Article

High school students' perceptions of third-party tracking and personalization

Anushka Patil[1], Shilpa Patil[2], Hadjer Benkraouda[3]

[1] Cupertino High School, Cupertino, California

[2] Intuit, Cupertino, California

[3] Polygence, Urbana, Illinois

SUMMARY

Most teenagers between the ages of 13 and 17 have used social media, which plays a significant role in their lives. As a result, almost all teenagers have been subjected to digital tracking. Tracking is prevalent in our everyday lives as large companies use it to monitor our searches to push forth propaganda, advertisements, and products. Tracking can take personal information from trusted users to promote customized advertisements that increase revenue for businesses. Though many studies have been conducted regarding adults, this work has failed to realize that many consumers are teenagers. Hence, in this study we expanded the data on third-party tracking by surveying high school students on the subject matter. We analyzed the data and made inferences to understand common problems in the online world. Our objective was to conduct a behavioral analysis using chi-square tests to assess the reliability of connections between high school students' characteristics and their feelings on corporate tracking mechanisms. We concluded that high school students' feelings towards third-party tracking are influenced by social class, with individuals with higher social class exhibiting greater knowledge of tracking; and screen time, with students with higher screen times feeling more apprehensive about their data being stolen. Our research opens up conversations regarding online safety and minors and contributes to potential tailored privacy solutions, behavioral insights, and technology designs for the future.

INTRODUCTION

As people browse the internet, many are subjected to various forms of targeting and third-party advertising by corporations. About 90% of teenagers between the ages of 13 and 17 have used social media, likely exposing most to this type of targeting (1). Most companies may tailor advertisements, articles, or emails toward a person's personal needs. Companies may label tracking as "personalization" or similar, which appears less malicious and more justified than the former term (2). Due to ambiguities in the word "tracking," we must first clarify here how we intend to use it.

In short, to access the data in a web browser, a data file known as a cookie is implanted within a device when a user opens the browser (3). Normally, cookies vanish when the

browser is closed, but in the case of tracking, the cookie stays on the user's device and records their online activities. First-party organizations install cookies from their own websites on the user's computer and only utilize the data they collect for their own purposes. Third-party organizations pay other businesses to receive information about the user's activities on their website and reflect those activities on the third party's own website (4). Ads are subsequently displayed on the user's feed to reflect this information.

In a study of 306 adults, participants were presented with ads that aligned with their interests, accompanied by explanations pertaining to one of ten reasons for the ad display (3). The results revealed increased discomfort as the targeting techniques became more invasive but showed slight alleviation when the targeting approaches were veiled under the guise of personalization (4). Another study shed light on the prevalence of third-party tracking mechanisms within websites affiliated with US hospitals (5). A cross-sectional study of publicly available website data showed how personal information on patients was allegedly getting leaked (5). Some users are unbothered by tracking or personalization since they may incorrectly assume that companies track their own users for the benefit of the user. Other users may be wary of the data collected on them despite having little to no information on the workings behind third-party organizations (5). Notably, many studies focus on data-tracking methods involve adults (4). Due to ethical constraints, there are not many studies done on teenagers in this field. However, studying teenagers is important as they are among the heaviest users of social media and technology due to their upbringing with these platforms.

Gen Z high school students consume more technology and social media than other generations, making them the most susceptible to tracking (6). Though information on third-party tracking is available for public disclosure, many teenagers may not be aware of these facts. Hence, our study centers around high schoolers, their knowledge and perception of tracking, and their experiences related to it. In this study, 117 students from numerous locations took an anonymized survey with questions regarding their comfortability with tracking. We then used this data to create inferences and connections between specific factors like gender with feeling, class with awareness, and screen time with concern in specific scenarios. Specifically, we focused on the feelings of high school students. Our goals were to identify demographic factors that contribute to student's perceptions of tracking and to help create a diverse research demographic. We found that high school students' feelings towards third-party tracking are influenced by their gender and social class and impact

their comfort levels with online engagement. Our results may influence future tailored privacy solutions, behavioral insights, and technology designs.

RESULTS
Gender and emotion

First, we examined high school students' gender and their feelings toward tracking. In the survey, the participants were given the following options to describe their feelings toward tracking: content, neutral, helpless, irritated, or distraught. No significant trend between feeling towards tracking and gender was seen (p = 0.074, chi-square) **(Table 1)**. Overall, "neutral" was the most popular choice for all genders (52.99 percent). Female identifying participants chose "irritated" second most, with "helpless" as the third most selected answer. For male and non-binary identifying individuals, 'helpless' was the second most selected answer. The experience of helplessness implies a degree of familiarity with the topic at hand, with negative connotations stemming from a perceived inability to exert control over personal data.

Social class and awareness

Second, we examined the correlation between social class and awareness of tracking. In this study we define social class by the students' household income. We assigned social class household income boundaries using average income data from Santa Clara country, where the study took place. In the survey, the students were given five class categories: poor (below $70k), lower middle class ($70k – $90k), middle class ($90k – $110k), upper middle class ($110k – $130k), and wealthy (above $130k). Widening the demographic of the middle class allows for leeway in relation to classifying social class since a large range of people fit within the middle class. Regarding knowledge of tracking, the participants were given the options: "I understand", "I somewhat understand", and "I don't understand." Most participants chose "somewhat" regarding their knowledge of tracking (75.21%) and "I understand completely" was the second most popular choice (13.68%) **(Table 2)**. We found that no one in the poor category chose "I understand", while this was the second most common answer among upper middle-class and wealthy students (13.42%). Additionally, no students in the wealthy category chose "I don't understand" as an answer. Though there were fewer poor and wealthy individuals, the number of subjects in these two categories were similar in number, yet they were distributed differently. We did see an overall correlation between household income and knowledge of tracking (p = 0.035, chi-square) **(Table 2)**.

Online engagement and comfortability

Lastly, we examined the duration of participants› screen time and their responses to three scenarios related to digital

tracking. We provided options for screen time that included: 30 minutes or less, between 30 minutes to an hour, 1 to 2 hours, 2 to 4 hours, or 4 hours or higher. Then, we presented three scenarios describing certain situations any of the subjects may find themselves in concerning third-party tracking.

The first scenario entailed a user logging into the YouTube platform and subsequently receiving customized advertisements based on their previously consumed content as a result of tracking via the same platform. This is a frequent occurrence that we hypothesized would not be alarming to participants. We asked participants to rate "how concerned would you be in this scenario?" on a scale from 1 to 10, from least to most concerned, respectively. This scenario shows that a majority of all participants (73.50%) selected within the range of 3 to 7, with option 5 being the most common choice **(Figure 1)**. 54.70% of participants selected screen times of 2 to 4 hours and 4 hours and above. 60.27% of the selections within the range of 3 to 7 (73.50%) included people with screen times of 2 hours and above. Notably, the collected data is indicative of a near symmetric distribution centered at 5.

In the next scenario, the participants› data was hypothetically purchased by a larger company and subsequently sold to another entity for the purpose of delivering customized advertisements. This scenario specifically describes the use of cookies in third-party tracking. Like in the first scenario, the distribution of the data was relatively symmetric with the most common response being 5 (19.7%). However, there was a shift in the data towards higher numbers with 7 and 8 being tied, both being 15.4%. **(Figure 2)**.

The final scenario presented the most extreme form of third-party tracking, voice-activated tracking, which monitors an individual›s vocal interactions and exploits them for data mining or promotional purposes. This scenario detailed the participant hypothetically talking about a failed school test near their phone and later finding an ad for tutoring on their feed regarding the subject they failed. This scenario garnered a more alarming response than previous scenarios, with the most popular response as 10 (25.60%) and the second most popular response as 9 (17.10%) for all screen time users **(Figure 3)**. Overall, our scenarios suggest that students with higher screen times seemed to feel more concerned about their data being tracked compared to those with lower screen times.

DISCUSSION

Studying feelings with survey questions is challenging. A level of trust must be built before easing into the heavier questions. To help build trust, we slowly increased the invasiveness of our scenarios. We observed several correlations, including of gender with feeling, social class with understanding, and hours online with perceived concern in

Gender	Content	Neutral	Helpless	Irritated	Distraught	Total
Female	1	24	9	14	4	52
Male	4	32	10	6	3	55
Non-Binary	1	6	2	1	0	10
Total	6	62	21	21	7	117

Table 1: Relationship between gender and emotion regarding tracking. Chi-table showing data from participants displaying no correlation, p=0.074.

| Level of Understanding	Poor	Lower Middle Class	Middle Class	Upper Middle Class	Wealthy	Total
I don't understand	1	1	5	6	0	13
I somewhat understand	1	3	37	45	2	88
I understand completely	0	0	5	10	1	16

Table 2: Relationship between financial class and understanding in regard to tracking. Chi-table showing data from participants displaying correlation, p=0.035.
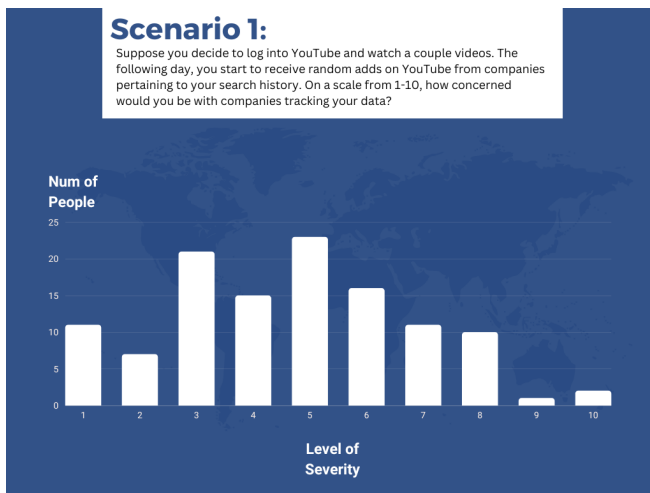
Figure 1: Perceived severity responses for the first scenario examining tracking by large corporations resulting from user's search history. Data from all participants (n = 117) on a numbered scale based on severity, with 10 as the most severe. This graph shows an approximately symmetric distribution with a mode of five.

Figure 2: Perceived severity responses for the second scenario examining tracking by large corporations resulting from user's accepting cookies on a website. Data all participants (n = 117) on a numbered scale based on severity, with 10 as the most severe. This graph shows an approximately symmetric distribution with a mode of five.

certain scenarios. The main purpose of our survey was to evaluate correlations between demographic variables and perceptions and attitudes of digital tracking among high school students.

In the first study on gender and attitudes of security, a participant choosing "content" may potentially suggest that they have a sense of security in having their personal information in the hands of larger companies though it is not confirmed. "Neutral" may imply that they don't have much of an opinion on the matter or they're neither satisfied nor unsatisfied with the concept of tracking. Choosing "helpless" could demonstrate anxiety towards the topic, which may indicate that they are aware of the consequences. "Irritated" may imply they are aware of the consequences and unsatisfied with them, while "distraught" may suggest they are deeply agitated with the matter at hand. It is possible that female students may be more vulnerable to third-party tracking due to evidence that finds they are more frequently targeted by material product marketing, such as makeup and apparel (4). Advertising is the most common form of third-party tracking, and buying material goods and seeing targeted advertisements can increase awareness of tracking (4). A previous qualitative study looked at women's feelings about a new safety app called Anshimi, which tracks its users in order to guarantee their safety due to rising violence against women in South Korea (10). The authors found that many of the women felt that the app is invasive as it tracks their every move, but did not draw a comparison between violence against women and targeted advertising (10). Rather they focused on the practice of data distribution, and how data from the app on these women were sold by the company of the app (10). The women willingly downloaded the app and gave up their data to this app, but they felt that the app was invasive due to their data being sold and used by large corporations to create tailored advertisements for them (10). It is important to acknowledge how the p-value garnered from our first study indicated no correlation. The Anshimi study correlates with the data point in our study on how the second most selected answer for female identifying individuals was "irritated" (10). The Anshimi study
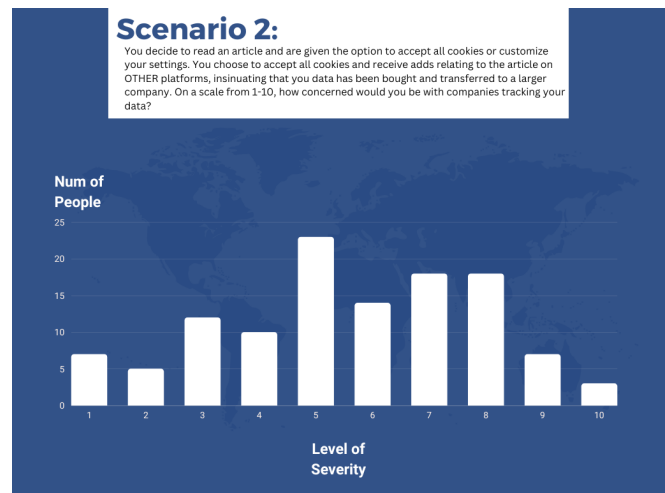
is the extent to which other studies done regarding feelings on tracking in relation to gender match up with the results we garnered, making the lack of correlation accurate for our particular study. To our knowledge there are no other studies done on male and non-binary identifying people that reflected the results we had such as feeling "helpless." It is crucial to acknowledge the limitations to this particular study due to the medium. A google form was used and it was difficult to conduct a thorough analysis of particular feelings on tracking due to emotions being incredibly dynamic and subjective to a person. Additionally, choosing the words that would be used to conduct the survey was difficult in particular as many people can feel many different things and the words available may not be what the participant is feeling. These limiting factors may have been the cause for lack of correlation. In the future, using a free response survey question could be used to improve this study.

The second study examined how social class and knowledge on tracking are correlated. Our results suggested a correlation between social class and knowledge of tracking, which may in turn influence feelings toward tracking. The more knowledge one has, the more of an experience-based opinion one can form on tracking (7). Our final study examined how the participants felt regarding the increase in severity of tracking. For this study, three scenarios were used with the severity of tracking increasing with each scenario. The participants had to select how comfortable they felt with each scenario from a scale from 1 to 10 with 1 being the most comfortable and 10 being the least comfortable. The first scenario had a relatively symmetric response shifted toward the left, or comfortable side, as it was the least severe form of tracking. The second scenario had a very symmetric response as the severity in tracking went up. The last scenario was skewed toward the right, or uncomfortable side, the most due to it being the most severe in tracking.

Overall, our findings suggest correlations between high school students› feelings towards third-party tracking and social class, with individuals with higher social class

Figure 3: Perceived severity responses for the third scenario examining tracking by large corporations resulting from user speaking near a device. Data from all participants (n = 117) on a numbered scale based on severity, with 10 as the most severe. This graph shows as skewed distribution with a mode of 10.

exhibiting greater knowledge of tracking; and screen time, with high school students with higher screen times feeling anxious about their data being stolen. This behavioral study serves as a contribution toward how the underrepresented high school demographic feels about tracking and ultimately contributes to diversifying the research done on feelings and third-party tracking, hopefully equipping students with the knowledge to make informed decisions about their online activity. Hopefully, in the future, this experiment can be replicated on a larger scale to garner a precise understanding of high school student's behaviors. Future applications of this research on a wider scale may include: informing educational initiatives on digital literacy and privacy awareness, guiding the development of privacy-focused technologies tailored for young users, and contributing valuable insights to policymakers and stakeholders regarding ethical data practices in educational contexts.

Additionally, choosing the words that would be used to conduct the survey was difficult in particular as many people can feel many different things and the words available may not be what the participant is feeling. These limiting factors may have been the cause for lack of correlation.

MATERIALS AND METHODS

When brainstorming questions for the survey, a great amount of consideration was put into each. We started out by listing quantifiable measurements we could ask of the participants to compare different statistics. We also thought of different ways in which participants could get tracked, and then created questions surrounding them like the study with the tracking scenarios. Upon beginning the survey, the participant was provided the consent form exhibiting pre-survey assurances and regulations regarding confidentiality, benefits, and the purpose of the project. Once the participant had confirmed parental consent, the first set of questions was asked while keeping the participant completely anonymous.

The survey started with "What do you identify as? (pronouns)" and had the options of she/her, he/him, they/them, and other. The next question was "What is your household or parent/guardian income level?" and had the options of: poor (below 70k), lower middle class (70k – 90k), middle class (90k – 110k), upper middle class (110k – 130k), and wealthy (above 130k). The next question was "How much social media do you use?" and had the options of: lower than 30 min, 30 min – 1 hour, 1 hour – 2 hours, 2 hours – 4 hours, 4 hours and above. The next question was "To what extent do you know how tracking works?" and the options were: I don't understand, I somewhat understand, and I understand it completely. The next question was: "How would you feel about companies knowing about your personal information?" and the options were: content, neutral, helpless, irritated, and distraught. For the scenarios, the participants were given three scenarios and had to rank them from 1 to 10 from most comfortable to least comfortable with 10 being the least comfortable. The first scenario was: "Suppose you decide to log into YouTube and watch a couple videos. The following day, you start to receive random adds on YouTube from companies pertaining to your search history. On a scale from 1-10, how concerned would you be with companies tracking your data?" The second scenario was: "You decide to read an article and are given the option to accept all cookies or customize your settings. You choose to accept all cookies and receive adds relating to the article on OTHER platforms, insinuating that you data has been bought and transferred to a larger company. On a scale from 1-10, how concerned would you be with companies tracking your data?" The last scenario and question in the survey overall was: "After discussing how poorly you performed on your exam today with your phone within arms reach, you start to see tutoring advertisements on all of your social media platforms, including those for tutoring in the subject you struggled with. How worried would you be with companies tracking your data, on a scale of 1 to 10?"

These questions were crucial in understanding the user's knowledge regarding tracking. The emotional reactions are mainly fact-induced shock factor from information that may be new to the user as the questions progressed. Lastly, participants were asked a series of scenarios that increase in severity that capture some of the realities of tracking. These last couple of questions helped examine the user's underlying feelings on tracking. When creating the scenarios, we sought to include realistic but unnerving circumstances that reflect everyday situations. The scenarios simulated day-to-day tasks that may seem normal but later spew the reality of each situation. For example, one of the scenarios proposed a situation in which the subject recently failed an exam and mentioned it aloud near their phone. The scenario then described the subject receiving multiple tutoring advertisements tailored to the subject with which they struggled. After reading each scenario, the user was asked to rate a score from one to ten based on how worried they would feel if companies tracked their data.

Once all the results have accumulated, Google Forms has a page that shows pie charts of exact percentages. These percentages are later used to aid in the creation of making connections. In order to create the p-values to see if there were connections, we used resources to learn how to make chi-square tables (9). We used Google Sheets to input our data from the Google form to create a chi table and then used equations from tutorials to conduct p-tests to garner p-values to see correlations in our data (9).

REFERENCES

1. Vogels, E. A. "*Teens and social media: Key findings from Pew Research Center Surveys*." Pew Research Center, 24 April 2023, https://www.pewresearch.org/short-reads/2023/04/24/teens-and-social-media-key-findings-from-pew-research-center-surveys/.
2. Farke, Florian M., et al. "Are Privacy Dashboards Good for End Users? Evaluating User Perceptions and Reactions to Google's My Activity (Extended Version)." *arXiv*, 28 May 2021, https://doi.org/10.48550/arXiv.2105.14066.
3. Shah, Vidhi Kirit. "User Acceptance of Online Tracking If 'forgetting' Was an Option." CURVE, 24 July 2020, https://curve.carleton.ca/931b9c64-2eda-4e1e-babd-7cff16f85470. Accessed Jun. 2023.
4. Dolin, Claire, et al. "Unpacking Perceptions of Data-Driven Inferences Underlying Online Targeting and Personalization." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, https://doi.org/10.1145/3173574.3174067.
5. Niforatos, Joshua D et al. "Prevalence of Third-Party Data Tracking by US Hospital Websites." *JAMA Network Open*, vol. 4,9 e2126121. 1 Sep. 2021, https://doi.org/10.1001/jamanetworkopen.2021.26121.
6. Natã M. Barbosa University of Illinois at Urbana-Champaign, et al. "Who Am I?: A Design Probe Exploring Real-Time Transparency about Online and Offline User Profiling Underlying Targeted Ads" *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol 5, No 3, 1 Sept. 2021, https://doi.org/10.1145/3478122.
7. Springtide. "Gen Z Pandemic Research and Resources." *Springtide Research Institute*, 20 Apr. 2023, www.springtideresearch.org/gen-z-pandemic-research-and-resources. Accessed Jun. 2023.
8. Turney, Shaun. "Chi-Square (X2) Table: Examples & Downloadable Table." *Scribbr*, 10 Nov. 2022, www.scribbr.com/statistics/chi-square-distribution-table/.
9. Zach. "How to Read the Chi-Square Distribution Table." *Statology*, 2 June 2021, www.statology.org/how-to-read-chi-square-distribution-table/#:~:text=The%20Chi%2DSquare%20distribution%20table%20is%20a%20table%20that%20shows,0.01%2C%200.05%2C%20and%200.10). Accessed Jun. 2023.
10. California, Chaeyoon Yoo University of, et al. "Anshimi: Women's Perceptions of Safety Data and the Efficacy of a Safety Application in Seoul." *Proceedings of the ACM on Human-Computer Interaction*, Vol 5, 1 Apr. 2021, https://doi.org/10.1145/3449221.