by **mistake**? Why?
   - Is there any information here that makes you **uncomfortable**? Why?
   - Is there any information **missing**, that you would have expected to see but isn't on here? Why?
   - Is there any information on here that you find **confusing**? Why?

5. Finally, show them the page comparing their categories with other people's categories. Be really, really careful not to ask leading questions about these categories!
   - Ask about the first set – the categories that are unique to them. *Note that if this list is long and the participant seems like they are getting tired of talking about categories, just ask them which ones stand out and talk about only those.*
     - Ask the **why** and **how** questions about these categories, for the ones you haven't already talked about with them.
   - Compare the first set (unique to you, Section 1) to the second set (common for other people in our survey, but not associated with you, Section 2):
     - What's the first thing that stands out to you? Why?
   - Next, ask them to take a look at Section 3 under "How your information compares with others":
     - This section compares the information that was assigned to you, with the information Google | Facebook assigned to other people who completed the survey. It shows information that everyone had, versus information that you and only a few others had.
     - I'd like you to look at this information and think for a minute about how accurate it might be. How would you say these lists compare, from that perspective?

3. Big-picture questions:

1. What do you think about what the purpose of the information is? Why do you think so? Tell me more about that…
   - *If they have trouble answering the above question, try:* How do you think it is being used by Facebook or Google right now? What do you think it could be used for?
   - If the participant says something about third parties using the information, here are some follow-up questions: Can you give me a specific example of that? Describe for me how that works. What would [third party] be doing with that information?

2. How does the information we've been talking about today from [Facebook | Google] compare with the information you typically put in when you fill out an online profile for a website account? Can you give me an example of the kind of information in the profile of one of your online accounts? How is that similar/different?

3. If there were someone who works for [Facebook | Google] sitting here with us, what would you like to ask them about this?

Interview Guide and Protocol

4. Questions about Facebook or Google usage:

1. Facebook
   - When was the last time you used Facebook? Tell me about that.
   - Is that typical of what you use Facebook for? Can you tell me about another example of when you used Facebook like that?
   - What other ways do you use Facebook? Can you give me a specific example? Tell me about that.
   - How do you use the "Like" button on Facebook? Can you give me an example? Tell me about that.
   - What are your impressions of the advertising that you see on Facebook? Can you give me an example? Tell me about that.
   - What would you do if you wanted to change or remove some of this information about you that Facebook has? *(If they don't know, they may ask you how to do it.)*

2. Google
   - Google has a lot of different "products"; in your survey you mentioned that you used products X, Y, and Z. Which one do you use the most?
   - When was the last time you used X? Tell me about that.
   - Is that typical of what you use X for?
   - What's another Google product that you use often? When was the last time you used Y? Tell me about that.
   - What are your impressions of the advertising that you see in product X or product Y? Can you give me an example? Tell me about that.
   - What would you do if you wanted to change or remove some of this information about you that Google has? *(If they don't know, they may ask you how to do it.)*

5. Wrap Up

1. Ask if the participant has any questions:
   - If the participant has asked any questions, now is the time to answer them. If you don't know the answer, make a note about it and tell them [redacted] will get back to them. You should be able to answer simple questions like "what is this study about", "when do you expect to be done", "can you email me a copy of the report" (yes!), etc. You should also be prepared to assist if the participant asks you how to delete/remove/edit categories.

2. Ask the participant what email address they want to use for receiving the gift card:
   - Tell the participant that you've reached the end of the interview. Then tell them that they'll be sent a $25 Amazon.com gift card for participating, by [redacted]. Ask them what email address they want us to use to send them the gift card, and write it down!
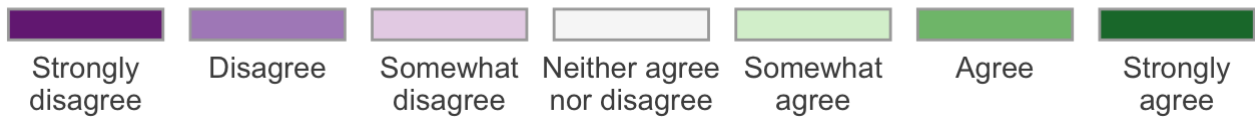
(Participant Name): Facebook Information

Below is the information you answered questions about when you filled out the survey. You answered three questions for each piece of information:

- [X] is relevant to who I am as a person
- [X] is an accurate description of my everyday activities
- It makes sense that [X] is associated with me

The colors in each table below represent the average of your responses to the three questions for each piece of information. The legend illustrates what the colors mean:

Legend

Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree

Life, Physical and Social Sciences	
Facebook access (OS): Mac OS X	
WiFi users	
Household income: top 10% of ZIP codes (US)	

Facebook access (OS): Mac Sierra	
Facebook access (mobile): smartphones and tablets	
Facebook access (mobile): Apple (iOS)devices	
Facebook access (browser): Chrome	
Education and Libraries	
Birthday in August	
Mobile network or device users	
Facebook access (mobile): tablets	
Recent mobile network or device change	
Potential mobile network or device change	
Multicultural Affinity: African American (US)	
Facebook access (mobile): all mobile devices	
Architecture and Engineering	

How your information compares with others

This page shows you how your Facebook information compares with the Facebook information about the other people who filled out our survey.

In total, 44 people from around the East Lansing area completed the survey and submitted their Facebook information.

Section 1

Here is the information that appeared in your file, **but nobody else's**:

Architecture and Engineering

Section 2

Here's the information that **was NOT in your file**, but was in in the Facebook files of **at least 13%** of the people who completed our survey:

Android: 360 degree media supported

Commuters

Engaged Shoppers

Facebook access (mobile): Android devices

Facebook access (network type): 4G

Facebook access (network type): WiFi

Facebook access: older devices and OS

Frequent Travelers

Gmail users

Household income: top 10%-25% of ZIP codes (US)

Likely engagement with US political content (liberal)

Parents (All)

US politics (moderate)

Uses a mobile device (25 months+)

Section 3

And finally, here's how similar your information was to the information in other people's Facebook files.

Everyone *had this information:*
Mobile network or device users, Potential mobile network or device change, Recent mobile network or device change, WiFi users

*You and **a few other people** had this information:*
Education and Libraries, Facebook access (OS): Mac OS X, Facebook access (OS): Mac Sierra, Life, Physical and Social Sciences
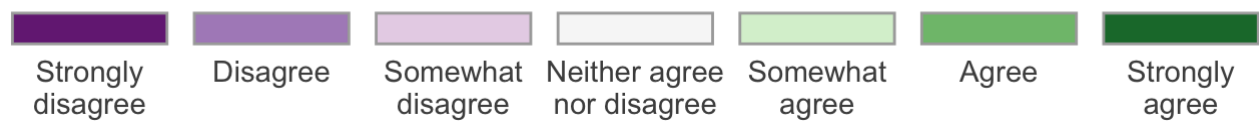
(Participant Name): Google Information

Below is the information you answered questions about when you filled out the survey. You answered three questions for each piece of information:

- [X] is relevant to who I am as a person
- [X] is an accurate description of my everyday activities
- It makes sense that [X] is associated with me

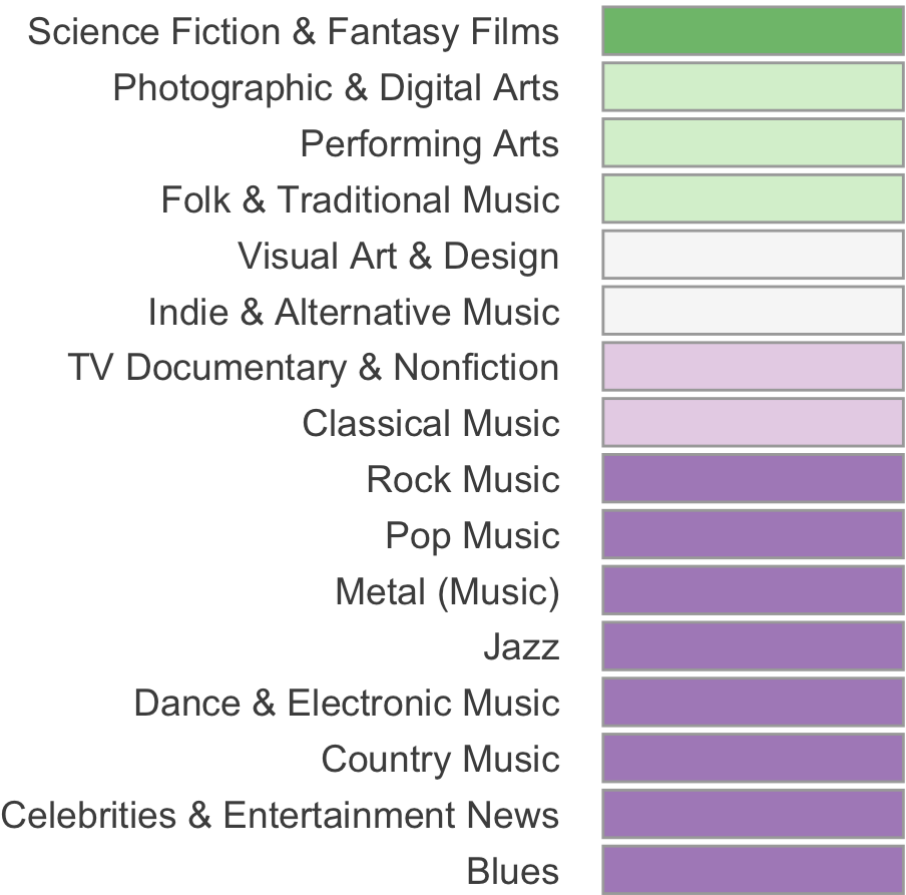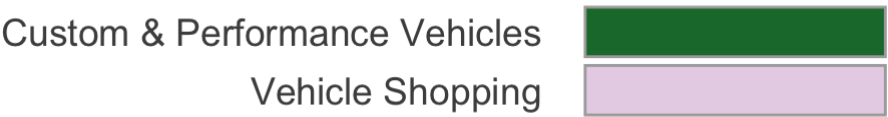The colors in each table below represent the average of your responses to the three questions for each piece of information. The legend illustrates what the colors mean:

Legend

Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree

Google has organized some of the information about you into a hierarchy. Where possible, this report presents those pieces of information grouped under the heading that Google has assigned. The first group, below, has no heading:
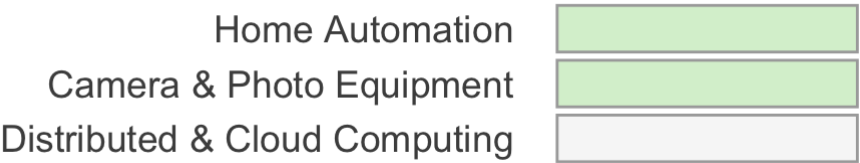
Female	
45-54 years old	
The Home Depot	
Etsy	
Williams-Sonoma	
L.L.Bean	
Target	
Office Depot	

Arts & Entertainment

Science Fiction & Fantasy Films

Photographic & Digital Arts

Performing Arts

Folk & Traditional Music

Visual Art & Design

Indie & Alternative Music

TV Documentary & Nonfiction

Classical Music

Rock Music

Pop Music

Metal (Music)

Jazz

Dance & Electronic Music

Country Music

Celebrities & Entertainment News

Blues

Autos & Vehicles

Custom & Performance Vehicles

Vehicle Shopping

Computers & Electronics

Home Automation

Camera & Photo Equipment

Distributed & Cloud Computing

Food & Drink

Gourmet & Specialty Foods

Restaurants

Fast Food

Cooking & Recipes

Hobbies & Leisure

Outdoors

Holidays & Seasonal Events

Jobs & Education

Colleges & Universities

Education

News

World News

News

Local News

Politics

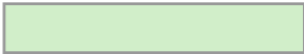# Online Communities

Social Networks

People & Society

Family & Relationships
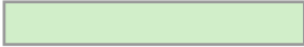
Pets & Animals

Dogs

Shopping

Shopping

Custom & Personalized Items

Sports

Baseball

Australian Football

Travel

Travel

How your information compares with others

This page shows you how your Google information compares with the Google information about the other people who filled out our survey.

In total, 54 people from around the East Lansing area completed the survey and submitted their Google information.

Section 1

Here is the information that appeared in your file, **but nobody else's:**

Custom & Personalized Items

Williams-Sonoma

Section 2

Here's the information that **was NOT in your file**, but was in in the Google files of **at least 50%** of the people who completed our survey:

American Football

Autos & Vehicles

Books & Literature

Business Services

Comics & Animation

Computers & Electronics

Coupons & Discount Offers

Credit Cards

Gifts & Special Event Items

Home & Garden

Home Improvement

Parenting

Price Comparisons

TV Game Shows

TV Reality Shows

TV Talk Shows

Section 3

And finally, here's how similar your information was to the information in other people's Google files.

Everyone *had this information:*
Shopping, Holidays & Seasonal Events, Performing Arts, Family & Relationships, Visual Art & Design, Cooking & Recipes, Blues, Gourmet & Specialty Foods, Politics, Travel

*You and **a few other people** had this information:*
Photographic & Digital Arts, Camera & Photo Equipment, TV Documentary & Nonfiction, World News, Social Networks, News, Baseball, Colleges & Universities, Custom & Performance Vehicles, Australian Football