Survey on Preference and Perceptions of Google Ads and Profiling

Dillon Shu
Carnegie Mellon University

Victor Waddell
Carnegie Mellon University

Danny Nguyen
Carnegie Mellon University

Haozhe Zhou
Carnegie Mellon University

Abstract

This study investigated perceptions of Google Ads profiles from the user's perspective. Participants completed a survey on their internet behavior and advertising preferences on Google. We found that users perceived their profiles as reasonably accurate (Self-reported data revealed high accuracy for demographics (age: 92.3%, gender: 90.00%, language: 87.5%). However, the reported accuracy of Ad categories varied (Education: 41.38% to Parenting: 63.64%). Reactions to profiling were diverse, with some valuing personalization and others stating privacy concerns. Despite such concerns, many users said they would continue using personalized ads. These findings suggest the need for balancing personalization and privacy, diverse anti-tracking tools, and increased transparency.

Keywords: Google Ads, user profiles, profiling, privacy, online advertising.

1 Introduction

The advent of digital advertising has significantly transformed how businesses communicate with their audiences, with Google emerging as a key player. Through their extensive collection and processing of user data, they are able to create detailed user profiles for personalized advertising. While Google's practices offer users perceived effectiveness and personalization, they also raise pressing concerns about the accuracy, privacy, and ethical implications of user profiling. Furthermore, there is an inadequate amount of research on Google's generated Ad profiles, especially from the perspective of the user. Motivated by the increasing focus on online privacy and data utilization, our research aims to investigate Google's user profiling practices, particularly how users perceive their Google advertising profiles derived from their usage data. This paper seeks to enrich the discourse on user profiling efficacy and dissect the immediate issues by focusing on Google's user profiling and how it is perceived by users themselves.

To delve into the accuracy and user perceptions of Google's profiling, we conducted a survey targeting users who navigate the internet while possessing a Google account with personalized ads enabled. This survey explored various topics, including the perceived accuracy of user profiles, the impact of these profiles on the user, and the broader ethical implications. We analyzed the data using a mix of quantitative and qualitative methods, ensuring a comprehensive understanding of the issue at hand.

Our findings revealed that the domain of digital advertising is multifaceted regarding balancing profiling accuracy, privacy, and upholding the interests of the user. While Google's profiling shows high accuracy in basic demographic categories, there are notable discrepancies in other areas, leading to mixed perceptions among users. Many participants expressed surprise at Google's data collection being greater or lesser than they initially perceived, though this did not necessarily translate into negative perceptions or behavioral changes. The nature of the user's survey responses towards digital profiling and privacy emphasizes the need for more tailored approaches in digital advertising practices.

2 Related Works

This section provides background on how existing research has defined user profiles. Afterward, we discuss what previous research has found about user profiling in practice, as well as the lack of literature specifically regarding Google's user profiling. We then highlight the potential privacy implications of advertising profiles before reviewing how previous research

has measured profile accuracy.

2.1 Defining User Profiles

Our team has conducted an analysis of various research works related to user personalization, with a specific focus on user profiling for advertisers. During this analysis, we have identified numerous definitions associated with online profiling and targeting such as user profiles that need to be firmly established. Some studies have defined a user profile as a snapshot and set of demographics and current interests that can describe and infer further data and interests regarding a user, which is considered static [1], [2], [3]. Other works describe a user profile as a process of obtaining information about the user and their interests in order to predict their future behavior [1], [4], [3]. For our research project, we have decided to align our definition of a user profile with various existing works that outlined the difference between static and dynamic user profiling and provided many examples of existing literature definitions [1], [3]. We will define it as a collection of user interests and preferences tracked over time that can change according to the user's behaviors and online activity [1], [3], [5]. We believe that this definition is more suitable for our research, as it allows us to account for the dynamic nature of user behavior and preferences in the online world.

2.2 Privacy, profiling, algorithms in practice

There are a variety of forms that user profiling can take. While the main algorithms that Google, Facebook, etc. use are a blackbox to us, the literature shows a need to collect sensitive information in explicit and implicit ways. Such information can be traced back to the individuals. One paper details a method of gathering sensitive information from simulated users using Google's Ad analytics [6]. Another proposes another method of using text analysis to determine the gender, age, and education of a user [7]. Together, these demonstrate how malicious methods can use data collected to identify users. The traditional techniques for generating and representing user profiles are based on keywords, semantic networks, or concept networks [3] [8]. The data used is often collected through a user's search history and activity, but it can also be provided directly by users. The core ideas behind user profiles are based on intentionality: behavioral modeling, interest modeling, and intentional modeling [1]. Even if the end goal is always targeted advertising, both the data collected and the inferences created differ heavily depending on the techniques used to develop the model and the motivation behind it.

2.3 Gap in Existing Literature

While existing literature has extensively covered approaches and methods in which user profiling can occur, limited information is available on how these profiles are perceived [5].

Furthermore, profiling based on demographics has also been limited in works while often found to lead to incorrect and biased profiles and inferences [5]. No recent research has covered the efficacy of either Facebook or Google, with the most recent paper covering Google's personal profiles being the aforementioned paper by Conti. et al [6]. Literature regarding Facebook is instead focused on the placement of ads [9] or political advertising and social effects [9] [10] [11], some of which are only tangentially related to user advertising profiles in the first place. Our study aims to provide a perspective of the efficiency of user profiling from primarily a consumer perspective.

2.4 Perception of Privacy

Our study aims to understand how users perceive the extent of Google's tracking by viewing their profiles. Previous literature has noted how users have inconsistent behavior regarding their privacy due to the complexity of online situations [12]. We limit users to engaging with only the case of their Google profile. The perceptions of users matter in determining what future actions or policies need to be taken regarding Google. Additionally, insights into the perceptions of Google's tracking can be used to infer perceptions of other platforms.

2.5 Measurement Methodology

While many works have investigated the accuracy of profiling algorithms, such as [13] [14] [15], they are mainly discussing from an algorithmic perspective instead of existing deployed commercial profiling techniques. To understand the methodology suitable for our interest in measuring user profiling accuracy from the user's perspective, we examined several related works and their study procedures. Neumann et al. [14] investigated the precision of age and gender inference in addition to audience-interest segments. The first field study showed that digital advertising campaigns using both data brokers and demand-side platforms (DSP) had an average audience targeting accuracy of 59 percent, while the second one showed that the accuracy of using only data brokers in identifying specific demographic attributes, such as gender and age, ranged widely with an average accuracy often below or near random selection. This result highlighted that although DSPs might optimize campaigns and potentially enhance targeting, the fundamental accuracy of the underlying data from data brokers can still be flawed. While the studies conducted by Neumann et al. are impactful, their studies rely on collaboration with DSPs and access to global data management platforms (DMP) and thus require significant effort to replicate such studies or use similar study designs. On the other hand, Tschantz et al. [15] presented their findings on Google Ads demographic inferences using 501 survey participants on Amazon Mechanical Turk (mTurk) by simply asking the participants' ground truth demographics and a copy of their

Google's Ad Settings to get the inference on the demographic information. Based on the survey, the authors showed a significantly higher accuracy when using Google's website or services than websites and services from other third parties. This study provides a highly practical methodology for obtaining matched ground truth information about real-world users and inferred results without privileged access to any type of advertising platform.

3 Methodology

We ran into several challenges when designing our study. First, access to what profiles are being inferred about each user can be challenging. Second, it can be tricky to ask users these questions as some of them might be sensitive to participants, so they don't want to be released. Our study design focused on answering our research questions while also considering the real-world intricacies of obtaining accurate responses on profiling.

RQ 1 What is the accuracy and coverage of personal profiling from Google Ads in inferring user attributes and targeting ads?

RQ 2 What Internet behaviors are more or less likely to lead to accurate profiling?

RQ 3 How do users feel about the level of profile inference made by Google Ads and what decision will they make towards Google Ads settings?

RQ 4 How does showing users actual examples of Google's data collection practices affect their beliefs about the extent of data gathered on them?

3.1 Procedure

Our study consisted of a prescreen survey followed by a main survey. The prescreen was distributed through Prolific and was administered using Qualtrics. Following pilot testing, the prescreen survey was designed take approximately 2-3 minutes for each participant, while the main survey would take approximately 10 minutes to complete. For the prescreening survey, a compensation of $0.30 per participant is set, with the main survey offering $1.50, subject to increase once we conduct and interpret the pilot survey outcomes. Our budget includes $75 for 150 prescreen participants, $75 for 60 main survey participants, and an additional $50 for Prolific platform fees. These figures will be finalized after a small sample pilot study is conducted on Prolific and we have an estimate of the conversion rate.
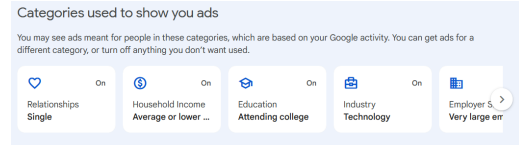

Figure 1: Example of demographic information collected shown to participants

3.2 Participant Demographics and Recruitment

Our study recruited participants using Prolific. Participants were eligible for the prescreen if they are 18 years of age or older and reside in the United States. To have been eligible for the main survey, participants must have taken the prescreen in addition to having personalized advertising enabled, which was determined based on their responses to the prescreen survey. Qualified participants were then invited to participate in the main survey immediately after they completed the prescreen, though they did not have to continue. Our study recruited 294 prescreen participants, of which 41 continued to the main survey. Although the sample is not representative, the prescreen survey was distributed using Prolific's gender-balance feature. The main survey maintained this even distribution somewhat, with 23 male, 17 female, and 1 non-binary participants.

3.3 Survey Design

Before participants could see the primary survey, they were asked to fill in some screening questions to ensure their Google Ads setting was enabled. While perceptions from people who already disabled the feature are insightful for our community, it would be challenging to measure the accuracy of profiling, which was aprimary focus of our study.

At the beginning of the main survey, we first asked participants to estimate the age of their Google account, as there is no method that can correctly ascertain the age of a Google account. Participants were then asked about their daily time spent using Google search and YouTube, as well as browsing other websites. We hypothesized that participants with older accounts and frequent Google searches and YouTube usage would have more accurate profiles. In addition, we asked participants about their habits of using anti-tracker or ad-blocker tools. We hypothesized that enabling these tools would weaken the profile accuracy.

Next, we asked participants a series of questions to determine the accuracy of Google's online profiling. These questions measured both subjective and objective correctness. First, participants were given a link that shows Google's assumed demographic information about them, shown in Figure 1, and asked how many categories were accurate and inaccurate out of the categories they had. This set of questions
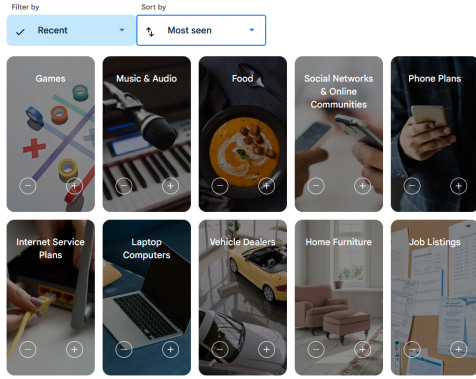
Figure 2: Example of ad categories information in Google's Ad Center shown to participants
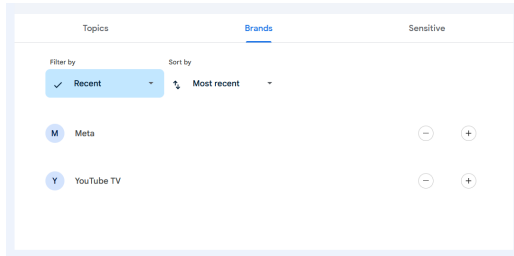

Figure 3: Example of brand information in Google's Ad Center shown to participants

provided an objective measurement of the accuracy of the participant's online profile. Afterward, participants were given a link that shows the topics that Google assumes they are interested in, as well as the brands whose advertisements they have seen, as shown in Figures 2 and 3. Participants were then asked about how accurately they think these topics describe their interests, providing a subjective opinion of accuracy. Additionally, participants were asked how many of the brands whose ads they have recently been shown relate to something they are interested in. These questions provided a subjective measurement of accuracy, collecting the participant's opinion about the accuracy of their profile. Participants were also asked how many of the advertisers they are most commonly shown are companies whose products or services they have actually used, which serves as an objective measurement.

Finally, we asked the participants to reflect on the information that they have learned. We wanted to how they felt about the extent of Google's tracking. We asked them to rate their knowledge and perception on a Likert scale and offer an open-ended response for specific surprises. Lastly, we wanted to determine whether or not this information changes their behaviors. Participants were shown pre-defined reasons to either continue using personalized ads or disable personalized ads along with an other option for free responses. This was followed by a final open-ended response for any other rea-

sons they have for their decision. These allowed our study to measure the subjective perception that users have of Google's practices.

3.4 Data Analysis

We performed both quantitative and qualitative analyses on our results, with a focus on the accuracy and user perception of Google's advertising profiles. For our qualitative analysis, we performed double-coding on all open-ended responses using a codebook drafted by the team, with the coders reaching an agreement rate of at least 85% for each question. As a result of our small sample size, we used a logistic model test to determine the statistical significance of our quantitative results that require a comparison across groups. Additionally, we wanted to determine the overall accuracy, so we wperformed t-tests for the accuracy of demographics and the relevance of ads against a standard null hypothesis. If there are differences between our observed mean responses for our sample and the standard value, then there is evidence to suggest that people perceive their profiles to be accurate/inaccurate and the ads are often relevant/not relevant to them. The results of these tests for each question reveal which specific aspects of Google's tracking are the most significant.

4 Results

The primary focus of this study was to determine the accuracy of Google's profiles, where we found that while the accuracy varies, it was generally more accurate in correctly identifying demographics information. Our study also aimed to determine user perception of their Google profiles. Here, we found mixed responses; some were surprised at how inaccurate profiles were, while others felt the exact opposite. We also found that users' behavior was not affected by going through their own profiles.

4.1 User Profile Accuracy

Using the participant's self-reported accuracy of their user profile, we determined Google's overall accuracy. We summarized the accuracy of each demographic and ad category in Figure 4. We found Google Ad profiles for basic demographic categories including Age, Gender, and Language achieved high average accuracy (92.3%, 90.00%, 87.5% respectively). On the other hand, the accuracy of prediction for other ad categories produced varied, with the highest being Parenting (63.64%) and the lowest being Education (41.38%).

We further examined the statistical significance of the profile accuracy by performing t-tests against random guesses. Our study found that Google's accuracy for demographic data was statistically significant, whereas the accuracy for the different ad categories Google defined varied more. The ad categories of Education, Employer Size, and Homeowner
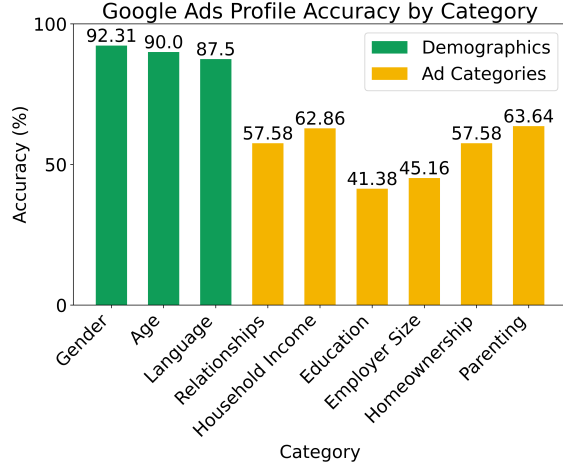
Figure 4: Average Accuracy of User Profile

that Google's accuracy for these categories is significantly better than randomly guessing. Then, we created a logistic model to compare the differences between categories and used Education as the baseline variable. We found that the other variables did not contribute significantly to predicting accuracy with p-values of 0.56, 0.56, 0.15, 0.77, 0.25, and 0.38 for Employer Size, Homeowner, Household Income, Industry, Parenting, and Relationship Status respectively. All this suggests is that the ad categories are not significantly different from each other.

Figure 4 highlights what the actual accuracy for each category is in a user's Google profile. We observe that the demographic categories have a high accuracy of around 90% whereas the ad categories vary between around 40% and 60%.

4.2 Profile Accuracy and Online Behaviors

This section will explore the relationship between the profile accuracy of participants' online behaviors. In our survey, we collected the following five factors, which we hypothesized to be relevant to profile accuracy: Account Age, Google Search, YouTube, Visiting Other Websites, and Anti Trackers.

To perform analysis, we converted the levels of each factor into numeric values. We also computed the mean of the profile accuracy across all demographic and ad categories. We then examine the Pearson correlation between each of the five factors and the average profile accuracy.

Factors	Correlation	95% CI
Account Age	.24	[-.07 , .51]
Google Search	.03	[-.27 , .34]
YouTube	-.07	[-.37 , .26]
Other Websites	.13	[-.18 , .42]
Anti Tracker	-.15	[-.44 , .16]

Table 1: Correlation of Online Behavior Factors with Profile Accuracy

The results of the correlation analysis, including the correlation coefficients and 95% confidence intervals, are shown in Table 4.2. As shown, a potential slight positive correlation might exist between account age and profile accuracy ($r = .24$, $CI = [-.07, .51]$). This suggests that, as expected, as accounts age increases their accuracy in profiling might increase. On the other hand, the use of anti-trackers appears to be slightly negatively correlated with profile accuracy ($r = -.15$, $CI = [-.44, .16]$), hinting that these tools might lower the accuracy of profiling.

However, these correlations do not show statistical significance $p > 0.05$. The confidence intervals for most factors are wide, indicating a high degree of uncertainty in these estimates. While the above preliminary results revealed the potential of certain trends, to draw more concrete conclusions, further studies with increased sample size are needed
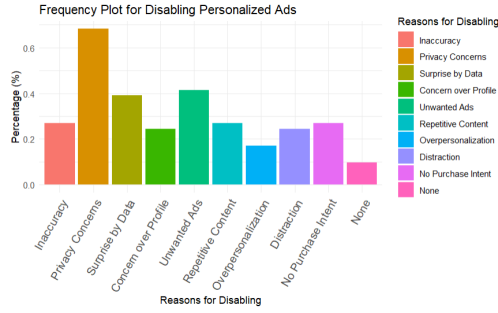
were not significant, whereas Household Income, Industry, Parenting, and Relationship Status were. We did not see significant differences between the demographic categories from each other nor were there significant differences between the ad categories.

For our t-tests with the demographic data, we found that their p-values were 6.18×10^{-12}, 3.48×10^{-10}, and 1.65×10^{-8} for gender, age, and language respectively. The t-tests were performed with a baseline accuracy of 50%. Since all of these p-values are less than a significance level of 0.05, we have evidence to suggest that Google's accuracy for demographic data is higher than a coin flip guess. We created a logistic model using the demographic data to predict accuracy to determine if there were differences between the categories. Using Age as a baseline, we found that Gender and Language did not significantly improve the model with p-values of 0.69 and 0.72 respectively. Since these p-values are greater than a significance level of 0.05, this suggests that there is not a significant difference between the demographic categories in terms of accuracy.

For the t-tests of ad categories, we found their p-values were 0.009, 0.001, 0.089, 0.0009, 0.202, 0.392, 4.33×10^{-6} for Relationship Status, Household Income, Education, Industry, Employer Size, Homeowner, and Parenting respectively. The t-tests were performed using a baseline value of $\frac{1}{M}$, where M is the number of options Google has for that category. Since Education, Employer Size, and Homeowner had p-values greater than a significance level of 0.05, this suggests that Google's accuracy for these categories is not significantly better than randomly guessing the category. Since Relationship Status, Household Income, Industry, and Parenting had a p-value less than a significance level of 0.05, this suggests

Figure 5: Reasons to Disable Personalized Ads

Figure 7: Beliefs about Google Before and After Survey

Figure 6: Reasons to Keep Personalized Ads

at over 60%. However, we saw in Figure 6 that the second most frequent response was that users would keep ads due to their relevance to their interests at over 50%. Some of the open-ended responses showed that some users cited laziness as a reason to keep ads. Others were not concerned due to inaccuracies in their own profile.
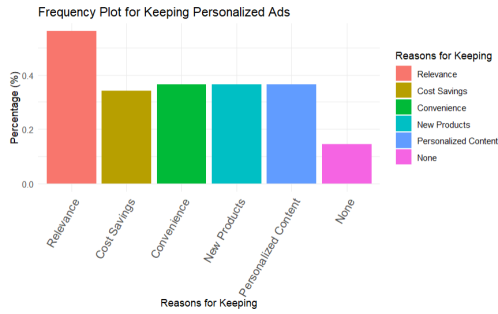
to enhance the statistical power,

4.3 Perceptions and Preferences Regarding Personalized Advertising

When users learned about this information, it did not encourage a significant change from keeping the same settings. Some users now believe that Google is not very invasive due to their inaccuracies, whereas some users believe it is very accurate. These results reinforce the idea of the Privacy Paradox, where users claim to care about privacy and take little action to protect it.

We asked participants how likely they were to disable personalized ads after reviewing their profile on a Likert scale. We performed a t-test with a baseline value of 3 indicating that they are neither likely nor unlikely to disable personalized ads. We found a p-value of 0.43, which is greater than a significance level of 0.05 and suggests that many users are neither likely nor likely to disable personalized ads. If this is the case, then the average user is unlikely to change from the defaults even if they have concerns.

Additionally, we asked participants to select pre-defined reasons why they would choose to disable personalized ads or continue enabling them. Participants also had the option to submit an open-ended response to this question. In Figure 5, we found that the most frequent response to these questions was that users would disable ads out of privacy concerns
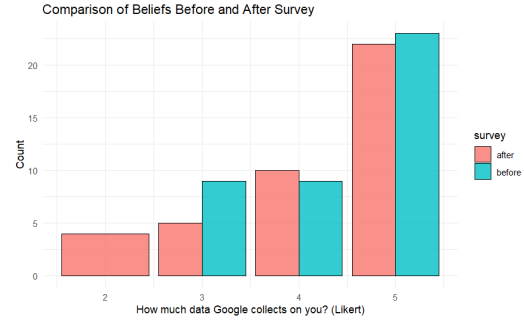
4.4 User Reflections

We found that many users were surprised to some extent about the accuracy of the information Google collected about them. Some were surprised it was inaccurate and others were surprised that it was accurate. People generally did not seem to change their beliefs, and many people still believe that Google collects a lot of information about them.

We asked participants how surprised they were at the information displayed in their Google profile on a Likert Scale. We performed a t-test using a baseline value of 3 to indicate that they were moderately surprised. We found a p-value of 0.26, which is greater than a significance level of 0.05. This suggests that the participants were not too surprised nor were they too unsurprised at their profile. Individually, some users were surprised at how inaccurate their profile was, and others were surprised at how accurate it was.

We asked participants before and after taking the survey how much data they believed Google collected about them on a Likert Scale. We performed a t-test to compare if there was a difference between their beliefs before and after the survey. We found this to have a p-value of 0.55, which is greater than a significance level of 0.05. This suggests that the average person did not change their beliefs about Google's data collection before and after seeing their profile. In Figure 7, some of the responses after the survey actually went down. This is likely due to individual inaccuracies in their profiles. Overall, most participants before and after the survey believe Google collects a lot of data on them.

5 Discussion

5.1 Privacy Paradox

The "Privacy Paradox" refers to the contradictory behavior of users who express concern about their privacy but continue to share personal information online [16]. This paradox is evident in our participant's responses to our questionnaire. While many users are aware of privacy issues and may feel uncomfortable with the extent of data collection, they often do not take significant steps to protect their privacy, such as using anti-tracking tools or adjusting privacy settings. This observation suggested the complexity of users' perceptions and behaviors towards making privacy decisions.

5.2 The Misconception: Blocking is Good

Research in the privacy field has been centered around providing end-users with the weapons to fight against tracking and targeting. Without denying the significance of such research and technology, it shall not be regarded as the only correct way to deal with tracking and targeting. We observed many participants outweigh the benefits of personalized online experience more than privacy concerns. This phenomenon highlights the need to reconsider ad-blocking tools to meet the requirement of various user's perceptions and wishes. It also suggests the potential for developing more sophisticated tools that balance privacy protection and personalized content with more fine-grained protection mechanisms.

5.3 Limitations and Future Works

In this project, our analysis is largely limited by the sample size and subsequently the statistical power. While the current sample size has yielded several important observations, as we described in Section 4, a larger sample would enable more detailed analysis, such as deriving conclusive results on the impact of online behaviors and brand categories. In addition, our participants are based on Prolific, which may or may not represent a generalized image of Google Ads users. Future studies might consider a larger population with more diverse samples, such as including various professional backgrounds and living locations.

6 Conclusion

In this project, we investigated the user's perceived accuracy and opinion regarding Google Ads profiles. Our study showed that Google Ads profiles overall produced reasonable accuracy, particularly for demographic data, indicating their effectiveness in targeted advertising. Second, we observed a diverse range of user perceptions and reactions to profiling, suggesting a balance between its benefits for personalization and concerns over privacy and ethics. These findings emphasize the necessity for careful consideration of the varied opinions and preferences of users in implementing both targeted ads and anti-tracking tools.

References

[1] C. I. Eke, A. A. Norman, L. Shuib, and H. F. Nweke, "A survey of user profiling: State-of-the-art, challenges, and solutions," *IEEE Access*, vol. 7, pp. 144 907–144 924, 2019. DOI: 10.1109/ACCESS.2019.2944243.

[2] C. Castelluccia, M.-A. Kaafar, and M.-D. Tran, "Betrayed by your ads!" In *Privacy Enhancing Technologies*, S. Fischer-Hübner and M. Wright, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 1–17, ISBN: 978-3-642-31680-7.

[3] S. Gauch, M. Speretta, A. Chandramouli, and A. Micarelli, "User profiles for personalized information access," in *The Adaptive Web: Methods and Strategies of Web Personalization*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 54–89, ISBN: 978-3-540-72079-9. DOI: 10.1007/978-3-540-72079-9_2. [Online]. Available: https://doi.org/10.1007/978-3-540-72079-9_2.

[4] S. Kanoje, S. Girase, and D. Mukhopadhyay, "User profiling trends, techniques and applications," *CoRR*, vol. abs/1503.07474, 2015. arXiv: 1503.07474. [Online]. Available: http://arxiv.org/abs/1503.07474.

[5] M. Degeling and J. Nierhoff, "Tracking and tricking a profiler: Automated measuring and influencing of bluekai's interest profiling," in *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, ser. WPES'18, Toronto, Canada: Association for Computing Machinery, 2018, pp. 1–13, ISBN: 9781450359894. DOI: 10.1145/3267323.3268955. [Online]. Available: https://doi.org/10.1145/3267323.3268955.

[6] M. Conti, V. Cozza, M. Petrocchi, and A. Spognardi, "Trap: Using targeted ads to unveil google personal profiles," in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2015, pp. 1–6. DOI: 10.1109/WIFS.2015.7368607.

[7] M. Trusov, L. Ma, and Z. Jamal, "Crumbs of the cookie: User profiling in customer-base analysis and behavioral targeting," *Marketing Science*, vol. 35, no. 3, pp. 405–426, May 2016, ISSN: 1526-548X. DOI: 10.1287/mksc.2015.0956. [Online]. Available: https://doi.org/10.1287/mksc.2015.0956.

[8] J. Hu, F. Jin, G. Zhang, J. Wang, and Y. Yang, "A user profile modeling method based on word2vec," in *2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 2017, pp. 410–414. DOI: 10.1109/QRS-C.2017.74.

[9] S. K. Manon Metz and S. Lecheler, "Personalization of politics on facebook: Examining the content and effects of professional, emotional and private self-personalization," *Information, Communication & Society*, vol. 23, no. 10, pp. 1481–1498, 2020. DOI: 10.1080/1369118X.2019.1581244. eprint: https://doi.org/10.1080/1369118X.2019.1581244. [Online]. Available: https://doi.org/10.1080/1369118X.2019.1581244.

[10] D. S. Raffael Heiss and J. Matthes, "What drives interaction in political actors' facebook posts? profile and content predictors of user engagement and political actors' reactions," *Information, Communication & Society*, vol. 22, no. 10, pp. 1497–1513, 2019. DOI: 10.1080/1369118X.2018.1445273. eprint: https://doi.org/10.1080/1369118X.2018.1445273. [Online]. Available: https://doi.org/10.1080/1369118X.2018.1445273.

[11] K. Hall, "Public penitence: Facebook and the performance of apology," *Social Media + Society*, vol. 6, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:216426286.

[12] A. Acquisti, L. Brandimarte, and G. Loewenstein, "Privacy and human behavior in the age of information," *Science*, vol. 347, no. 6221, pp. 509–514, 2015. DOI: 10.1126/science.aaa1465. eprint: https://www.science.org/doi/pdf/10.1126/science.aaa1465. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.aaa1465.

[13] T. Theodoridis, S. Papadopoulos, and Y. Kompatsiaris, "Assessing the reliability of facebook user profiling," in *Proceedings of the 24th International Conference on World Wide Web*, ser. WWW '15 Companion, Florence, Italy: Association for Computing Machinery, 2015, pp. 129–130, ISBN: 9781450334730. DOI: 10.1145/2740908.2742728. [Online]. Available: https://doi.org/10.1145/2740908.2742728.

[14] N. Neumann, C. Tucker, and T. Whitfield, "Frontiers: How effective is third-party consumer profiling? evidence from field studies," *Marketing Science*, vol. 38, pp. 918–926, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:210040385.

[15] M. C. Tschantz, S. Egelman, J. Choi, N. Weaver, and G. Friedland, "The accuracy of the demographic inferences shown on google's ad settings," *CoRR*, vol. abs/1808.07549, 2018. arXiv: 1808.07549. [Online]. Available: http://arxiv.org/abs/1808.07549.

[16] S. Kokolakis, "Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon," *Computers & security*, vol. 64, pp. 122–134, 2017.

A   Appendix

A.1   Consent Form and Screening Questions (Prescreen Survey)

A.1.1   Consent Form

This survey is part of a course project conducted under Danny Nguyen at Carnegie Mellon University and is funded by Carnegie Mellon University.

Summary and Purpose

You will be asked to complete a short survey about your perceptions on your Google advertising profile. We ask that you please answer the questions truthfully.

Procedure

You will be provided a link to an online survey using Qualtrics. We expect this survey to take no more than 2 minutes. When filling out the survey, do NOT disclose any private or personally-identifiable information about yourself or anyone else. Following the completion of the survey, you will be given a link to click as proof of completion of the survey within Prolific. Please note that this is necessary for payment.

Participant Requirements

Participation in this study is limited to individuals age 18 and older who live in the United States.

Risks

The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or during other online activities.

Benefits

There may be no personal benefit from your participation in the study but the knowledge received may be of value to humanity.

Compensation & Costs

You will be compensated $0.30 USD following the completion of this survey. No payment will be given if you do not complete the survey. There will be no cost to you if you

participate in this study

Future Use of Information

In the future, once we have removed all identifiable information from your data (information or bio-specimens), we may use the data for our future research studies, or we may distribute the data to other researchers for their research studies. We would do this without getting additional informed consent from you (or your legally authorized representative). Sharing of data with other researchers will only be done in such a manner that you will not be identified.

Confidentiality

The data captured for the research does not include any personally identifiable information. Your IP address will not be captured. When answering questions during this survey, DO NOT disclose any private or personally-identifiable information about yourself or anyone else.

Right to Ask Questions & Contact Information

If you have any questions about this study, you should feel free to ask them by contacting the Principal Investigator at danguyen@andrew.cmu.edu. If you have questions later, desire additional information, or wish to withdraw your participation please contact the Principal Investigator by e-mail in accordance with the contact information listed above. If you have questions pertaining to your rights as a research participant; or to report concerns to this study, you should contact Lujo Bauer. Email: lbauer@andrew.cmu.edu.

Voluntary Participation

Your participation in this research is voluntary. You may discontinue participation at any time during the research activity. You may print a copy of this consent form for your records.

A.1.2 Screening Questions

Screening Q1. I am age 18 or older.

- Yes

- No

Screening Q2. I have read and understand the information above.

- Yes

- No

Screening Q3. I want to participate in this research and continue with the survey

- Yes

- No

A.2 Prescreen Survey

Q1. What is your age?

- 18-35

- 36-53

- 54+

Q2. Which of the following best describes your gender?

- Male

- Female

- Non-binary/Other

- Prefer to self describe

Q3. Do you have a degree or have you been employed in a technical field (such as engineering, computer science, or a similar field)?

- Yes

- No

Please answer the following questions using your primary Google account or the Google account you use the most frequently.

Q5. To the best of your knowledge, what is the age of your Google Account?

- Less than 3 months old

- Between 3 months and 1 year

- Between 1 to 3 years

- Over 3 years

Q6. Which of the following best matches what you see after clicking the following link? [link redirected to Google's personalized ad settings page]

- Screenshot of page with personalized ads enabled

- Screenshot of page without personalized ads enabled

- My page is not similar to either of the above.

Q7. Prior to taking this survey, did you know about personalized ads on Google?

- Yes

- No

Q8. If you qualify, would you like to participate in a short 5-minute survey about Google's personalized ads for $1 compensation? Note: You will be taking this survey immediately after the completion of this survey, and it should be completed in one sitting.

- Yes

- No

A.3 Consent Form and Screening Questions (Main Survey)

This survey is part of a course project conducted under Danny Nguyen at Carnegie Mellon University and is funded by Carnegie Mellon University.

A.3.1 Consent Form

Summary and Purpose
You will be asked to complete a short survey about your perceptions on your Google advertising profile. We ask that you please answer the questions truthfully.

Procedure
You will be provided a link to an online survey using Qualtrics. We expect this survey to take no more than 12 minutes. When filling out the survey, do NOT disclose any private or personally-identifiable information about yourself or anyone else. Following the completion of the survey, you will be given a link to click as proof of completion of the survey within Prolific. Please note that this is necessary for payment.

Participant Requirements
Participation in this study is limited to individuals age 18 and older who live in the United States.

Risks
The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or during other online activities.

Benefits
There may be no personal benefit from your participation in the study but the knowledge received may be of value to humanity.

Compensation & Costs
You will be compensated $1.50 USD following the completion of this survey. No payment will be given if you do not complete the survey. There will be no cost to you if you participate in this study.

Future Use of Information
In the future, once we have removed all identifiable information from your data (information or bio-specimens), we may use the data for our future research studies, or we may distribute the data to other researchers for their research studies. We would do this without getting additional informed consent from you (or your legally authorized representative). Sharing of data with other researchers will only be done in such a manner that you will not be identified.

Confidentiality
The data captured for the research does not include any personally identifiable information. Your IP address will not be captured. When answering questions during this survey, DO NOT disclose any private or personally-identifiable information about yourself or anyone else.

Right to Ask Questions & Contact Information
If you have any questions about this study, you should feel free to ask them by contacting the Principal Investigator at danguyen@andrew.cmu.edu. If you have questions later, desire additional information, or wish to withdraw your participation please contact the Principal Investigator by e-mail in accordance with the contact information listed above. If you have questions pertaining to your rights as a research participant; or to report concerns to this study, you should contact Lujo Bauer. Email: lbauer@andrew.cmu.edu.

Voluntary Participation
Your participation in this research is voluntary. You may discontinue participation at any time during the research activity. You may print a copy of this consent form for your records.

A.3.2 Screening Questions

Screening Q1. I am age 18 or older.

- Yes

- No

Screening Q2. I have read and understand the information above.

- Yes

- No

Screening Q3. I want to participate in this research and continue with the survey

- Yes

- No

A.4 Main Survey

Survey Instructions: Before you begin, please ensure you are logged into the Google account you use most frequently that has personalized advertising on. This survey will refer to the settings and information associated with your primary Google account.

Q1.

To the best of your knowledge, what would you estimate the age of your Google Account?< 6 months 1 year 1 3 years 3 5 years 5+ years

Q2. How often would you say you browse the Internet for each of the following on an average day?

-

-

-

•

Q3. Do you use any ad block or anti-tracking tools when browsing the Internet?

1. YouTube

    - Do not use service
    - Less than 1 hour
    - 1 to 3 hours
    - 3 to 5 hours
    - More than 5 hours

2. Google Search

    - Do not use service
    - Less than 1 hour
    - 1 to 3 hours
    - 3 to 5 hours
    - More than 5 hours

3. Other websites

    - Do not use service
    - Less than 1 hour
    - 1 to 3 hours
    - 3 to 5 hours
    - More than 5 hours

Q4. How much did you think Google collected about you for advertising purposes?

- 1 - None
- 2
- 3 - Some
- 4
- 5 - A great deal

Survey Instructions: Please open and refer to the information found by clicking this link when answering the following questions

Q5. Under the section titled Your Google Account Info, how many options do you see? For example, in the screenshot below 3 options are displayed.

- 0
- 1
- 2
- 3

Q6. Which of the following details do you think Google has gotten right?

1. Gender

    - Accurate
    - Inaccurate
    - Not Enough Info
    - Not Shown

2. Age

    - Accurate
    - Inaccurate
    - Not Enough Info
    - Not Shown

3. Language

    - Accurate
    - Inaccurate
    - Not Enough Info
    - Not Shown

Survey Instructions: Please go to the section labeled Categories used to show you ads. Here, you might see categories such as 'Relationship', 'Household Income', etc.

Q7. Count the total number of categories listed. If there are more categories off the screen, use the arrow key to scroll through and continue counting. How many options do you see?

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7

Q8. Which of the following details do you think Google has gotten right?

1. Relationships

    - Accurate
    - Inaccurate
    - Not Shown

2. Household Income

   - Accurate
   - Inaccurate
   - Not Shown

3. Education

   - Accurate
   - Inaccurate
   - Not Shown

4. Industry

   - Accurate
   - Inaccurate
   - Not Shown

5. Employer Size

   - Accurate
   - Inaccurate
   - Not Shown

6. Houseownership

   - Accurate
   - Inaccurate
   - Not Shown

7. Parenting

   - Accurate
   - Inaccurate
   - Not Shown

Q9. Overall, how accurate do you think the information is about you? What information was incorrect? *(free response question)*

Q10. Under the section titled "Activity used to personalize ads", which of the following options are enabled/on?

1. Web App Activity

   - Currently Enabled/On
   - Currently Disabled/Off

2. YouTube History

   - Currently Enabled/On
   - Currently Disabled/Off

3. Areas where you've used Google

   - Currently Enabled/On
   - Currently Disabled/Off

Survey Instructions: Please click on this link (link to Google Ad Center Personalized Topics):
Set "Filter By" to Recent. Then set "Sort" by Most Seen. Observe the topics shown on your page. It should look something like this... (screenshot of linked page) **Q11.** How accurate would you say these topics are in describing your interests?

   - 1 - Not accurate

   - 2

   - 3 - Neutral

   - 4

   - 5 - Very accurate

Q12. How many topics do you see? *(free response question)*
Q13. How many of them are you interested in? *(free response question)*

Survey Instructions: On the same page, please select the Brands tab. Set "Filter By" to Recent. Then set "Sort" by Most Seen. Your page should look something like this. (screenshot of previously linked page with filtering now set to "Recent")

Q14. Scroll through the brands listed. How many of the brands listed would you consider to be relevant to your interests?

   - 1 - None

   - 2

   - 3 - Some

   - 4

   - 5 - Almost all

Q15. How many brands do you see on your page? *(free response question)*

Q14.1. Of the number of brands listed, how many of them interest you?*(displayed if participant responded with <10 for Q15)*

   - 0

   - 1

   - 2

   - 3

   - 4

   - 5

   - 6

   - 7

   - 8

- 9

- 10

Q14.4. Of the top 10 brands listed, how many of them interest you?*(displayed if participant responded with >=10 for Q15)*

- 0

- 1

- 2

- 3

- 4

- 5

- 6

- 7

- 8

- 9

- 10

Q15.1. Of the number of brands listed, how many have you previously interacted with or plan to interact with? This can including buying products if they sell them or creating an account if it is a web service. *(displayed if participant responded with <10 for Q15)*

- 0

- 1

- 2

- 3

- 4

- 5

- 6

- 7

- 8

- 9

Q15.4. Of the top 10 brands listed, how many of them interest you?*(displayed if participant responded with >=10 for Q15)*

- 0

- 1

- 2

- 3

- 4

- 5

- 6

- 7

- 8

- 9

- 10

Q16. How much do you now think Google collected about you for advertising purposes?

- 1 - None

- 2

- 3 - Some

- 4

- 5 - A great deal

Q17. To what extent were you surprised by the information you've reviewed about your generated user profile today?

- 1 - Not at all surprised

- 2

- 3 - Somewhat surprised

- 4

- 5 - Extremely surprised

Q18. What surprised you the most about the information you've reviewed about your generated user profile today? *(free response question)*

Q19. Google allows you to disable personalized advertising. You can do that with the following link (Link to Google Ad Center page for disabling personalized ads). You can turn off the types of tracking with the link and note that the disable personalized ads is located at the bottom of the page.

After going through your advertising profile, how likely are you to disable personalized advertising?

- 1 - Extremely unlikely

- 2

- 3 - Neutral

- 4

- 5 - Extremely likely

Q20. Why would you disable personalized advertising? (Select all that apply)

- Inaccuracy: The ads and brands do not accurately reflect my interests.

- Privacy Concerns: I'm concerned about how my data is being collected and used.

- Surprised by Data: I was surprised by the extent of information collected about me.

- Concern Over Profile: The information collected about me is concerning.

- Unwanted Advertising: I prefer not to see advertising while browsing.

- Repetitive Content: The same ads and brands are shown too frequently.

- Overpersonalization: The ads feel too tailored and intrusive.

- Distraction: Personalized ads disrupt my online experience.

- No Purchase Intent: I don't intend to purchase products/services from ads.

- None of the above

- Other (free response field)

Q21. Why would you disable personalized advertising? (Select all that apply)

- Relevance: The ads and brands that I see reflect my interests

- Cost Savings: The purchases I have made from personalized ads saved me money

- Convenience: Personalized ads save me time from searching for products

- New Products: Personalized ads have introduced me to new products similar to those I like

- Personalized Content: The personzalized content recommended to me matches my interests

- None of the above

- Other (free response field)

Q22. Are there any other reasons that you would/wouldn't disable personalized ads that were not mentioned above? *(free response question)*