Youths' Perceptions of Data Collection in Online Advertising and Social Media

CAMI GORAY, University of Michigan, USA
SARITA SCHOENEBECK, University of Michigan, USA

This project illuminates what data youth believe online advertisers and social media companies collect about them. We situate these findings within the context of current advertising regulations and compare youth beliefs with what data social media companies report collecting based on their privacy policies. Through interviews with 21 youth ages 10-17 in the United States, we learn that participants are largely aware of how their interactions on the website or app are used to inform personalized content. However, certain types of information like geolocation or how long data is retained is less clear to them. We also learn about what school and family factors influence youth to adopt apps and websites. This work has implications for design and policy related to companies' personal data collection and targeted advertising, especially for youth.

CCS Concepts: • Social and professional topics → Adolescents; • Security and privacy → Privacy protections; • Human-centered computing → Human computer interaction (HCI).

ACM Reference Format:
Cami Goray and Sarita Schoenebeck. 2022. Youths' Perceptions of Data Collection in Online Advertising and Social Media. Proc. ACM Hum.-Comput. Interact. 6, CSCW2, Article 475 (November 2022), 27 pages. https://doi.org/10.1145/3555576

1 INTRODUCTION

"Today, the content moderation and data collection practices of Big Tech remain largely hidden to consumers...Users also remain in the dark about what data is being collected about them, how it's being used and to whom it's being sold and for what purpose."
- U.S. Senator Roger Wicker at the U.S. Senate hearing "Protecting Kids Online: Facebook, Instagram, and Mental Health Harms" on September 30, 2021

Young people's online experiences are shaped by how algorithms filter content—content for social media posts, search results, advertisements and more. The algorithms driving many recommender systems, such as the YouTube search results or the TikTok For You stream, track users' interactions, circulate content, and curate personalized experiences. They influence what content the user sees, as well as how the user is "seen" (and monetized) by the platform. Furthermore, companies may use the data traces to build algorithmically-mediated profiles of users and target them with ads. How this process of personalization works, though, is oftentimes a mystery for both adults and children. Importantly, much of this process is invisible to the user [41]. In this study, we investigate young people's perceptions of data collected by both online advertisers and social platforms. We contrast this with what data tech companies state they collect, as documented in their privacy policies.

Authors' addresses: Cami Goray, University of Michigan, 105 S. State St, Ann Arbor, MI, USA, 48109, cgoray@umich.edu; Sarita Schoenebeck, University of Michigan, 105 S. State St, Ann Arbor, MI, USA, 48109, yardi@umich.edu.

2573-0142/2022/11-ART475 $15.00
https://doi.org/10.1145/3555576

While journalists, activists, and academics for years have called for growing attention to dubious data collection practices of social media, recently, there has been a heightened interest among lawmakers in how the ad targeting that results from such practices can impact youth. Specifically, U.S. lawmakers have called to update the Children's Online Privacy Protection Act (COPPA) which was enacted in 1998, many years before today's sophisticated ad profiling techniques were conceived. Under COPPA's provisions, websites are prohibited from collecting personal information from children under the age of thirteen without "verifiable parental consent." Tech companies, too, have taken notice. U.S. Federal Trade Commission lawsuits over children's data tracking against TikTok (2019), YouTube (2019), and Xanga (2006) may have prompted tech companies to re-calibrate their targeting restrictions for young users. Instagram announced that they would limit advertisers to only target youth under age based on age, gender, and location.[1] In a more aggressive measure, Google announced that they would completely block ad targeting based on age, gender and preferences of youth under age 18.[2] Facebook announced that they would remove detailed targeting options that relate to topics people may perceive as sensitive, like health causes, sexual orientation, religious practices, and political beliefs.[3] Clearly, the tech industry is experiencing a sea change in its politics of who is categorized and targeted, and how.

The categories platforms label users with generate a language that constantly parses and re-creates their digital identities. The process itself of classifying users' interests is inherently political [18]: users who are "marked" are assumed to be subordinate to the unmarked user, which is the default. Researcher Joy Buolamwini refers to the way that algorithmic systems inherently reflect the priorities, preferences, and prejudices of the builders of the technology as the "coded gaze" [16].

Taking a historical perspective, we see how infrastructures of tracking and classification have evolved [63]. Many institutional systems in the United States, including education, health, and prison systems, already leverage digital and non-digital infrastructures to categorize and track youth. How these systems classify youth can have major implications for their options later in life, shaping the educational and occupational opportunities and trajectories that might become available–or unavailable–to them [12, 29]. Micro-targeting behavioral advertising is an extension of these institutional systems in how it integrates assumptions and predictions [37] about what preferences and pathways young people should be exposed to via ads.

Targeted advertising relies on algorithmic profiling, which is the systematic and purposeful recording and classification of data related to individuals [15]. Users' data profiles are used to predict what targeted advertisements will be most relevant to them. In a review of harms detected by algorithmic audits, Bandy illustrates how discrimination in online advertising can occur when algorithms make assumptions about who is searching, for example, when companies rely on sensitive personal attributes like race and gender to target users [7]. Research has revealed that up until recently Facebook still labeled users with sensitive ad preferences relating to religious beliefs, ("Quran", "Christianity"), political opinions ("nationalism," "socialism"), and sexual orientation ("LGBT community") [17].

How algorithmic systems can better communicate processes to users is a persistent challenge in HCI [3]. More work needs to be done to understand how users want information about ad targeting processes communicated to them. To shed light on how young users conceptualize the data profiles companies have on them, we ask the following research questions:

RQ1a: To what extent are youth aware of online advertisers and social platforms collecting data about them?

[1]https://about.instagram.com/blog/announcements/giving-young-people-a-safer-more-private-experience
[2]https://blog.google/technology/families/giving-kids-and-teens-safer-experience-online/
[3]https://www.facebook.com/business/news/removing-certain-ad-targeting-options-and-expanding-our-ad-controls

RQ1b: If they are aware, what information do youth believe social platforms and online advertisers have about them?
RQ2: How do youth react to a company's ad profile of them?
RQ3: How do youth consider trade-offs between usefulness and creepiness in profiling?

The first two research questions (RQ1a, RQ1b) focus on youths' expectations and understandings of what companies know about them (What is the gap in what they know vs. what they want to know?), whereas the subsequent research question (RQ2) focuses on how they respond to seeing an example of one form of data that the companies does store on them: their ad inferences. Together, these two questions are meant to bridge the discussion between data collected and data inferred/assumed. The last research question (RQ3) concentrates on a list of identity attribute categories. Even though companies may not explicitly collect personal information about sensitive identity attributes, past research has shown how people's preferences can link them to that identity. The data elements reportedly used by data brokers in the Federal Trade Commission's 2014 report, "Data Brokers: A Call for Transparency and Accountability"[4] motivated the inclusion of many identity attribute categories in the list for this study.

This research aims to inform policy for transparent personal data collection, specifically, how companies should communicate what data they collect and how they make inferences about their users. It follows a call by Milkaite et al. for the need for more research to "not only concentrate on the data that children give knowingly but also focus on their understanding of data traces (which are usually left unintentionally) and inferred data which is rooted in given data and its traces" [52].

2 RELATED WORK

In this section, we draw upon prior work on youths' awareness of their data traces. We then synthesize literature on advertising inferences and targeted advertising, framing the findings in terms of how they relate to current understandings of research and policy on youth privacy.

2.1 Youth Awareness of Data Traces

Adolescence is marked as a time when youth gain more independence, both offline and online. How do youth make sense of their digital worlds? A growing body of literature has examined children and teens' awareness of their online digital traces and their understanding of data processing [11, 42, 55, 67]. We see trends across age ranges from children to teens. Sun et al.'s study revealed that younger children ages 4-10 believe data to be static and isolated within the technology device, and they interpret privacy risks as interpersonal [67]. They argue that digital literacy education on commercial privacy i.e., personal data that is harvested and used for business and marketing purposes, should be given more attention. Likewise, according to Livingstone et al., commercial privacy is an area that children are least aware about and are least likely to comprehend [42]. Meanwhile, while teens are more likely to be aware that their data is sold [36], they tend to treat it as a peripheral, secondary concern [55]. A 2014 study with teens 12-17 found that children are less concerned about their data being collected by marketers than their parents [26]. Furthermore, while teens may understand the purposes of social media data for marketing, they do not necessarily associate this with the term 'data' [11]. Bowler et al. argue that their conceptualization matters because awareness of data as concrete, tangible and locatable makes it easier to manage [11].

Algorithms play a central role in how youth encounter information online. Automated algorithms mediate the distribution of information in news feeds and ads, creating dynamic experiences that are tailored to individual users. Some research has revealed that online behavioral advertisements

[4]https://www.ftc.gov/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014

may subtly shape people's perceptions about themselves. For instance, the types of gendered advertisements people encounter may color their self-confidence and leadership ability [27]. Marketing science literature suggests that people may perceive that behaviorally targeted ads function as an implied social label [66]. This phenomena is critical to study with young people because one's identity is especially fluid during adolescence [19]. During adolescence, personality traits begin to stabilize and individual differences in self-esteem emerge [48]. French likens personalized advertisements, search results, and friend suggestions as a type of 'algorithmic mirror' that reflects a system's approximation of the self back to the person, a form of feedback for the person [27].

Yet the feedback fed to online algorithmic systems is usually not confined to the individual user. Through collaborative filtering, a person's data profile is influenced by how similar users behave online. Those users may not necessarily be in an individual's network. In contrast, with "shadow profiles", which are profiles that companies create about a person, the data comes from someone in their social network through the user granting the site access to their social contacts [28]. One study found that personal information of non-users, such as their sexual orientation and relationship status, of the now defunct social networking site Friendster could be predicted based on the activity of other users on the site [28]. With regard to youth's experiences online, Villebro et al.'s survey discovered that teens were largely uncomfortable with the idea that a platform could target their friends with ads based on their own activity on the site [69]. These studies establish that the boundaries of one's data profile may extend beyond their own digital activity and that in reality, one's online digital traces may be outside of one's control.

Youths' expectations of how companies handle their data are not one-size-fits-all. Villebro et al. discovered that teens' privacy expectations vary widely based on the social platform [69]. For instance, more youth were fine with Facebook using their data to target them with ads compared to Snapchat. Moreover, significantly more youth trusted Snapchat not to sell their data to third parties compared to Facebook. The results prompt further exploration of what platform affordances signal the trustworthiness of commercial data collection. They also motivate research into comparisons between platforms, which our study addresses.

Together, these findings highlight that youths' awareness of data traces and consequent information sharing norms are highly contextual [53]. Secondly, users are unaware of the diverse ways firms collect and use their data, resulting in information asymmetries [1].

2.2 Online Advertising Inferences

Targeted advertising is becoming more pervasive and more sophisticated. It allows advertisers to reach users based on demographics, psychographics, and other profiling information. Specifically, online behavioral advertising (OBA) is advertising that is targeted at individuals based on the observation of their behavior over time, oftentimes via automated processing.[5] A survey of online behavioral advertising (OBA) literature reveals that there is an especially large gap in consumer knowledge about OBA [10].

Many social platforms rely on advertising to generate revenue, in a kind of symbiotic relationship. Ad exchanges host "real time auctions" for individual ad impressions on web pages where bidding happens within milliseconds. A wide range of data types may be shared with advertisers, third-parties and data brokers, including browser and search history, email and social media activity, behavior tracked on smartphones, and ofline activity [5, 72].

Prior work on OBA has pointed to four key problems: (1) a lack of user awareness and participation [2, 68], (2) insuficient user understanding of how profiling works [49, 68], (3) profiles that misrepresent the user [2, 76], (4) and a lack of control over profiling [35, 49]. Barbosa et al.

[5]https://iapp.org/resources/article/behavioral-advertising-2/

synthesized these problems as relating to engagement, interpretation, representation, and action [8]. To ameliorate these issues, an emerging line of work has proposed various user-facing transparency tools for OBA [21].

One potential transparency tool is the privacy dashboard. Privacy dashboards allow users of online services to review and control data collection [25]. A small number of companies provide privacy dashboards that display to users information that the company has inferred about them. For example, Google Ad Settings is one such transparency tool [22]. However, it is not obvious how the inferences are made based on the data collected. Furthermore, the data dashboard is curated; it does not necessarily contain everything that the company has on the user, only what they choose to share with the user [20]. Some companies may even intentionally choose not to collect data about their users, like their age, in order to avoid responsibility, what a Data & Society report labels as "strategic ignorance" [39]. As noted by Stoilova et al., individuals, institutions, and businesses make inferences based on data collected, yet few studies on youth's privacy have looked into data traces and inferred data. As a result, there is a lack of perspective from children to inform policy deliberations regarding the risks of profiling [65]. We are left to wonder at what point assumptions made about users based on various fragments of their digital identity infringe on their privacy and autonomy. Our proposed study aims to address this gap by focusing on children (particularly teens who have been absent in policy), OBA, and algorithmic inferences.

Oftentimes, inferences about users manifest in "afinity profiling." According to Wachter, afinity profiling is profiling which does not directly infer a user's sensitive data ("special category data"), such as personal traits or membership in protected groups, but rather uses other data to measure the user's "afinity" for groups [70]. In other words, afinity profiling looks for a similarity between the assumed interests of a user and the interests of a group. Many of the profile categories that appear on Google Ad Settings could count as afinity profiling. Importantly, Wachter points out that these categories, which are inferred from online behavior, can act as proxies and may potentially signal sensitive information such as religious or political beliefs, sexual orientation, race or ethnicity, physical or mental health status, or sex or gender identity without users ever being aware [70]. Some user studies have shown users' caution about afinity profiling. For instance, in Woodruff et al.'s workshops on algorithmic fairness with Black and Hispanic users and low-socioeconomic status users, participants reported disliking personalized ads based on the online behavior of people with similar characteristics because of its potential for stereotyping [74].

Another line of work within the HCI and privacy communities has focused on ad settings and data profiles as windows to understanding people's attitudes and conceptualizations of OBA [14, 21, 31, 57, 59, 72]. A 2019 Pew Research survey found that just over half of adults reported feeling uncomfortable with how their Facebook ad settings categorized them [32]. Work by Rao et al. showed specific types of data that concerned users; users were most concerned about the level of detail and the inclusion of sensitive information like health or credit information in their BlueKai, Google, and Yahoo data profiles [59]. Other work has specifically untangled the user's process of sensemaking. Rader et al. concluded that users' rationales of the inferences they saw on Facebook and Google Ad data profiles were influenced by their self-perception, past online behavior, and interests of family members [57]. The findings suggest interesting research directions related to how one's social environment shapes their perception of algorithmic processing and/or personalisation.

Related research on users' perceptions of online advertising and algorithmic processing has paid attention to ad explanations [6, 23, 73], which appear as a small icon in an ad accompanied with the text "Ad Choices," "Why am I see this ad?" or "Why this ad?" In response to research showing significant inconsistencies around how ads should be explained, Wilkinson et al. propose a classification scheme of ad explanations that organizes explanations based on input parameters

(e.g., inferred attributes vs. stated attributes) and input source (e.g, app/site, advertisers, third parties), among other attributes [73]. Studies have demonstrated that users like more personalized explanations. For example, Eslami et al. observed that users were more satisfied with the explanation when it clearly conveyed a "linkage" to their identity [23]. Barbosa et al.'s study takes the discussion on transparency/visibility of user profiling one step further by emphasizing the need for visibility at the right time [8]. They discovered that users were not aware of their inferred identities created by Google or Facebook and wished such algorithmic identities were more available to them. They desired to be invited to review, and at certain moments to control, pieces of their ad-targeting profiles. What this tells us is that users not only want the information visible, but they also want to provide feedback when necessary. This reflects a research direction in algorithm studies oriented towards algorithm contestability [44]. There is still much work to be done on how ad explanations and ad dashboards can better address users' needs.

To reiterate, all of these studies were done with adult users and while technically many companies' policies require users to be 13 to use their platform, research has yet to address adolescent populations and their perspectives on data profiling.

2.3 Youth and Advertising Regulation

Designing ad explanations and privacy settings for youth presents unique challenges. Up until recently, research on children's understanding of advertising has been limited to television and print; modern-day advertising appears in a more variable format [58]. Teenagers are in a unique stage of their development when their peer relationships are beginning to have a profound influence on their choices and preferences [13]. Moreover, the part of their brain that controls inhibitions does not fully mature until late adolescence or early adulthood [56]. Radesky et al. argue that children may be less aware of privacy risks and thus need additional protection with regard to their personal data. For instance, disclosure of data collection should be prominently provided at appropriate literacy and developmental levels [58].

Besides youth's cognitive development, other factors related to the nature of targeted advertising directed at children make them vulnerable. In the US and UK Google Play stores, news apps and apps targeted at children have the highest amount of third-party trackers [9]. Additionally, gamified advertising is becoming popular, especially in apps geared towards children [4]. Gamified ads are less conspicuous and have game-like interactive components that can mislead children.

Laws governing technology regulation have a lot of catching up to do. Recently, U.S. lawmakers have introduced bills to update COPPA, including a bill that would expand existing law to include teenagers under 18, prohibit companies from using targeted advertising (as opposed to contextual advertising) directed at children, and require companies to get consent from teens before collecting their data. Unfortunately, COPPA is not always enforced, as demonstrated by the number of violations discovered in children's apps [60]. It also does not apply to websites that are targeted to a general audience, as it is still dificult to reliably distinguish between adults and children online [45].

3 METHOD

3.1 Recruitment

We conducted semi-structured interviews (n=21) to understand youths' perceptions of both social media and online advertising data collection and profiling practices. Our inclusion criteria required participants to be between 10-17 years old and live in the United States. We chose not to restrict

participation based on whether they used certain social media in order to leave room for cross-platform comparison, in line with similar research with youth [69]. The study was approved by our institution's ethics review board.

Because the participants were minors, we required assent from the child and consent from their guardian. All interviews were conducted via Zoom, audio-recorded, and transcribed for analysis, with participants' consent/assent. They were conducted between June and August 2021 and averaged 47 minutes. We set a limit for the interview to one hour to avoid participant fatigue, and we gave participants the opportunity to pause or end the interview at any time. Participants were compensated $25 for their time and expertise.

Recruitment consisted of several waves. The first three participants responded to a study post on the Nextdoor app. The first author also reached out to local community centers to invite participants from outside of university communities. The majority of participants came from a research recruitment website at a large university. While the research recruitment website is not restricted to people who live local to the university, the majority of participants were local residents. The racial demographics self-reported by participants was: 15/21 white, 2/21 Asian, 2/21 Black, 1/21 Hispanic, and 1/21 Mixed Race. Eleven of the 21 participants identified as female and 10 as male. Ten of the 21 participants were between the ages of 10-13, while 11 of the 21 participants were between the ages of 14 and 17. We did not ask about family income or parent education. Results from this study should be interpreted in the context of our participant demographics which may not reflect a wide range of experiences in the U.S. or globally. Additionally, the study was advertised to be about perceptions of social media and online advertising. Thus, participants who were willing to sign up may have different online habits and perceptions than those who did not participate.

In the United States, COPPA applies to children under 13 years, which means many sites restrict, at least ostensibly, younger children from creating accounts. However, it is no secret that many children under 13 years old self-declare their age to be older than it actually is in order to access the platform. A 2016 survey by Common Sense Media found that 23 percent of tweens age 8-12 had a social media account [38]. For this reason, our study includes 10 to 17-year-olds.

3.2 Procedure

We chose semi-structured interviews as the key method for this project because they allowed follow-up questions, enabling the child to elaborate on their answer. Special considerations were taken into account when interviewing the youth. Throughout the interviews, the researcher was cognizant of not leading the participant or using technical jargon; instead, they tried to mimic the language that the participants themselves use to describe algorithms in social media and advertising recommender systems. Through several iterations, interview questions were revised to be less "test-like," (i.e., like they are testing the participant's knowledge), but still address the constructs of interest. The first author conducted several pilot studies to refine the questions. In the consent form participants were asked to find a quiet, private place for the child to do the interview. A couple of the younger youth participated in the interview with their parents in the room; however, the majority of participants gave the interview by themselves.

The interview was divided into four parts that moved from general to more specific. At the beginning of the interview, we asked about participants' technology use broadly: which apps/websites they used and why. The aim was to understand the participants' motivations for their online activity, whether related to school or leisure. We also asked about whether the participants had their own Google account (which would be used later in the interview for the Google Ad Settings Activity). Account sharing impacts one's personalized experience online. This part of the interview led to discussions about how technology was shared among siblings and monitored by parents. Privacy risks associated with account sharing have been documented in the HCI literature [54].

Additionally, participants 12 and under were asked about their opinions about and experiences with social platforms specifically geared towards children, e.g., YouTube Kids. Because the youth we interviewed ranged in age from 10 to 17, we heard from youth who had various levels of autonomy in their digital life.

For the second part of the interview we proceeded to ask about participants' perceptions of the specific websites/apps they had mentioned. We asked whether they believed X website had data about them. If they responded yes, we probed with a series of questions to ask what kind of data the website had about them.

The third part of the interview focused the discussion on online advertising. In a parallel format to the previous section, we asked similarly structured questions about whether they believed online advertisers had data about them. This part of the interview was designed to learn about the participants' understandings and expectations of the data profiles companies had on them. In contrast, during the Google Ad Settings activity, we directed participants to a data dashboard that contained data–or rather, inferences from that data–that a company had on them. After they had time to review the ad categories on the page, we asked them about their overall impression of what they were seeing and then proceeded to ask more detailed questions. For example, we asked them to talk about five to ten categories that stood out to them and why. We did not want to force them to disclose anything they were uncomfortable sharing, so we did not ask them to share their screens. At the end of this part of the interview, we also asked if they would choose to keep their ad personalization settings on after seeing this page. We wanted to learn whether seeing this information would influence them to configure their settings or whether it had no effect. Three participants who did not have their own Google account or who shared their account with a parent did not participate in this activity.

Finally, we conducted an identity mapping activity in Miro, a virtual whiteboard software. The activity prompted participants to imagine how they would feel if advertisers and/or social platforms assumed a certain identity attribute about them in order to personalize the content or ads that they saw. The nine identity characteristics discussed in this activity were age, gender, race, personality, socioeconomic status, sexual orientation, mental/physical disability, political affiliation, and religion. The majority of the identity attributes are considered protected characteristics in the U.S., i.e., it is against the law to discriminate against someone based on these characteristics [70]. (However, as noted in the literature review, while ostensibly companies do not target based on this information, the inferences and affinity profiles can link them to certain identities.) Google's privacy policy reports that they collect demographic information such as age, gender, and language as well as geolocation.[6] They do not show personalized ads based on race, religion, or sexual orientation. TikTok's privacy policy reports that they collect age and geolocation.[7] Instagram's privacy policy reports that, in connection to Facebook data, "You can choose to provide information in your Facebook profile fields or Life Events about your religious views, political views, who you are "interested in," or your health. This and other information (such as racial or ethnic origin, philosophical beliefs or trade union membership) could be subject to special protections under the laws of your country." [8]

Participants were asked to map the attributes on the board in two dimensions: usefulness and creepiness. In other words, how useful was it to them for advertisers to assume X attribute in order to personalize the content they saw? How creepy was it to them for advertisers to assume X attribute in order to personalize the content they saw? The pairing of these two dimensions reflect

[6] https://policies.google.com/privacy
[7] https://www.tiktok.com/legal/privacy-policy-us?lang=en
[8] https://privacycenter.instagram.com/policy/

the tension that users of online advertising algorithmic systems experience: they appreciate being "seen" by advertisers, yet also find such intimate surveillance as disturbing–an intractable trade-off of personalization [61]. This approach follows prior work that explores how users negotiate the utility versus creepiness of targeted ads [68]. We gave participants the freedom to choose which identity attributes they wanted to map. The participants were asked to talk through the attributes they selected, one by one. The activity was meant to help the participants better visualize and compare how they felt about each of the identity attributes.

3.3 Data Analysis

The first author analyzed interview transcripts using open coding and axial coding. The first author wrote notes immediately after each interview. In the first pass through the data, they conducted a line-by-line analysis and coded salient quotes related to the research questions. They then compared all participants' answers for each of the four above outlined interview sections, documenting similarities that they discovered. The codes were then iteratively grouped together in an outline to construct a hierarchy of themes. Finally, they looked at the privacy policies, both adult and child versions, in the main social platforms discussed in the interviews: YouTube, TikTok, and Instagram. Specifically, they focused on the language in the sections related to data collection and data sharing. As described in the results, this was useful to find (mis)alignments with participants' beliefs about companies' data collection practices.

The process for analyzing the data from the identity mapping activity was somewhat different. After each interview, the first author took a screenshot of the Miro board to capture the participant's Useful vs. Creepy map. We then compiled all of the screenshots into a single document. During this stage, we inspected participants' responses both in the transcript and on the map and noted any (in)consistencies in how they felt about each of the identity attributes. We looked at the data from both an intra- and inter-participant perspective. For example, for the intra-participant perspective: Did the participant's reasoning differ between identity categories? For the inter-participant perspective, we honed in on an individual identity attribute and compared all participants' responses.

4 RESULTS

The findings from this study are organized into four sections. In the first section, we share findings about participants' motivations for creating accounts and using the websites and apps that they discussed during the interview. (4.1 Technology Adoption and Motivations) The subsequent three sections correspond to the three main research questions of this project. RQ1: To what extent are youth aware of online advertisers and social platforms collecting data about them? (4.2 Awareness of Data Tracking); RQ2: How do youth react to a company's ad profile of them? (4.3 Google Ad Settings Activity); RQ3: How do youth consider trade-offs between usefulness and creepiness in profiling? (4.4 Identity Attribute Mapping Activity)

4.1 Technology Adoption and Motivations

The social apps most frequently reported by participants were YouTube, TikTok, and Instagram. Six participants reported using YouTube for as long as they could remember or from a very young age. In some cases, it was interesting to learn what the participants considered to count as "social media." For example, P18 did not count YouTube as social media. When asked how long they have been using YouTube and TikTok they replied, "Pretty much like my whole life I think, um, and, whereas social media actually is very recently because I got my phone in eight[h] grade" (P18). Three participants said that they started using TikTok during the pandemic because they were bored.

Table 1. Participant social media and websites used.

P	Apps and Websites
1	Instagram, Snapchat, TikTok
2	Netflix, Disney Plus, Google, TikTok, Snapchat, YouTube
3	Google, YouTube, Netflix, Disney Plus
4	YouTube, Instagram, Facebook Messenger
5	Google, YouTube, Messenger Kids, Roblox
6	TikTok, Snapchat, text messages
7	Google Classroom, YouTube
8	IXL, Wikipedia, online games, YouTube
9	Instagram, Facebook, Snapchat, TikTok
10	Google Classroom, Messenger Kids, YouTube, Pandora, OverDrive
11	YouTube, Roblox, Instagram
12	Instagram, Twitter, TikTok, Snapchat, YouTube
13	YouTube, Roblox, Quora, TikTok, Google Classroom
14	YouTube, online games
15	YouTube, Google, Gmail, Instagram, Netflix
16	YouTube, Disney Plus, Netflix, Hulu, Google Hangouts
17	YouTube, Instagram, TikTok
18	TikTok, YouTube, Snapchat, Instagram, Khan Academy, Google Apps, Discord
19	YouTube, Roblox, Hulu, Netflix, Discord
20	YouTube, Snapchat, Instagram, Google
21	TikTok, YouTube, Snapchat, Instagram, Gmail

4.1.1 Account Creation: Influence of Family and School. Online accounts are explicit profiles created by the user. In many ways, their purpose is inherently for personalization. Though many websites technically prohibit users under 13 from creating accounts, we want to acknowledge the complexity of technology adoption. Two prominent themes that arose in the interviews were the influence of family and school.

HCI scholars have recognized account sharing as pervasive in the household [47, 54]. At least two participants in this study did not remember when they first got their Google account because it was so long ago. Those who did sometimes recounted the role of family members. For P2, her family created her account as a form of diary or life scrapbook. They made her Google account when she was a baby so that her parents and grandparents could send her photos and read the messages aloud to her. P4 said that she recently got a smartphone but up until then, she shared her phone with her mother. "I shared it with my mom, but like I just recently got my smartphone like a year ago on my birthday. Because I got my driver's license, so if I am driving alone so I have my phone with me. Before that, I used to share it with my mom" (P4). One participant talked about how he noticed when his sister had accessed his YouTube account based on the videos recommended on his homepage. He then described his clever strategy to adjust the algorithm: he selected a video he liked and turned on autoplay so that a series of similar videos would appear (P7). Another participant talked about how they shared with their sibling an account for Roblox, which is an online gaming platform. "Roblox, so me and my brother are basically share[ing] an account. So like I had one, when I got it for my ninth birthday, I got a Roblox account, and it had parental settings so I couldn't like play certain games. I couldn't use chat in Roblox. So, um, I abandoned in that account, so now I only play on my brother's, which I can play all games, but I cannot chat in it. I used my brother's

account because I want to play certain games" (P5). P21 said when he was a high school freshman when his friend made his Snapchat account without his knowledge: "It wasn't even me who made the account. It's my friend who made the account for me. And I didn't normally use it for like a few months until I started getting used to it" (P21). These sorts of legacy accounts highlight how one's social circle influences account creation.

While motivations for entertainment, relief for boredom and access to friends/peers were to be expected, school, surprisingly, was also an underlying force behind why some participants used social platforms. For example, a participant talked about how one of the few non-restricted websites that she enjoyed when she was logged into her school Google account was YouTube (YouTube is a subsidiary of Google). She said, "I have my own account, but it's a school account. I can still watch YouTube on it, but like, I can't do anything else with it. And I also, but I also have my mom's account, but which I can do whatever I want on it until whenever I sleep" (P14). Likewise, the types of apps that P13 downloaded on her first phone were greatly influenced by the apps she was exposed to on her school computer. "Well, first my sister had TikTok and I kind of watched it with her, and then a couple, like, I don't know exactly when, but then I got a phone and I kind of just got what I did on my computer because I had a school computer and I basically did every single app that I had on there and I signed in with my same account and then I put it on my phone" (P13). P20 described how he used YouTube for "just like video games or just like gaming videos or anything like that, or anything that helped me in school, too" (P20). This group of mostly middle and high school youth are part of a generation that were among the first users of the learning management systems like Google Classroom and online-based learning. The COVID-19 pandemic likely accelerated this process.

4.1.2 Attitudes Towards Kids' Platforms.

There is a debate among U.S. lawmakers and tech companies about the design of child-oriented social platforms–and whether they should be built at all. We have already seen how child-oriented platforms have evolved. In August 2021, YouTube turned off autoplay by default on YouTube Kids, YouTube's platform specially designed for younger audiences. The design of recommender systems in particular seems like a salient area of focus. This is not surprising, considering an analysis of YouTube's design mechanisms found that recommendations, ads, and autoplay were the three features that made users feel less in control of their experience [43].

In this study, the younger participants were asked whether they were familiar with child-oriented social platforms like YouTube Kids or Messenger Kids. Overall, there were mixed attitudes towards having a separate version of their favorite social platform but geared towards children. The participants in this age group mostly referred to their younger siblings who used it, while they had moved on to the main YouTube site. P14 described how music was an important part of her life, and the music that she liked to listen to (e.g., rap music) would likely not be found on the kids' version (P14). She continued to say that she felt like it would not make a difference if there were two versions of the app because kids could be bullied on the kids' version. P16 also preferred the main YouTube platform over YouTube Kids because she felt like YouTube Kids mostly contained kids younger than her (P16). The main YouTube platform had all the people and channels she liked. However, she only watched subscriber videos of kids her age. Although he had not used it himself, P19 felt like it could be beneficial for younger users, saying, "it's safer because they're younger and because they want to do more stuff really...[o]nce they hit five or six, they just want to be independent" (P19).

4.2 Awareness of Data Collected

RQ1: To what extent are youth aware of online advertisers and social platforms collecting data about them?

Participants' responses to the question of what data, if any, they believed online advertisers and social platforms had on them varied. Participants of all ages most frequently cited data related to account information, such as their name, email, and age. However, they also discussed less intuitive forms of data collection, like data about their tastes/preferences/personality and data collected through user interactions. Some participants made inferences based on what they had noticed was recommended to them.

4.2.1 Personality, What they "Liked". Participants' answers suggested that they believed that the social platforms they used and/or advertisers knew something about their personality. This was before we had conducted the identity profiling activity that involved discussing personality profiling. Personality was central to P13's response regarding YouTube and TikTok:

> "I think it kind of knows a little bit [about] my personality...I think it would like [to] know that I'm on YouTube, I like watching a lot of Among Us videos, and then Roblox videos, and then just funny videos, and it would not show me like as much of, like, the other stuff... And then on TikTok I think it just kind of comes up with things that they feel like I would really like, which it's been right a lot, but not fully all of it" (P13).

Similarly, when asked about what he thinks Google knows about him, one young participant's response suggested that the search engine knew about his personality: "A nerd. Because I search up nerd topics. And it's like, and it's like, 'Okay, this kid is a smart guy. He searches up topics that nobody cares about. He must be so smart' " [chuckles] (P8).

Not too far off from the topic of personality, four participants (P1, P5, P17, P18) explained that social platforms or online advertisers knew 'what they liked.' For example, P5 said that online advertisers knew "what she liked" based on what ads in gaming apps she clicked on (P5).

P21's answer showed how he relied on the interface cues in YouTube to communicate what type of data the platform had on him. To P21, a specific feature in his feed was especially telling.

> "Probably, I think that's how the recommended videos come up anyway. And I think when you go on YouTube, there's like a, like a, like a bar on the top that shows, you know, it'll say like music or like, what else would you say, maybe like gaming or sports...Yes, or based on whatever you've been watching recently. So, and I don't think they'd be able to have any of that without collecting any data from what you're watching or anything about you in general" (P21).

HCI research over the years has increasingly focused on how signifiers in a system help the user construct a mental model, even if it is imperfect [30]. The "bar" he is referring to at the top of the YouTube homepage contains a series of constantly updated meta tags of music genres, artists, and types that the user can select to filter content. His experience echoes the work of Karakayali et al. who refer to recommender systems as "intimate experts" that not only guide user preferences, but play a role in shaping taste and guiding self-discovery [34]. Interestingly, when asked in a follow-up question on what sort of data YouTube had on him, P21 replied that he had no idea, even though his comment suggests otherwise.

Along with P21, P20 observed how the platform made his interests and/or likes explicit. When P20 was asked what data, if any, Instagram had on him he said, "Yes. But it's not as bad as YouTube." When asked to explain, he replied, "On YouTube, when they– I guess the information or whatever–they make you want to buy something. On Instagram, it's just they're trying to kind of help you to like see your interests. Yeah. And your needs" (P20). When asked to elaborate on any specific types of data that Instagram had on him, he responded, "just your interests, really." Although he did not specifically mention ads in this comment, he did talk about how the platform "make[s] you want to buy something"; in reality, both YouTube and Instagram partner with advertisers. His comment

signified a need to further compare how users imagine any differences in data collection between companies.

As another example of cross-platform comparison, recall P21's observation of YouTube. When P21 was later asked about what data, if any, TikTok had on him, he replied,

> "Same thing as what I said about YouTube is, you know, I wouldn't–There's no way they constantly show all these videos that appeal to my humor if they didn't collect this data, especially since some of these videos are similar or they have some type of like continuous joke. So, yeah, it's hard to think that they're not collecting data. So, but it's not like they're collecting probably personal data, it's probably more so data on what I see is funny or interesting than what you've, what you see on the app" (P21).

We can see from this quote that P21 is picking up on clues from what he is seeing in his feed, like a consistent stream of humor-filled videos based on a similar joke. He is clearly aware that the platform is collecting data, but he does not associate it with "personal data." This resonates with themes in Bowler et al's work, like what the term data encapsulates and why it matters [11]. Our findings prompt us to go one step further and question when data becomes 'personal.'

How do the platforms learn what P21 finds funny or interesting? As we will see in the next section, many participants were attuned to the user interactions they made on social platforms and with advertisements.

4.2.2 User Interactions & Passive Data Collection. When explaining what data platforms and advertisers had on them, participants frequently entered into a discussion about how the data was captured. They mentioned user interactions like clicks, time viewing the video, or search history. Their responses indicated that they had a reasonable understanding of how the user interactions impact their recommendations. According to P6,"I think they collect information by seeing if I skip the ad really quickly or watch it through. And if I use the link in the ad, and buy the thing, or just look at the link because I think it just depends on what the person–how the person kind of interacts with the ad" (P6).

P11 framed his response in terms of what the YouTube video creator could see, specifically for which videos he pressed the Like button. In contrast to other participants, though, he did not imagine evidence of his musical tastes being curated in aggregate.

> "I think they definitely like can see when I will like a video, like I press the Like button and they can, they can like, well the YouTuber can see, and I'm pretty sure YouTube also can see when I like a video. But in that case, really I don't think, like, other than that I don't think that they know what I like or don't like. Well, they couldn't because they might like see when I click on a video like what type of video I like because like by the own– by the YouTubers, they could like possibly see, on which video I click on and they'll be like, 'Huh, so he likes this YouTuber a lot.' So that's why I think so, but in other cases, I don't, I don't think they know what I like because I've watched a lot of different things. And, yeah, and they don't, they don't personally know me, so they don't know what I like or not" (P11).

In reality, the creator of the YouTube video cannot see who liked or disliked their video; that information is anonymous. His comment, "they don't personally know me," demonstrates how interpersonal privacy is most salient to him.

P18's direct response to the question about what data Instagram had on her was centered on interactive ads. "It's kind of hard to say with Instagram because like the only type of interactive stuff that I get a lot and I don't know how to explain it but, like, the ads that I said in between stories or in

between posts, but everything else like I pretty much know people that I'm following or like celebrities or whoever" (P18). Data collection to her was tied to a "type of interactive stuff."

As an example of passive data collection, two older participants speculated that TikTok collected geographic data. Interestingly, both of them said that although TikTok claims that they do not collect geographic data, their experience with the app has made them suspect differently. When asked about what TikTok knew about her, P18 said, "So I think obviously like who we follow matters, like videos from them, and similar content and like the type of videos that you like. So, if you like astrology, it probably honestly knows a lot. And, or the tarot readings–I kind of just scroll through them, so I stopped seeing them, but I think it tracks all those things, but as well as like location. I know it doesn't, but I kind of think it does because, like, wherever I go, it shows me Best Places to Visit near this like, you know, so I think it tracks that" (P18). Similarly, P9 said that even though TikTok says that they do not collect geographic information, she sensed that it knew her location because many videos in her feed referenced [her state] (P9). In its privacy policy, TikTok states that it does in fact collect location data. Their experiences with getting mixed signals, though, emphasize the need for more clarity about data collection.

P18's comment also provided evidence that she was aware of how time viewing TikTok videos was a form of feedback that influenced what she saw in her feed. She observed, "And, or the tarot readings–I kind of just scroll through them, so I stopped seeing them." The default of watching a video through serves as positive reinforcement that you like the video and want more videos like it. The opposite, what she did, was to scroll through them.

When the question was oriented toward online advertisers, again, a common theme that emerged was how they interacted with the ads, and whether they gave the data explicitly or implicitly. Three participants (P5, P14, P18) referred to interactive ads, i.e., ads designed to elicit user input, like the survey that appears at the beginning of a YouTube video or the embedded survey on the YouTube homepage. P14 declared that yes, online advertisers did have data on her because "yes, because, well, sometimes when you scroll down, it's actually questions like, 'Is this video above, like, a good fit for you?' or something, and then you can hit a check mark or dislike or something until you see something interesting" (P14). Oftentimes, participants directed the conversation from talking about online advertisers in general to talking about data collected through ads. Like P14, P18 also discussed interactive ads. "I couldn't answer that because, like I skipped the ads. Like I honestly don't interact with them. I know there are a lot of interactive ads, but I don't really do that, any of that, even when you're using sources like Brainly or something like that" (P18). As previously mentioned, P5 reported playing a lot of gaming apps–and seeing a lot of ads in those apps. She said that the apps know what she likes based on if she clicked on one of the ads that appear during the games she plays (P5).

Besides viewing and clicking on ads, multiple participants talked about search history as evidence that data was collected about them. For example, one participant sensed certain product pages she viewed were tracked, so she came up with a strategy to circumvent unwanted ads. "I definitely have used [the incognito browser] for shopping and stuff because like if it's something I don't want an ad for like, I'm buying one t-shirt, I don't need hundreds of ads for t-shirt like sometimes I'll go there and just order my one t-shirt because then it'll be that I get way less ads as opposed to just if I did it my normal, normal Google browser" (P9). P9's experience was not unique. P12 shared a similar experience of searching for a bike and then moments later seeing an ad for a bike. Moreover, P2 had noticed ads mimic what she was searching. "I feel like it's um kind of...weird. Cause sometimes they do like what you were searching. And that's just kind of weird to like get ads to exactly what you're searching" (P2).

P4 also mentioned her search history, as well as her emails, websites she visited and videos that she watched:

> "It's just maybe based on my search history or something that were in such emails or something that some pages that I go to mostly, so they may be gathered that are based on like 'Okay, this person goes to this website a lot.' Or maybe if I watch a particular type of the videos more than others which I may not understand ... maybe they are collecting the data based on the like okay, 'This person is on this website a lot and they're watching these things a lot'" (P4).

Taken as a whole, their comments suggest that many participants were cognizant of captivation metrics, also known as measures of user retention, which we will discuss later in the paper.

4.2.3 Uncertainty of Data Tracked. Nearly two-thirds of participants (n=13) explicitly expressed uncertainty about what data was tracked or how data was tracked. In general, participants had a more challenging time articulating what data advertisers collected. Three participants were not aware that any of their data was tracked by online advertisers, but not necessarily social platforms. These participants were typically younger–between 10 to 12 years old. For example, when asked if he thought online advertisers had any data about him, P11 replied, "No. They're just, they're just trying to get me to buy Verizon mobile or T-mobile or Google phone, like a Google Fi phone plan or something like that, just to get more money and then who knows where it's going to go after that" (P11). The younger youth interviewed also tended to have a limited awareness of how data is exchanged between websites, apps, and advertisers; the participants saw their data as contained in the app in isolation.

Meanwhile, regarding social platforms, there were other participants who did believe data was collected about them, but they were unsure what that data was exactly. Participants who expressed much uncertainty about what data was collected also tended to refer to limitations as to what data they visibly input into their accounts, such as their age or email address. P12's response to what YouTube knew about him overlapped with what TikTok knew about him. "I think they probably collect your age, and like I don't know what it could collect. Cause they can only like see what you put in your like account or whatever. So like age. And like I feel the same thing about TikTok, like whatever you like, like similar stuff on your ForYou page" (P12). P14 also reported on account information. "Yeah, I don't know because I'm like, um the only thing that they will probably know's my email cause like you have to type that in" (P14).

P21 did believe that YouTube, TikTok, Instagram, and online advertisers collected data about him; however, he said that he had no idea about what that data might be (P21). He expressed confusion about how data was transferred between platforms.

> "I don't know how the advertisers do it, again, there's a joke saying that there's, you know someone in our phones listening to us, or even on, you know, Alexa, like. So yeah, probably, I just, again, I have no idea how they would do it. Especially because I don't necessarily watch videos or anything about things I was trying to buy. So, yeah. Yeah. And I'll be on YouTube and sometimes on my phone, under a video there's like, like a list that you can scroll through, and it's a bunch of stuff. And sometimes it'll be like the same shoe I was looking at or the same kind of shirts I'll be looking at. So I don't know how that happens because I'll be looking at Safari and [it] transfers to YouTube" (P21).

P21's reflection reveals that he senses, or at least jokes about sensing, how he feels he is surveilled by advertisers. The impact of what it means to feel constantly "followed"–followed through different websites or different devices (Alexa and the phone)–will be explored later in the paper. His comment also highlights the barriers in trying to map out the distributed, pervasive nature of online advertising.

P21 continued to talk about how he viewed targeted advertising as both helpful and scary. Note that he said this before we did the identity mapping activity on the useful/creepy board later in the interview.

> "Advertising–I think it's helpful because since it's targeted, more of the ad, more the advertisements will be based off of things you're interested in or things you're looking for, um, yeah so I think it can be helpful, but maybe at some point it can be scary, because I'm pretty sure most people have had this experience where you know they'll be talking about something, and five seconds later they'll see an ad pop up about it, and you'll be thinking, 'There's someone in my phone listening to me or whatever,' so it can be scary. I'm not lying about that. But the pros, you know, it can be helpful, directing you to things that you could be interested in" (P21).

Prior work has explored online advertising as both "helpful" and "scary." For example, P21's sentiment echoes the results of Ur et al.'s study on OBA that found users' perceptions of them being either useful, smart, or creepy depended on the situation and the company [68]. However, a major difference from the 2012 study was that many participants in this study were indeed aware of search history and browser history as influential factors in the ads they saw, as described earlier in this paper.

P1 also reported a "telepathic" experience with ads. When asked how he felt about online advertisers he said,

> "Just...they reap the stuff you like..I don't know..I guess it's on your phone and stuff...stuff you look up. I think it's fine. I'm like. But I don't like how they like know...like one time it was weird. Like one time I spoke about something, it just popped up. And I was like 'what?!' That is weird...It was like one time I was like cooking cause I was like mad hungry and there was like food popping up on ads and I was like 'What?' That's crazy" (P1).

P1's response conveyed a range of sentiments. On the one hand, he started off with a pretty indifferent attitude towards advertisers. However, the tone of his comment also demonstrated a language of surveillance through his comment–the advertisers "reap the stuff you like"–and also through his reflection on feeling followed. Of course it is dificult to know whether the ads appeared out of coincidence or whether he had searched for a similar topic a long time ago. What perhaps adds an additionally layer of mystery is when the algorithms are used to predict internal states–or at least the perception that algorithms are used to predict internal states.

Another participant explained a memorable ad he saw of something he had searched for, except it appeared in an unexpected place: his inbox. "I was searching out to buy a video game or something. And then I got an email, like a video game being on sale for something" (P20).

Several participants (P15, P18, P20) used their assumptions about social platforms to draw parallels about how they imagined what data online advertisers possessed and how. As P15 reflected on advertisers' data collection practices, he compared it to those of YouTube."I guess it's probably like the same thing as YouTube like your search history, or something like that because like sometimes I'll be like thinking of something and then I search it out, like an hour later I see an ad about like that thing" (P15). He then recounted a memorable experience, "Well yeah, like, like I think it was two days ago I was thinking about buying these sunglasses. And then I went on YouTube and then there's just a bunch of like sunglasses try on like channels. They're just all over my page. It's kind of weird. I mean, to an extent, I think it's a good thing, because it like personalizes the things you see, but I don't exactly know the limitations of like the data they collect from you."

4.3 Google Ad Settings Activity

RQ2: How do youth react to a company's ad profile of them?

4.3.1 Indifference and Appreciation Towards Targeted Ads. Participants talked about memorable experiences with ads, their opinions of targeted advertising, and their reactions to seeing their ad dashboard, Google Settings. Over half of the participants, regardless of age, were either indifferent towards targeted ads and/or expressed appreciation for the benefits that came from receiving targeted ads, namely products that were useful and interesting to them. For example, at the end of this activity, participants were asked whether seeing this information influenced how they felt about the company that made the Ad Settings dashboard (Google). Did it make them trust the company more, less, or the same–no change? All but two participants' responses were "no change." When they did give a rationale why, participants said something along the lines of P1's response, who said, "I think the same. They're not doing anything wrong. They just want to get ads to you that are for you. There's nothing wrong with that" (P1). These results mirror those of Pangrazio and Selwyn (2018), who found that teens interpreted targeted advertisements as a necessary part of search and social media [55].

When asked if there was anything that they saw on the Google Ad settings page that they would change–either add new categories, take away categories, change the specificity, or even opt-out of personalized advertising–again, all but two participants who did the activity said that they would not change anything. Targeted ads to them were at best, something that could occasionally deliver information about useful products, and at worst, an annoyance. These findings mirror Farke et al.'s study, which exposed users to Google's My Activity [25]. Only 25 percent of the 153 participants reported that they would change any settings in My Activity or their behavior. They also tended to view the data collection more beneficially after looking at their My Activity.

P12 reflected on the information in his own profile, which he presumed was a result of his search history. "I don't think there's like really damage in access to my information. Like, they just like know my age, like, they just like know what I look up. I don't even like remember looking most of this stuff up that's on there" (P12). First, his reaction suggests some confusion about data retention or how some of the inferences were linked to him: "I don't even like remember looking most of this stuff up that's on there." Secondly, his comment about Google Ad Settings knowing his age is interesting because the estimated age category for every single participant that reported it on their page was inaccurate. Most created their accounts long before they turned 13 and many put their age as much older. The inconsistency between their sign up "age" which they may have forgotten and their actual age might explain why they were surprised to see ad inference categories for Life Insurance and Housing. P21, along with two other participants, felt like there would be much more at stake if they had a job or financial information online to be concerned about.

For this activity, participants were only asked to share the profile categories that they were comfortable with sharing. In terms of what stood out to participants, four expressed surprise to see profile categories related to searches they conducted and activities they had done a long time ago. Some participants were surprised to see highly specific products or companies. P20's reaction changed from confusion as to why he was seeing topics that were wildly off to surprise of how specific categories matched his interests. After looking through the Ad Settings page, he said, "So my first impression of it is it looks really off. And it wasn't that accurate, but when I went more down there were things I saw that only I do" (P20). Specifically, he noted he was surprised to see the Web Design and Development category because he took a web design class the year before, which seemed to him like a long time ago.

4.3.2 Concerns about Targeted Advertising: Identification and Distraction. While most participants expressed indifference or appreciation, there were a couple of exceptions and caveats. For instance, P4 simultaneously expressed concern and appreciation for the 'targeted' part of online targeted advertising.

> "Yeah, being like an immigrant and coming from a different culture and country, I think having targeted advertising is not something that I'm really happy with or something like I would want. I would like some things like advertising, a particular, not a particular people but the whole people in general, which, which I think I feel more comfortable with that, instead of like, because if there is something that I don't really use it or I don't really need it. I would rather not watch that advertisement" (P4).

Her concern appears to stem from the connotation that targeted advertisements make her identifiable. She continued on to say that she wished targeted advertisements were more general.

One exception case in which a participant expressed vehement dislike towards ads was P8, who self-reported he had ADHD. He did not like targeted advertisements because he found them extremely distracting."Like sometimes when I'm like, not on YouTube, I'm like reading Wikipedia and it's like video ads for horror movies or something, and it's like–it's like I can't pay attention. Yeah. It's like screwing up my life" (P8). Earlier in the conversation, he said that he tried not to pay attention to advertisements when he sees them. While Wikipedia does not host advertisements and his memory was presumably of another site, his overarching concerns about ad distractions are salient. His comment emphasizes the urgency of studying the effects of "advergames," [4] gamified advertisements that stealthily draw users' attention, especially the attention of children.

4.4 Identity Attribute Mapping Activity

RQ3: How do youth consider trade-offs between usefulness and creepiness in profiling?

The final segment of the interview consisted of the Identity Attribute Mapping Activity where participants mapped how useful and/or creepy they found it was to be categorized according to X attribute (e.g, age, gender).

4.4.1 Divergent Views About Personality Used for Profiling. Participants had especially divergent views on the personality attribute. They had conflicted feelings of it being simultaneously both useful for formulating recommended ads or content, but also potentially creepy depending on who had access to that information.

In marketing, the targeting of people based on personality is referred to as psychographics. Data brokers may create psychographic profiles, which are profiles that consist of a user's presumed attitudes, values, and activities, interests, and opinions. Some scholars argue about the nature of consent in psychographic targeting practices when algorithmic inputs are indeterminable and data uses are unpredictable [50].

In terms of usefulness, two participants distinguished personality from the other characteristics for being more accurate in matching them with relevant information and ads. For example, P15 mapped personality as extremely useful and not at all creepy. "I think personality would like obviously be the biggest factor in deciding what you would like...because like gender, race, and age are just like very general, like they can apply to a lot of people, but as far as personality goes, I think it's more personalized" (P15). He continued, "Like, I feel like they're [age, gender, race, religion, political afiliation] all like aspects of your personality. So, like the only way to be truly useful is if they were like combined, so that's my personality so far."

P12 shared a similar sentiment as P15, but in his reflection implicitly charged personality as a loaded term that encompassed interests and behavior. He commented, "I think it's like really useful,

because I think, I, like, what your personality is like what interests you and like how you act and stuff like like you like easier to find out what you want to watch or what ads are important, like ads like make you want to buy something" (P12).

In contrast, P20 felt like personality was not that useful or creepy for the very reason that it could be volatile. "Oh personality. Oh, that's kind of hard to understand really because people could have multiple personalities, or they can always change" (P20).

A couple of participants hedged their appreciation for the personality attribute by qualifying that their curated feeds were so accurate that they were also a little creepy. For example, P6 commented, "Yeah, so with personality, I think it's extremely useful, because, like, a lot of the things on my page are like exactly my personality, so it's a little creepy, that they kind of know. And, yeah, I think it's just all the videos on my page, again, like, are just things that I'm interested in" (P6). P16 also said that she had the impression that a lot of videos in her TikTok feed were relevant to her personality. When asked if she could give an example, she said, "Well, there's just, again, a lot of fashion videos, and I'm kind of obsessed with fashion, and a lot of girls my age, talking about just things in general. So, yeah, it's a little creepy though" (P16).

One of the participants argued that it was creepy for online advertisers to make assumptions about users, especially their personality: "Um, it's just for like ads, they shouldn't be assuming what people are...personality– it would be like creepy to like assume someone's personality when you don't like actually know them" (P2). P2 mapped personality as the second most creepy attribute, just behind political afiliation.

In contrast, when P10 was asked why he mapped personality as not at all creepy, he suggested he did not find it creepy because he had nothing to hide: "because like I'm a nice guy" (P10).

When talking about the personality attribute P21 said, "I'm not sure how it would be useful. It's more like a feeling it could be. But I think it is a bit creepy though, thinking that advertisers may have an idea of who I am as a person" (P21).

When P5 was asked why she placed personality in the top left as extremely useful and not at all creepy she replied, "Because it doesn't really tell you much about myself. Um, just my personality. It doesn't tell them my age, weight, or my, of my medical records or anything like that" (P5).

4.4.2 Age-Based Ads Were Perceived Useful. Participants felt that age was a useful characteristic because the content they liked would not be applicable to people of a different age. P9 claimed age in the context of Instagram would be "Definitely be more on the useful side than non useful because, as much as like you always lie about your age for things and stuff, but based on like the current like you don't really want, if you're like, 15 you're really not interested in seeing stuff that 13-year-olds post all the time. You really want to see more kids your age, so I definitely feel like that can be more useful" (P9). P12 and P4 shared a similar view of age. "It's pretty useful because like, then they know, like, as like they're going to know like someone who's like 13 likes different stuff and someone who's like, 30, and then they'll know like like that I think it is pretty useful," P12 said for YouTube, Instagram, and TikTok (P12). According to P4, with regards to content feeds in YouTube and Instagram, "If I'm 17, I'll be more interested in things like that most 17-year- olds will do, instead of something that, like, a 30-year-old couple or a 30-year-old person will do, like, just, I think it's just different interests based on your age" (P4).

For P3, her evaluation depended on whether that information, age, could be linked to other information about her, like her gender. She said YouTube estimating her gender would be:

> "Useful, extremely useful, extremely useful, but also extremely creepy... If they have your age and your gender, it would be more narrow[ed] down to see who you are and where you live. So I feel like that's more creepy. But, age, there's a ton of 10-year-olds in the world" (P3).

What she is describing is the impact of how data is combined, processed, and aggregated. This specific scenario exemplifies data aggregation in Daniel Solove's Taxonomy of Privacy [64].

> "Aggregation is the gathering together of information about a person. A piece of information here and there is not very telling. But when combined together, bits and pieces of data begin to form a portrait of the person. The whole becomes greater than the parts. This occurs because combining information creates synergies. When analyzed, aggregated information can reveal new facts about a person that she did not expect about her when the original, isolated data was collected." (p. 507)

4.4.3 Attitudes Towards Socioeconomic Status Depends on Scope of Data Use. Several participants mentioned that the level of creepiness that they assigned to an identity category would depend on who would have access to that information (P5, P6, P8, P9). This most commonly came up when they were discussing the characteristics of mental/physical disability, sexual orientation, and socioeconomic status.

Most of the older participants reasoned that collecting information about users' socioeconomic status could be beneficial to advertisers. When talking about social platforms, though, they tended to have mixed reactions. P6 felt somewhat ambivalent about socioeconomic status with regards to TikTok, seeing it as "not that useful and not that creepy." But in her explanation for why she mapped the characteristics the way she did, she discussed how homogeneous the people in her feed seemed to be, just like how she saw few non-White people in her feed. "Some people on my ForYou page are middle class and some are on the wealthier side. And I haven't seen many like people on the poor side, because most people that are on the ForYou page are famous and famous tech stars make money" (P6).

P10, who declared himself to be cybersecurity conscious, argued that having that information about a person would be extremely creepy, especially if it got into the wrong hands. In describing YouTube, he said, "The social economic status–I find it extremely creepy. Because I'm like, maybe like a bad guy who hacked into the YouTube framework, and like, if I was older. My job, and life, stuff like that." When asked how they would know that, he replied, "Um, it could be, they could hack in and look at my settings and stuff. Oh, that's one of the biggest problems with the internet, you may get a hacker, cyber hacker or something" (P8). His description exemplifies how he associates privacy risks with a person on the other end, rather than the distributed nature of ad tracking.

5 DISCUSSION

5.1 Comparison with Actual Data Collected as Stated in Privacy Policies

One of the aims of this paper is to compare youths' expectations and understandings of what data companies collect about them versus what data they actually collect. In other words, how do their expectations align or misalign with reality? To find out what data companies collect–or report they collect—we turn to the privacy policies of the three main platforms we discussed in this paper, YouTube, TikTok, and Instagram.

What was clear to participants was that the platform retained account information like their name, email, age, and content they uploaded. Additionally, one unexpected finding from this study was the number of participants who were cognizant of passive data collection from social platforms and advertisers. Passive data collection is data collection in which information is gathered automatically—often without the user's knowledge—as the user navigates from page to page on a website. It contrasts with active data collection, which is when the user needs to actively provide the data [24]. While that is the definition of passive data, in reality, this study suggests that there seems to be certain kinds of passive data that enters into the user's awareness. This diverges from past work that showed that participants were largely unaware of their search history and browser history being used in ad targeting [68]. Participants in our study frequently referred to interaction

metrics like time viewing a video before they swiped or watching the same video repeatedly as key information that companies collect on them. Likewise, in terms of advertisements, they frequently mentioned things they had searched for or whether they clicked on an ad in the past. Anthropologist Nick Seaver calls this type of log data captivation metrics because they capture how a particular user interacts with a system over time [62]. Here, we see participants' awareness that their captivation is measured. Participants also talked about specific user interactions. When P18 talked about the data she believed Instagram had on her, she immediately referred to a "type of interactive stuff"– interactive ads. The participants' understanding seemed to be pretty perceptive. According to Google's Privacy Policy, they collect information about views and interactions with ads, for example, whether the user mouses over the ad or interacts with other parts of the web page.[9] One hypothesis for why youth are attuned to captivation metrics is because of the surge in popular apps that are built for rapid feedback and meta-awareness of behavior. For example, the microinteractions in TikTok. Users swipe in TikTok to make a decision about whether the video attracts them, with the default of view time being a positive signal. Furthermore, features like Apple's ScreenTime make users' time spent on apps salient. Several participants in the study were content creators, which may have motivated them to think more about interaction metrics.

What was less clear to participants was whether geolocation data was collected and how long data was stored. Recall P18 and P9's mixed signals about whether TikTok collected their geolocation data. They said that TikTok claimed it did not collect geolocation data, but they speculated it did based on what they saw in their feed. According to their privacy policy, TikTok does in fact collect "geolocation-related data" such as the user's approximate location, IP address and with the user's permission, GPS. They report they use this for ad recommendations, among other uses. Regarding data retention, recall P20's surprise to see inferences about the web design class he took a long time ago or P12's saying he did not remember looking up many of the topics reflected in the ad settings inferences.

What was entirely absent from participants' responses was any mention of whether the company collected their social contacts' information (e.g., the contacts in their phone) when they created their accounts. As detailed earlier in the literature review, people tend to find the idea of their online profiles being linked to the people in their network creepy. However, the privacy consequences of allowing certain features may not always be clear to users, let alone what those consequences would be for other people. TikTok and Instagram state in their privacy policy that they collect this data, but only with users' permission. This seems like a pressing area to address, as newer social platforms find ways to entice users to give their contact lists. When the audio-chat app Clubhouse first launched, it would only allow users to send invites if they granted the site access to their contact list. Since then, Clubhouse has tweaked its invitation process, but it serves as a reminder of the networked nature of privacy online [46].

5.2 When is Personalization Resisted?

A recurring theme that emerged from interviews was participants' confusion about how data was transferred between websites or between devices. Recall P21's surprise at seeing product ads for a shirt he was looking at through Safari later appear on YouTube. This sensation of being followed was also evident in how multiple participants shared 'telepathic' experiences with ads for a product that appeared just as they were thinking about it: P15 recounting how he was confused to see sunglasses he had previously searched on Google later appear on YouTube or P1 describing how timely ads for food appeared when he was hungry. P20 was not sure what to make of something he had searched for appear as an ad in an email. Multiple participants used words like "weird" and

[9] https://policies.google.com/privacyfootnote-content-views

"scary" when talking specifically about what data they presumed was collected about them. This contrasted with how they talked about targeted advertising more broadly; as described earlier in the results, many participants carried an indifferent attitude or an appreciative attitude. Similar to Leon et al., a major motivation for participants in our study to grant advertisers to access their data was that it would give them more useful, relevant ads [40].

The unpredictability of a technology has been cited as a key factor of creepiness [75, 77], as is lack of control [77]. Algorithmically-mediated ad systems are extremely opaque–in what data is collected, inferred, and how it is shared–and consequently extremely unpredictable. We saw this unpredictability reflected in interviews with participants, as one participant remarked, "I mean, to an extent, I think it's a good thing, because it like personalizes the things you see, but I don't exactly know the limitations of like the data they collect from you" (P15). Granted, Wozniak et al. make it clear that creepiness is distinct from acceptability; users are willing to accept the technology, despite the creepiness [75].

A larger question related to this sense of being followed is under what conditions do people accept or resist personalization? P15 reported using an incognito browser, but only to search for products because she did not want to repeatedly see advertisements for the same shirt. Past work has shown how context matters in people's willingness to share information for advertising purposes [71]. Besides the sensitivity of the information, people care about how it will be used, for example, if it is shared with other websites or combined with personally identifiable information [40]. For P15, e-commerce, or product searches, was a domain context that she avoided personalization for ads. Furthermore, we can consider the role of signals in the interface that communicate that the user is being tracked. P21 was pleased to see the YouTube meta tags that describe the categories that go into his recommended videos. For many online algorithmic systems, some level of transparency and explainability are necessary before contestability can be achieved, i.e., the ways a user can provide feedback to the system or resist its impacts.

5.3 Design Recommendations

Many participants in this study expressed confusion about the types of data that social media and advertisers collected, but were also uncertain about who would have access to it. Participants frequently discussed that the "creepiness" of a given attribute in the Identity Mapping activity would depend on who would have access to it. TikTok's description of how they share information with advertisers in their privacy policy is also vague: "Advertising, marketing, and analytics vendors, who may receive the information you choose to provide, the information we obtain from other sources, and the information we collect automatically but would not receive your payment information or message data."[10]

Im et al.'s afirmative consent framework provides some direction for how algorithmic advertising systems can be more transparent [33]. They argue that the five principles of afirmative consent–consent as voluntary, informed, revertible, specific, and unburdensome, can be applied to sociotechnical systems. One way to design for the user to be informed is to consider their awareness of the audience. The Audience Intel block of the framework states that "Systems provide feedback as soon as the real audience diverges from the likely imagined audience." An imagined advertising system built on afirmative consent would make the identity and data uses of advertisers or any third-party trackers explicit. Barbosa et al's study shows how algorithms can be used to determine the opportunistic moments to intervene to give the user a chance to review how their online activity may influence their ad targeting profile [8]. For example, it could show how visiting this site could influence the (sensitive) inferences associated with the person. The downside of this approach is

[10]https://www.tiktok.com/legal/privacy-policy-us?lang=en

that it could lead to information overload and/or digital resignation for the user. To mitigate this, users could be provided with both local and global controls to customize what the algorithmic system monitors [3].

Companies have made progress in updating their policies to reflect the needs of children and teens. For example, Google has turned off location history for all users younger than 18 and restricts their ability to turn it back on. Videos uploaded to YouTube by users age 13-17 are private by default. Likewise, the accounts of users age 13-15 on TikTok are set to private by default and only approved followers are able to view the users' content. Instagram said that accounts created by children under 16 will be private by default. What protections do the companies have for commercial data collection? Instagram and Facebook will now use age, gender, and location as the only parameters to target users under the age of 18. They previously used people's interests and activities across other websites. TikTok has a feature called Family Pairing which allows parents and teens to customize the child's safety settings. TikTok has a separate privacy policy for users under 13 and collects account registration information like birthday and username as well as how they interact with the app. They use the information that is collected automatically to provide contextual advertising. The assumption that the companies are making, though, is that users truthfully give their birth date or age when they register for an account. As prior work and our study point out, this is rarely the case. Recently, scholars and policymakers have debated, and sometimes argued strongly for, the use of age verification as a means to ensure children's safety online. While there are some technical challenges (is this at the device level, OS level, browser level, what happens with shared devices), the predominant concerns are social and political. Tracking of children has been misused to harm children, especially those at the margins, in systems like education, policing, child welfare, and immigration, and introducing strict, close tracking of children online could introduce the same outcomes, even if unintentionally. One option may be to verify children's ages then delete the verification information, and require that companies show evidence of deletion, to minimize the risk of misuse of that data for other purposes. In general, we recommend that there be greater oversight and accountability for targeted advertising online for all users–adults and children–and that children may additionally benefit from special protections where possible.

5.4 LIMITATIONS AND FUTURE WORK

This was a qualitative project involving a sample size of 21 participants who do not represent the diversity of youth in the U.S. or globally. Further research is needed to generalize these results. This study documented the expectations and reactions of participants towards data collection and processing, rather than their behavior. One next step for a study could be to focus on youth behavior with regards to commercial privacy that could explore what influences whether a child is likely to opt-in/opt-out to third-party trackers when they download an app through the app store. Likewise, an experimental study could explore what factors influence whether a child allows a site access to their social contacts list or address list on their phone, taking into account dark patterns. Finally, as companies continually update and expand their data collection practices, a future study could ask how revealing users find each algorithmically-mediated data stream, in a similar vein to how Merrill et al.'s study explored how revealing people found various biosensor technology [51]. For example, TikTok collects biometrics like faceprints and voiceprints.

Another limitation of the study design was that the interview format may have contributed to social desirability bias. Participants may not have been as candid speaking to another person. Ultimately, we determined that interviews were the most appropriate method for this age group as opposed to surveys because it allowed the participants to ask follow-up and clarification questions.

6 CONCLUSION

In this paper, we investigated the gap between what data youth believe social media companies and advertisers collect about them (and what they do with it) and what data they actually collect. We also analyzed youths' attitudes towards online targeted advertising to understand how that shapes their perception of data collection. We find that youth are aware of certain kinds of data collection based on user interaction and passive data collection, but are unaware or are uncertain about other kinds of data collection practices like data retention or geolocation detection. People, including younger users, use different accounts for different life stages and purposes and their advertising experiences can be mutually constituted with those purposes. We advocate for greater visibility and accountability into how platforms shape youths' online advertising experiences.

7 ACKNOWLEDGEMENTS

We thank the youth who volunteered their time to share their experiences with us for this study. Thank you to the reviewers for their insightful comments and suggestions. We also thank Drs. Florian Schaub, Gabi Marcu, and Nazanin Andalibi for providing feedback on earlier drafts of this work. We thank the friends and colleagues who provided feedback that improved this work, including Ben Zhang, Yixin Zou, Rahaf Alharbi, Dr. Carol Scott, Yuanye Ma, Lia Follet, and members of the UMSI community.

This material is based upon work supported by the National Science Foundation under Grants #1763297 and 1552503.

REFERENCES

[1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2020. Secrets and likes: the drive for privacy and the difficulty of achieving it in the digital age. Journal of Consumer Psychology 30, 4 (2020), 736–758.

[2] Lalit Agarwal, Nisheeth Shrivastava, Sharad Jaiswal, and Saurabh Panjwani. 2013. Do not embarrass: Re-examining user concerns for online tracking and advertising. In Proceedings of the Ninth Symposium on Usable Privacy and Security. 1–13.

[3] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In Proceedings of the 2019 chi conference on human factors in computing systems. 1–13.

[4] Soontae An and Susannah Stern. 2011. Mitigating the effects of advergames on children. Journal of Advertising 40, 1 (2011), 43–56.

[5] Athanasios Andreou, Márcio Silva, Fabrício Benevenuto, Oana Goga, Patrick Loiseau, and Alan Mislove. 2019. Measuring the Facebook advertising ecosystem. In NDSS 2019-Proceedings of the Network and Distributed System Security Symposium. 1–15.

[6] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Investigating ad transparency mechanisms in social media: A case study of Facebook's explanations. In NDSS 2018-Network and Distributed System Security Symposium. 1–15.

[7] Jack Bandy. 2021. Problematic machine behavior: A systematic literature review of algorithm audits. Proceedings of the acm on human-computer interaction 5, CSCW1 (2021), 1–34.

[8] Natã M Barbosa, Gang Wang, Blase Ur, and Yang Wang. 2021. Who Am I? A Design Probe Exploring Real-Time Transparency about Online and Ofline User Profiling Underlying Targeted Ads. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5, 3 (2021), 1–32.

[9] Reuben Binns, Ulrik Lyngs, Max Van Kleek, Jun Zhao, Timothy Libert, and Nigel Shadbolt. 2018. Third party tracking in the mobile ecosystem. In Proceedings of the 10th ACM Conference on Web Science. 23–31.

[10] Sophie C Boerman, Sanne Kruikemeier, and Frederik J Zuiderveen Borgesius. 2017. Online behavioral advertising: A literature review and research agenda. Journal of advertising 46, 3 (2017), 363–376.

[11] Leanne Bowler, Amelia Acker, Wei Jeng, and Yu Chi. 2017. "It lives all around us": Aspects of data literacy in teen's lives. Proceedings of the association for information science and technology 54, 1 (2017), 27–35.

[12] Elizabeth V Brestan and Sheila M Eyberg. 1998. Effective psychosocial treatments of conduct-disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. Journal of clinical child psychology 27, 2 (1998), 180–189.

[13] B Bradford Brown and James Larson. 2009. Peer relationships in adolescence. (2009).

[14] Moritz Büchi, Eduard Fosch-Villaronga, Christoph Lutz, Aurelia Tamò-Larrieux, and Shruthi Velidi. 2021. Making sense of algorithmic profiling: user perceptions on Facebook. Information, Communication & Society (2021), 1–17.
[15] Moritz Büchi, Eduard Fosch-Villaronga, Christoph Lutz, Aurelia Tamò-Larrieux, Shruthi Velidi, and Salome Viljoen. 2020. The chilling effects of algorithmic profiling: Mapping the issues. Computer law & security review 36 (2020), 105367.
[16] Joy Adowaa Buolamwini. 2017. Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers. Ph. D. Dissertation. Massachusetts Institute of Technology.
[17] José González Cabañas, Ángel Cuevas, and Rubén Cuevas. 2018. Unveiling and quantifying facebook exploitation of sensitive personal data for advertising purposes. In 27th USENIX Security Symposium (USENIX Security 18). 479–495.
[18] Kelley Cotter, Mel Medeiros, Chankyung Pak, and Kjerstin Thorson. 2021. "Reach the right people": The politics of "interests" in Facebook's classification system for ad targeting. Big Data & Society 8, 1 (2021), 2053951721996046.
[19] Elisabetta Crocetti. 2018. Identity dynamics in adolescence: Processes, antecedents, and consequences. European Journal of Developmental Psychology 15, 1 (2018), 11–23.
[20] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated experiments on ad privacy settings. Proceedings on privacy enhancing technologies 2015, 1 (2015), 92–112.
[21] Claire Dolin, Ben Weinshel, Shawn Shan, Chang Min Hahn, Euirim Choi, Michelle L Mazurek, and Blase Ur. 2018. Unpacking perceptions of data-driven inferences underlying online targeting and personalization. In Proceedings of the 2018 CHI conference on human factors in computing systems. 1–12.
[22] Julia Earp and Jessica Staddon. 2016. " I had no idea this was a thing" on the importance of understanding the user experience of personalized transparency tools. In Proceedings of the 6th Workshop on Socio-Technical Aspects in Security and Trust. 79–86.
[23] Motahhare Eslami, Sneha R Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. 2018. Communicating algorithmic process in online behavioral advertising. In Proceedings of the 2018 CHI conference on human factors in computing systems. 1–13.
[24] Danica Facca, Maxwell J Smith, Jacob Shelley, Daniel Lizotte, and Lorie Donelle. 2020. Exploring the ethical issues in research using digital data collection strategies with minors: A scoping review. Plos one 15, 8 (2020), e0237875.
[25] Florian M Farke, David G Balash, Maximilian Golla, Markus Dürmuth, and Adam J Aviv. 2021. Are Privacy Dashboards Good for End Users? Evaluating User Perceptions and Reactions to Google's My Activity. In 30th {USENIX} Security Symposium ({USENIX} Security 21). 483–500.
[26] Yang Feng and Wenjing Xie. 2014. Teens' concern for privacy when using social networking sites: An analysis of socialization agents and relationships with privacy-protecting behaviors. Computers in Human Behavior 33 (2014), 153–162.
[27] Megan Rebecca French. 2018. Algorithmic mirrors: An examination of how personalized recommendations can shape self-perceptions and reinforce gender stereotypes. Stanford University.
[28] David Garcia. 2017. Leaking privacy and shadow profiles in online social networks. Science advances 3, 8 (2017), e1701172.
[29] Carolyn Greene, Jane B Sprott, Natasha S Madon, and Maria Jung. 2010. Punishing processes in youth court: Procedural justice, court atmosphere and youths' views of the legitimacy of the justice system. Canadian Journal of Criminology and Criminal Justice 52, 5 (2010), 527–544.
[30] Kevin Hamilton, Karrie Karahalios, Christian Sandvig, and Motahhare Eslami. 2014. A path to understanding the effects of algorithm awareness. In CHI'14 extended abstracts on human factors in computing systems. 631–642.
[31] Samantha Hautea, Anjali Munasinghe, and Emilee Rader. 2020. 'That's Not Me': Surprising Algorithmic Inferences. In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. 1–7.
[32] Paul Hitlin and Lee Rainie. 2019. Facebook algorithms and personal data. Pew Research Center (2019).
[33] Jane Im, Jill Dimond, Melody Berton, Una Lee, Katherine Mustelier, Mark S Ackerman, and Eric Gilbert. 2021. Yes: Affirmative Consent as a Theoretical Framework for Understanding and Imagining Social Platforms. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 1–18.
[34] Nedim Karakayali, Burc Kostem, and Idil Galip. 2018. Recommendation systems as technologies of the self: Algorithmic control and the formation of music taste. Theory, Culture & Society 35, 2 (2018), 3–24.
[35] Clare-Marie Karat, Jan O Blom, and John Karat. 2004. Designing personalized user experiences in eCommerce. Vol. 5. Springer Science & Business Media.
[36] Caroline Keen. 2020. Apathy, convenience or irrelevance? Identifying conceptual barriers to safeguarding children's data privacy. New Media & Society (2020), 1461444820960068.
[37] Ian Kerr and Jessica Earle. 2013. Prediction, preemption, presumption: How big data threatens big picture privacy. Stan. L. Rev. Online 66 (2013), 65.
[38] ALEXIS R Lauricella, Drew P Cingel, Leanne Beaudoin-Ryan, Michael B Robb, M Saphir, and Ellen A Wartella. 2016. The Common Sense census: Plugged-in parents of tweens and teens. San Francisco, CA: Common Sense Media (2016).

[39] A Lenhart and K Owens. 2021. The unseen teen: The challenges of building healthy tech for young people. Data & Society (2021).
[40] Pedro Giovanni Leon, Ashwini Rao, Florian Schaub, Abigail Marsh, Lorrie Faith Cranor, and Norman Sadeh. 2015. Privacy and behavioral advertising: Towards meeting users' preferences. In Symposium on usable privacy and security (SOUPS). 22–24.
[41] Rebecca Lipman. 2015. Online Privacy and the Invisible Market for Our Data. Penn St. L. Rev. 120 (2015), 777.
[42] Sonia Livingstone, Mariya Stoilova, and Rishita Nandagiri. 2019. Children's data and privacy online: growing up in a digital age: an evidence review. (2019).
[43] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 1–17.
[44] Henrietta Lyons, Eduardo Velloso, and Tim Miller. 2021. Conceptualising Contestability: Perspectives on Contesting Algorithmic Decisions. Proceedings of the ACM on Human-Computer Interaction 5, CSCW1 (2021), 1–25.
[45] Milda Macenaite and Eleni Kosta. 2017. Consent for processing children's personal data in the EU: following in US footsteps? Information & Communications Technology Law 26, 2 (2017), 146–197.
[46] Alice E Marwick and Danah Boyd. 2014. Networked privacy: How teenagers negotiate context in social media. New media & society 16, 7 (2014), 1051–1067.
[47] Tara Matthews, Kerwell Liao, Anna Turner, Marianne Berkovich, Robert Reeder, and Sunny Consolvo. 2016. "She'll just grab any device that's closer" A Study of Everyday Device & Account Sharing in Households. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 5921–5932.
[48] Dan P McAdams and Bradley D Olson. 2010. Personality development: Continuity and change over the life course. Annual review of psychology 61 (2010), 517–542.
[49] Aleecia McDonald and Lorrie Faith Cranor. 2010. Beliefs and behaviors: Internet users' understanding of behavioral advertising. Tprc.
[50] Terrell McSweeny. 2018. Psychographics, Predictive Analytics, Artificial Intelligence, & Bots: Is The FTC Keeping Pace? Geo. L. Tech. Rev. 2 (2018), 514–514.
[51] Nick Merrill, John Chuang, and Coye Cheshire. 2019. Sensing is Believing: What People Think Biosensors Can Reveal About Thoughts and Feelings. In Proceedings of the 2019 on Designing Interactive Systems Conference. 413–420.
[52] Ingrida Milkaite, Ralf De Wolf, Eva Lievens, Tom De Leyn, and Marijn Martens. 2021. Children's reflections on privacy and the protection of their personal data: A child-centric approach to data protection information formats. Children and Youth Services Review 129 (2021), 106170.
[53] Helen Nissenbaum. 2011. A contextual approach to privacy online. Daedalus 140, 4 (2011), 32–48.
[54] Borke Obada-Obieh, Yue Huang, and Konstantin Beznosov. 2020. The Burden of Ending Online Account Sharing. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 1–13.
[55] Luci Pangrazio and Neil Selwyn. 2018. "It's not like it's life or death or whatever": Young people's understandings of social media data. Social Media+ Society 4, 3 (2018), 2056305118787808.
[56] Cornelia Pechmann, Linda Levine, Sandra Loughlin, and Frances Leslie. 2005. Impulsive and self-conscious: Adolescents' vulnerability to advertising and promotion. Journal of Public Policy & Marketing 24, 2 (2005), 202–221.
[57] Emilee Rader, Samantha Hautea, and Anjali Munasinghe. 2020. " I Have a Narrow Thought Process": Constraints on Explanations Connecting Inferences and Self-Perceptions. In Sixteenth Symposium on Usable Privacy and Security ({SOUPS} 2020). 457–488.
[58] Jenny Radesky, Yolanda Linda Reid Chassiakos, Nusheen Ameenuddin, Dipesh Navsaria, et al. 2020. Digital advertising to children. Pediatrics 146, 1 (2020).
[59] Ashwini Rao, Florian Schaub, and Norman Sadeh. 2015. What do they know about me? Contents and concerns of online behavioral profiles. arXiv preprint arXiv:1506.01675 (2015).
[60] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. 2018. "Won't somebody think of the children?" examining COPPA compliance at scale. Proceedings on Privacy Enhancing Technologies 2018, 3 (2018), 63–83.
[61] Minna Ruckenstein and Julia Granroth. 2020. Algorithms, advertising and the intimacy of surveillance. Journal of Cultural Economy 13, 1 (2020), 12–24.
[62] Nick Seaver. 2019. Captivating algorithms: Recommender systems as traps. Journal of Material Culture 24, 4 (2019), 421–436.
[63] Robert Soden, David Ribes, Seyram Avle, and Will Sutherland. 2021. Time for Historicism in CSCW: An Invitation. Proceedings of the ACM on Human-Computer Interaction 5, CSCW2 (2021), 1–18.
[64] Daniel J Solove. 2005. A taxonomy of privacy. U. Pa. L. Rev. 154 (2005), 477.
[65] Mariya Stoilova, Rishita Nandagiri, and Sonia Livingstone. 2021. Children's understanding of personal data and privacy online–a systematic evidence mapping. Information, Communication & Society 24, 4 (2021), 557–575.

[66] Christopher A Summers, Robert W Smith, and Rebecca Walker Reczek. 2016. An audience of one: Behaviorally targeted ads as implied social labels. Journal of Consumer Research 43, 1 (2016), 156–178.
[67] Kaiwen Sun, Carlo Sugatan, Tanisha Afnan, Hayley Simon, Susan A Gelman, Jenny Radesky, and Florian Schaub. 2021. "They See You're a Girl if You Pick a Pink Robot with a Skirt": A Qualitative Study of How Children Conceptualize Data Processing and Digital Privacy Risks. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 1–34.
[68] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In proceedings of the eighth symposium on usable privacy and security. 1–15.
[69] Mikkel Villebro, Irina Shklovski, Luca Rossi, and Alex Bjørstorp. 2018. Comfortably Numb: Danish Teens' Attitudes, Towards Social Media Platforms. In Proceedings of the 9th International Conference on Social Media and Society. 187–196.
[70] Sandra Wachter. 2020. Affinity Profiling and Discrimination by Association in Online Behavioral Advertising. Berkeley Tech. LJ 35 (2020), 367.
[71] Yang Wang, Huichuan Xia, and Yun Huang. 2016. Examining American and Chinese internet users' contextual privacy preferences of behavioral advertising. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 539–552.
[72] Miranda Wei, Madison Stamos, Sophie Veys, Nathan Reitinger, Justin Goodman, Margot Herman, Dorota Filipczuk, Ben Weinshel, Michelle L Mazurek, and Blase Ur. 2020. What Twitter Knows: Characterizing Ad Targeting Practices, User Perceptions, and Ad Explanations Through Users' Own Twitter Data. In 29th {USENIX} Security Symposium ({USENIX} Security 20). 145–162.
[73] Daricia Wilkinson, Moses Namara, Karishma Patil, Lijie Guo, Apoorva Manda, and Bart Knijnenburg. 2021. The Pursuit of Transparency and Control: A Classification of Ad Explanations in Social Media. In Proceedings of the 54th Hawaii International Conference on System Sciences. 763.
[74] Allison Woodruff, Sarah E Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. 2018. A qualitative exploration of perceptions of algorithmic fairness. In Proceedings of the 2018 chi conference on human factors in computing systems. 1–14.
[75] Paweł W Woźniak, Jakob Karolus, Florian Lang, Caroline Eckerth, Johannes Schöning, Yvonne Rogers, and Jasmin Niess. 2021. Creepy Technology: What Is It and How Do You Measure It?. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 1–13.
[76] Yaxing Yao, Davide Lo Re, and Yang Wang. 2017. Folk models of online behavioral advertising. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 1957–1969.
[77] Jason C Yip, Kiley Sobel, Xin Gao, Allison Marie Hishikawa, Alexis Lim, Laura Meng, Romaine Flor Ofiana, Justin Park, and Alexis Hiniker. 2019. Laughing is Scary, but Farting is Cute: A Conceptual Model of Children's Perspectives of Creepy Technologies. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 1–15.

Received January 2022; revised April 2022; accepted May 2022