

USER ACCEPTANCE OF ONLINE TRACKING
IF 'FORGETTING' WAS AN OPTION

by

Vidhi Kirit Shah

Submitted in partial fulfillment of the requirements
for the degree of Master of Computer Science

at

Carleton University
Ottawa, Ontario
January 2020

© Copyright by Vidhi Kirit Shah, 2020

Table of Contents

List of Tables	v
List of Figures	vi
Abstract	vii
Acknowledgements	viii
Chapter 1 Introduction	1
1.1 Contributions	3
1.2 Collaboration Statement	4
1.3 Thesis Organisation	4
Chapter 2 Background	5
2.1 What is web tracking?	5
2.2 Online Tracking technologies	5
2.2.1 Cookies	5
2.2.2 Browser Fingerprinting	7
2.2.3 Web beacons	8
2.3 Users perception of Web tracking	9
2.3.1 Users' unawareness about web tracking	9
2.3.2 Users desire to control ads	11
2.3.3 Users feel embarrassed when seeing some targeted ads	12
2.4 Facebook and Google User Tracking	12
2.5 Anti-tracking tools	14
2.5.1 Private browsing or Incognito mode	15
2.5.2 Deleting Cookies	16
2.5.3 Do Not Track	17
2.5.4 Ad Blockers	18
2.6 Ad preference	18
2.6.1 Issues associated with ad preference managers	19
2.7 Summary of background	19

Chapter 3	Creepiness	20
3.1	What is creepiness?	20
3.1.1	Psychological factor leading to the creepiness	21
3.2	Online tracking is creepy	21
3.2.1	Research evidence	21
3.2.2	Analysing Factors leading creepiness in the context of online tracking & targeted ads	25
3.3	Forgetting and creepiness	27
3.4	Storing User Data	28
3.4.1	File systems	28
3.4.2	Relational database	29
3.4.3	Machine learning	29
3.4.4	Distributed system	29
3.5	Data Deletion Process in different data structures	30
3.5.1	Google cloud storage	30
3.5.2	Facebook F4	30
3.5.3	Recommender system	31
3.5.4	Machine learning and AI	31
3.6	Challenges in Data Deletion Process	32
3.6.1	Cryptography Erasure	32
3.6.2	Data overwriting	33
3.6.3	Artificial Intelligence	33
3.7	Inspiration behind the research study	34
Chapter 4	Methodology	35
4.1	Research Question and Approaches	35
4.2	First study	36
4.2.1	Participants recruitment	36
4.2.2	Study design	36
4.2.3	Information Video	38
4.3	Second Study	38
4.3.1	Participants Recruitment	39
4.3.2	Study Design	39
4.4	Thematic Qualitative Analysis	40

Chapter 5	Results	42
5.1	Demographics	42
5.2	Results from the first study	44
5.2.1	Are targeted ads really of any help?	44
5.2.2	What is online tracking: Companies use cookies to track	45
5.2.3	It is Creepy when I realize the amount of data they track without my consent	46
5.2.4	To avoid ads: Ad blocker is universal	48
5.2.5	Flash Cookies: I didn't know those	49
5.2.6	Google opt out: No I wasn't actually aware of it	50
5.2.7	Some level of control helps	51
5.2.8	Views on Forget	52
5.2.9	Scale Questions	54
5.3	Results from the second study	58
5.3.1	Online Tracking: I think they know pretty much everything	58
5.3.2	Anti-Tracking: You really don't have a choice	59
5.3.3	Do not track: Chats and health-related ads should be off the table	60
5.3.4	Forget: It gives you a sense of comfort	61
5.4	Overall analysis of both the studies	62
Chapter 6	Discussion	66
6.1	Contribution	68
6.2	Limitations	69
6.3	Recommendations and Future work	70
Bibliography		72
Appendix A	Study 1	81
Appendix B	Study 2	116

List of Tables

5.1	Demographics of Study 1	43
5.2	Demographics of Study 2	43
5.3	Responses for Scale Question 1 to Question 4.	54
5.4	Responses for Scale Question 5 to Question 7.	55
5.5	Responses for Scale Question 8 to Question 10.	55
5.6	Responses for Scale Question 11 and Question 12.	56
5.7	Responses for Scale Question 13 to Question 15.	56
5.8	Response for Scale Question 16.	56
5.9	Response for Scale Question 17.	56
5.10	Response for Scale Question 18.	57
5.11	Response for Scale Question 19.	57
5.12	Results Summary	63

List of Figures

2.1	How cookies work [39].	6
2.2	Ad-block plus talking about social media icons(Ref. [79])	14
2.3	Google Chrome Incognito mode(Ref. Google)	15
2.4	Message to enable cookies on Google’s login page.	16
2.5	Do Not Track message from the browser(Ref. [43])	17
2.6	Ad-block plus filter rules (Ref. [79])	18
4.1	Representation of Forget in Study 1	37
4.2	Representation of Forget in Study 2	40

Abstract

Users are constantly bombarded with targeted ads based on their online behavior. Researchers have found that users often feel uncomfortable when they see targeted ads based on sensitive topics. The hypothesis behind this study is that perhaps the problem isn't the tracking, per se, but the retention of so much information about users in an opaque fashion. Hence this thesis aims to understand if giving users fine-grained control over retention of online tracking data (related to specific targeted ads) would change the acceptability of online tracking. We decided to test this hypothesis by conducting two user studies on views of online tracking and data retention and whether a hypothetical "forget" option that deleted all data related to the displayed ad would change those views. Our results indicate that such control does make online tracking more acceptable to users. This research thus motivates further work on fine-grained user tracking data deletion from an implementation and user experience perspective.

Acknowledgements

Firstly I would like to thank Professor Anil Somayaji for investing his valuable time in the completion of this thesis. I am extremely grateful for his immense support and encouragement. I had many break downs during this entire process of the thesis but he kept on encouraging me and made me believe in my abilities. I also want to take this opportunity to apologize to him if I have not stood up to his mark. It is said that without a guru (teacher) no one can ever cross over the shore and I have realized it myself because without him I would have never completed my biggest accomplishment.

I would also like to thank my parents, who are god for me. They have always been so supportive and caring. They are simply beyond compare and I can't express my gratitude for them in words. I look up to them in every aspect of my life. It was a tough journey for me as I am not physically close to them. I haven't seen them from the past few years because I was working hard to fulfill their dream. Through this tough journey of mine, they have been standing with me like a rock. I love you maa paa and my darling sister who makes me laugh when I am stressed.

Last but not least my better half without him I wouldn't have survived. He has stood for me in every difficult time I have ever seen. He is always there to support me. He has equally worked hard for my dreams. He is the best person I can ever get as a life partner. I would also like to convey heartfelt affection to my in laws who have encouraged me throughout my studies. I also want to say a special thanks to the committee: Dr. Sonia Chiasson, Guy-Vincent Jourdan and David Mould who took out their valuable time for making this thesis successful.

Thank you all my teachers, family and friends who were always there to support me. I dedicate this thesis to each one of you.

Chapter 1

Introduction

On the Internet, advertisers track people to target advertisements. Online tracking is a practice used by advertising networks and web analytics to record online data which includes users browsing history, IP address, operating system, search keywords, products bought, age group, sex, time spent online and a lot more. The intention behind collecting an abundant amount of data is to be able to tailor ads based on the user's online behavior. There are various tracking tools used by these networks: cookies, web beacons/web bugs, browser fingerprinting and others.

Targeted advertisements are inherently privacy compromising, yet they are also ubiquitous and constitute the most common online advertisements and an increasingly large chunk of the global advertising market [12]. Many users feel that targeted advertisement threatens their privacy [67, 3, 24, 16]. A study by Rainie [27] revealed that about two-thirds (68 percent) of Internet users said they disapprove of search engines and web sites tracking their online behavior to aim targeted ads at them. Moreover, countless users in the study performed by Agrawal et al. [3] revealed being more worried about observing embarrassing commercials on the web than about their browsing history being tracked by third parties, which implies that in certain communities, embarrassment from online ads is a concern. Cranor et al. [28] observed that users described the idea of tracking and targeted ads as creepy and scary. To deal with these concerns users have increased their use of anti-tracking and ad-blocking tools such as Ad-block plus [79], Ghostery [49], U-block origin [53], incognito mode/private mode, DNT (Do Not Track) [99] and others [5].

Anti-tracking tools are intended to help online users maintain their privacy and avoid online tracking. Several researchers indicated that users felt that they didn't have enough control over tracking even though there exist anti-tracking tools [28, 55]. Agarwal et al. [3] talked in his study on how users lack awareness of the mechanics of targeted advertisements and web tracking and that current consumer choice mechanisms for controlling tracking and education efforts to raise awareness have had limited impact in helping users exercise choice for targeted ads. Another research by Kaye [61] that looked at consumer awareness of the Digital Ad Alliance's Ad Choices privacy icons which aims to provide transparency and notice about internet advertisement to users,

found out that only 6 percent of consumers were aware of this opt out program. Previous attempts to regulate tracking have largely failed, and existing technical solutions to avoiding tracking lead to an escalating arms race that helps nobody. The current tools focus on ensuring user privacy more often limit access to personalized services. The fundamental issue is that present strategies are not sufficient to guarantee client privacy, and the clients have to trust in the privacy policies of web sites (assuming any), which is not secure by any means [63]. Supporting the argument, Agarwal et al. [3] stated that the current techniques to control online tracking or avoiding ads don't address every one of the concerns brought by the users and that there is a requirement for innovation to fill this gap.

Now, one may also recommend that choosing suitable advertising content ought to be the responsibility of ad networks but this has not happened so far. In several researches, online users have mentioned that they would like to stop irrelevant and embarrassing advertisements, something that they still don't have great controls for. They need to have something that controls web tracking as well as can be used as an advertisement blocking instrument [3, 55, 71, 28, 80].

Based on the above findings we can assume that we need to develop ways to help the user maintain their privacy which can also help them reduce the feeling of embarrassment and creepiness when they see targeted ads. But to design such tools, we first need to understand what is the user's perception of creepiness. Creepiness is a feeling you get when you are nonconsensually observed while engaging in private behavior. Although users have often talked and discussed creepiness, there have not been many efforts on understanding creepiness in the context of tracking and targeted ads.

Our insight is that perhaps the problem isn't tracking, per se, but the retention of so much information about users in an opaque fashion. Hence this thesis aims to understand if giving users fine-grained control over retention of tracking data could change online users acceptance about online tracking and targeted ads. To test this idea, we propose 'forget' as a hypothetical opt-out mechanism. 'Forget' allows online users to delete all the tracked data behind a specific ad. We chose 'forget' in our study as it was a conceptually simple mechanism that could be available to the user at the time when they encountered a creepy ad. In principle, data retention could be controlled by users in other ways, say through interaction with an activity browser or through privacy policies. But we thought that allowing user to 'forget' their tracked data would possibly be a better choice to understand their acceptance towards online tracking.

The idea of forgetting is inspired the observation that feelings of embarrassment could be reduced if one could request that others forget things that they have observed. Human brains cannot be instructed to forget certain memories deliberately, we forget when our subconscious considers a memory unimportant but many can still be recalled with the right stimulus [87]. But this is not the case with computers. We can instruct computers to forget things that we don't need anymore. We can delete data manually or automatically. But there are some challenges in deleting data from computer and data storage which we discuss further in Chapter 3.

We conducted two studies with an overall 22 participants. The research had scale questions and open-ended questions. We used a video from the Wall Street Journal [104] which represented the concept of tracking and cookies clearly and understandably. We further used paper mock ups to explain current ad preference mechanisms and how 'forget' would work. The study was audio recorded to make transcripts. We performed a qualitative analysis of the collected data to evaluate our results. Our results for the first study were inconclusive as we discuss in Chapter 5. Hence, we performed a follow-up study in which we added a better explanation of how 'forget' worked. Overall, our results suggest that 'forgetting' could enhance the trust of online users on the process of online tracking and can increase their acceptability towards targeted ads, suggesting that further research is warranted into fine-grained mechanisms for user control of ad tracking data.

1.1 Contributions

This thesis makes the following key contributions:

- This research is the first to propose 'forget' as an option for addressing user concerns regarding targeted ads.
- This work is the first attempt to understand user's perception about 'forgetting' or deleting data associated with targeted advertisements.
- Our research results also confirm the results obtained from past studies about user's acceptance of online tracking and targeted ads, specifically that providing user control can increase user acceptance.

1.2 Collaboration Statement

I did all this work all on my own under the supervision of professor Anil Somayaji but part of my background Chapter 2 was based on a class project of Computer Security and Usability taught by Dr. Sonia Chiasson.

1.3 Thesis Organisation

The thesis is organized as follows:

In Chapter 2, we provide background on various tracking tools used by ad companies and the tools available to users to avoid tracking and online ads. We also review past work on user perceptions of online tracking and advertisements.

Further in Chapter 3, we discuss what is creepiness as per human psychology and what are the factor leading to creepiness and whether the factor match the existing web tracking system. We also discuss the challenges associated with forgetting user data.

In Chapter 4, we present the methodology of the two user studies conducted as part of this work.

In Chapter 5, we present our results. This chapter has two sections each discussing results from two separate studies. We evaluate our study results and also compare them with past work. We categorize the results in the form of themes.

In the final Chapter 6, we discuss our results, contributions, limitations and recommended future work.

We also have attached two appendices which contain the study materials for our first and second user studies, respectively.

Chapter 2

Background

In this chapter, we discuss what is web tracking, how it works, technologies used for tracking, and tools to avoid tracking. We also discuss users perception about online tracking tools, anti-tracking tools, and ad preferences.

2.1 What is web tracking?

Web-based tracking is the act of following, recording, and storing web history [88]. Tracking is done with the help of different technologies: cookies, browser fingerprinting, web beacons and others. We explain key tracking technologies in the following sections. Advertisers use tracking information to generate personalized ads instead of random ads [23] while web analytic allow website performance and user activity to be monitored, helping companies optimize user experience and revenues rather than an individual user [4]. Every website user visited, time spent on each website, online shopping, watched videos, an online search is been tracked by various tracking tools [56].

2.2 Online Tracking technologies

In this section we talk about various tracking tools been used by advertisement networks, a few of which include cookies, browser fingerprinting, and web beacons. While there are other tracking mechanisms, these are commonly used and are sufficient to show the breadth and depth of current tracking technology.

2.2.1 Cookies

Cookies are text files that are saved in client browsers at the request of web servers. Once saved, they are sent back to the designated server as part of any subsequent request. Cookies are used by many websites to keep track of traffic flowing to their site, record user's preferences, and maintain sessions as a given session can involve many separate connections to a website. Several websites

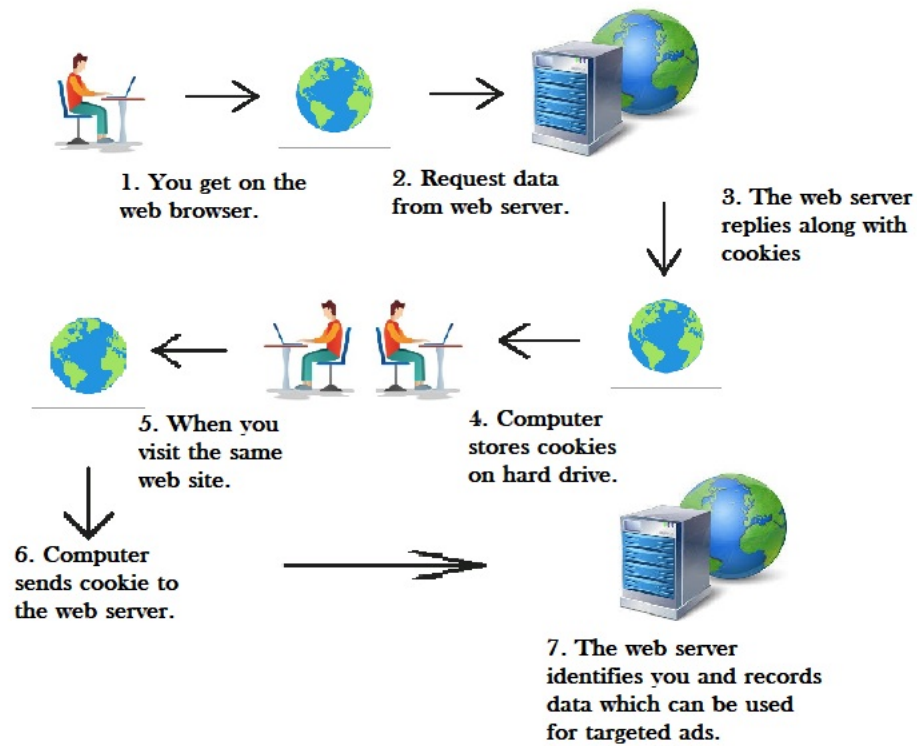


Figure 2.1: How cookies work [39].

have advertising material served by third party sites. These advertisements can store cookies for third party site which contain information from the containing site. This data is then used by ad companies for target advertising based on user browsing history [40]. Figure 2.1 explains how cookies work.

Consider the following scenario. A client will be prompted for data, for example, their name, postal address, and email address as part of signing up with an online retailer. The data will be stored in a database at the retailer and a unique identifier for it will then be stored in a cookie file that then saved in the client's browser profile. When the user visits the same retailer later, the recorded cookie will be sent, allowing the user to be identified even if they have changed IP addresses by connecting to a different network. The retailer can then use the cookie data to customize the user's experience, say by suggesting items based on their mailing address or purchase history. When content from web servers other than the retailer (such as an advertising network, social media feed, or web analytics service) are included as part of the page, they too can set cookies (called third-party cookies) than can be used to track the user on any website that embeds similar content [49, 40, 96].

As cookies are frequently used to recognize visitors to web sites, many people believe them to be a potential privacy invasion. The principal concern is that this is managed without one's consent. By utilizing cookies, organizations can get individual data, for example, purchasing habits, email address or the segments of a site that were seen before. This data can be joined into mailing records for direct advertising purposes or it tends to be sold to third parties.

HTTP cookies are not the only way for web servers to store identifying data in clients. Adobe Flash has its own cookie functionality [2], which can be particularly problematic as Flash cookies never expire [100] and most users don't know about them even if they know about regular cookies. While Flash cookies are only read when Flash content is accessed, they can be used to restore HTTP cookies that have been removed, thus "re-spawning" cookies [100, 96]. As Flash has decreased in usage, HTML5 local storage [107] has come into wider use, providing even more capabilities for storing identifying data on clients [7].

2.2.2 Browser Fingerprinting

Fingerprinting can be utilized to recognize clients and track them across sites when cookies are unavailable or are unreliable. As the name infers, a "unique mark" of the browser is calculated from aspects of the browser's configuration. Browser fingerprinting uses a variety of features including

browser type and version, operating system version, browser plugins, time zone, language, resolution of the screen. This information may appear to be generic and not specific to an individual, yet this is not the case. Panopticlick [32] found that solitary one out of 286,777 different browsers will have the same browser fingerprint. Cookies can be deleted from a computer but we cannot delete a browser's fingerprint, making it an effective tracking tool [86].

2.2.3 Web beacons

A web beacon also is known as a web bug, pixel tag or a clear GIF is a 1x1 pixel graphical image that is included in a web page or is embedded in an email [78]. They are often unidentified as they are too small to be noticed. Their purpose is to allow third parties to check if a user has accessed the web content on the web page or read an email [38, 98]. Gralla [48] explained that web beacons are everywhere, even in the places we'd least expect them. He said that the website of New York's Metropolitan Museum of Art has 28 of them. They are ubiquitous. At first, the companies using web beacons for tracking were mainly advertisers; later social media sites also started to use such tracking techniques, for instance through the use of buttons that act as tracking beacons [78, 96]. They are available in GIF format which is universally recognized on almost every browser. Beacons are often used with third-party cookies. Whenever a web page with the web beacon is loaded the beacons is been loaded from the third-party server. The beacon's GET request sends the client's information to the third party. By correlating beacon requests, clients can be tracked within and across websites, revealing what pages were visited when and for how long. Note that web beacons require no data storage on a client. User activity tracked by web beacons are used by several marketers, ad companies, and third party websites to publish targeted advertising to users [48, 78, 98, 94, 75].

According to Know Privacy's 2009 report on Internet privacy [50], the 50 most popular websites on the Internet all contain at least one third-party web beacon; some sites carry as many as 100. Their prominence shows how common and pervasive tracking is. It is very difficult to detect web beacons visually on any site as they are too small to notice; they can only be detected by observing the network traffic associated with a web page and are only documented (sometimes) in privacy policies—there are no pop-up messages about web beacons (unlike cookies). This shows how user activity is been tracked without user consent or knowledge [101].

2.3 Users perception of Web tracking

In this section, we try and understand what are users' views on web tracking and targeted ads. From the literature, we found that users have negative opinions of online tracking and targeted ads when it comes to sensitive topics. They felt that ads on sensitive topics embarrassed them while they were not in a private space and they felt that it was an invasion in their personal space [22, 28]. Users also felt that certain ads based on sensitive topics were creepy—we discuss this in the next chapter. Further, we also observed that users were not completely aware of the anti-tracking techniques or ad-blocking techniques they can use and those who were aware described the issues associated with using it [75, 68, 3]. Moreover, they expressed their requirement to have control over their own tracked data which they currently don't have.

Users have also expressed their concerns on targeted ads which are based on sensitive topics such as health. Carrascosa et al. [22], in his study to measure online behavioral advertising, found that the advertising market targets behavioral traits associated with very sensitive topics including health, politics, and sexual orientation despite the fact that such tracking is illegal in several countries. Statistically, 88% of the analyzed personas got targeted ads associated with all the keywords that defined their behavioral traits. Jaiswal et al. [3] found that around 85% of users have a negative picture for third-party tracking and targeted advertisement. Subsequently, in 2017, Parra-Arnau et al. [80] found that two out of the three Internet users are worried about their online behavior being scrutinized without their knowledge and consent. Similar user perceptions were observed earlier in 2007 when Facebook users criticized Facebook Beacon for sharing their online browsing data with their friends without the users' consent [68]. Plane et al. [84] explained how targeted ads can be discriminatory, leading certain groups to see better offers like job ads based on personal characteristics such as gender. Recently Vlajic et al. [106] also highlighted that children are also been tracked online, and the children's digital advertising market has been experiencing an astonishing 25% year on year growth and is expected to reach the US \$1.2 billion by 2019.

We can categorize user perception into different groups as follows:

2.3.1 Users' unawareness about web tracking

It is been observed that users are unaware of web tracking and presences of tracking tools on the websites. Below are some of the statements from previous user studies for web tracking and targeted ads which clearly defines their unawareness about web tracking techniques. Agarwal

et al. [3] interviewed 53 users to understand their concerns regarding web tracking and targeted advertising. He concluded in his study that users were unaware of the tracking technologies used by ad companies as one of the users has to say the following:

“Is there a way to know that third parties are tracking me on a website? or Do they track just the fact that I visited a website or anything else? or Do you mean, against one cookie, there are several websites that Google knows I am visiting?” [3]

Notenboom [75] in his article explained how one can detect web beacons in an email. The aim of this article was to understand whether online users were aware of the presence of web beacons in emails or not. One of his follower expressed the curiosity after reading his article as he mentioned:

“On knowing about blocking images from being downloaded when opening web email. I’ve since tried to find more information about those hidden web beacons (or web bugs) that can track an email recipient if the images are not blocked. Going a step further, would you kindly be able to explain how to detect the URL of the aforementioned web beacons or bugs from an email’s source code?” [75]

Martin [68] in his study talked about Facebook beacons and how it affected user’s privacy. He concluded that more than 50% of Facebook users were unaware of third party beacons were tracking them and were sending their off the Facebook data to their friends. In his study, one of the participants mentioned that experience of beacons sending his data without even asking permission from the user. Below is a statement by the Facebook user from Martin’s [68]study:

“So here I am, burning some brain cells and taking some time to relax playing a game on Kongregate when a little window pops up in the corner of my screen and says —Kongregate is sending this to your Facebook profile: Nate played Desktop Tower Defense 1.5 at Kongregate. Which immediately elicited a —Hell no from my mouth. Maybe what shocked me was the way it was worded, essentially saying that Kongregate was sending the data without even asking my permission (even though there is a No Thanks button in the corner) but needless to say, I was not too thrilled about my surfing habits showing up on my Facebook profile.” [68]

2.3.2 Users desire to control ads

Several researchers has suggested that users desire to have control over tracking and tracking data. Jin et al. [55] theorize that online users are progressively open to targeted ads in the event that they can examine, control, and understand advertisement choices; however, they also suggest that simple transparency may be sufficient. Below are some statements recorded by other researchers regarding users wanting control over ads.

Users sometimes want specific kinds of ads yet they cannot make such requests. Agarwal [3] found in his study that almost all the users somewhere or the other were not satisfied with the current anti-tracking mechanism as one of the 53 participants said the following:

“Can I choose ads from certain service providers? Like for example, I want travel ads, is that possible to say? If I can specify what are the kind of ads I want to see, or which are the service providers I’m ok to see ads from, then it would help. If it completely blocks out everything, then not.” [3].

Irrelevant, repetitive, and badly timed advertisements timed and repetitive ads can lead to user dissatisfaction. McDonald [70] found that 40% of users wanted to control the ads they were shown. One of their 14 participants expressed their desire to control certain ads which were repetitive as below:

“There should be a time after which I can push It., that is, I can say, ‘Now, change my profile’ because if I have already purchased a hard disk and then every day it is showing me that buy-a-hard-disk ad. It is irritating.” [70]

Ads that are unrelated to a user’s browsing history or online behavior but are displayed to the user can be very annoying. Cranor [27] conducted a study to understand users belief on targeted ads and he found out that 46% of online users often felt uncomfortable when they saw ads which were not based on their preferences especially in public space. As one of his few participants gave the below statement:

“I think what is embarrassing is that the ad reveals what kind of person you might be. The other person who would be with me when this ad came up would interpret more into that. Everybody has knowledge that your ads are being customized to your taste. I would obviously not like” [27]

2.3.3 Users feel embarrassed when seeing some targeted ads

Several researchers found that a ‘feeling of embarrassment’ was the most common and repeated comment users had about targeted advertisements. We came across several user statements and finding of researchers that specifies that the user feels embarrassment rather than fear of been tracked. Sambasivan [88] did a study on women in Asia to understand their perception of online tracking and their privacy practices. He found that more than half of 199 users often felt creepy and embarrassed when they would see an ad based on sensitive topics like sex, health or religion. One of the participants expressed how he is was embarrassed when he saw an ad displaying sexual content:

“Quite often I am watching something on the Internet and suddenly a porn ad or video pops up. I immediately lock my screen in that case and look around if anybody has seen this or not. I then open it again when nobody is around, view it and then delete or close it. My brother and parents would definitely not like the idea of me watching porn.” [88]

Agarwal [3] also concluded in his study how ads based on sensitive topics made the user feel embarrassment. Almost 74% of the study participants expressed their experience where they felt embarrassed by the ad content. We refer to one of his participant’s statement as he described:

“Whenever I open my Yahoo mail, it constantly shows available girls less than 25 in your area. I don’t have a clue how it can do that, it is extremely embarrassing at work. I assume that other people may also know what this thing is and may assume that the reason I’m being shown these ads is that of my previous history.” [3]

2.4 Facebook and Google User Tracking

In this section, we talk about Google and Facebook to understand how much they track online users and how. It is important to understand Google and Facebook in particular when studying user tracking because they both control so much of the online advertisement market. Google and Facebook hold the largest share of total US digital ad spend, with 38.6% and 19.9%, respectively [34, 97]. Their policies influence other players as well as much of the technology such as Google Chrome, Gmail, Android, Facebook Connect, Instagram, Facebook apps and others. User perception of online tracking is thus tied to how people perceive these corporations.

Know Privacy [42] estimates that Google tracks 88.4% of the traffic on the web. While Harmon [50] stated that Google's beacons are particularly ubiquitous: Google can track activity not only on sites with Google Analytics installed but also on sites that subscribe to Google-operated advertising services which include AdSense and DoubleClick. While Google Analytics isn't the only third-party analytics tool, it is the most popular.

Google Analytics, like most third-party analytics services, starts with the site operator adding a JavaScript snippets to their website which loads the Google Analytics code from Google's servers. Thus every time that page is loaded, Google Analytics is loaded into the client browser. This code instruments the page so that it can record all user activity; it also records characteristics of the browser environment. This recorded activity allows site operators to better understand who is accessing their website, how they got to the website, and what they did when they visited. [46, 47]. This information also goes to Google. Note that Google Analytics uses the DoubleClick cookie as part of its analysis [46].

While Facebook has been found tracking the online user behavior of non-Facebook users. The small Facebook like icon on every website collects data about the user even if they don't hold a Facebook account to create a shadow profile and use these data which includes an IP address to encourage a user to join Facebook [108]. Figure 2.2 taken from ad-block plus settings page also demonstrated that social networking sites record data with social media icons present on different sites which are often unnoticed [79].

Martin [68] talked about Facebook's beacon controversy which happened in November 2007, when Facebook began offering a free tool, Beacon, to third-party websites for tracking user activity. It recorded Facebook user online activities and proactively broadcast off-Facebook activities to designated Facebook friends. In response to this, Facebook received intense criticism. The case was ended by a permanent termination of the system and an establishment of a Privacy Foundation. Facebook finally paid \$9.5 million in total to resolve the privacy concerns around its users and settled the case. Facebook beacon was just a third-party analytics service (like Google Analytics) with a default opt-in to reporting on social media. That connection to social media, however, made a huge difference. Below we referenced the statement given by Mark Zuckerberg on the entire controversy of beacons. This statement is included in this paper as it suggests that Facebook beacons existed and were used for tracking off Facebook activities :

“When we first thought of Beacon, our goal was to build a simple product to let people share information across sites with their friends. It had to be lightweight so it wouldn't

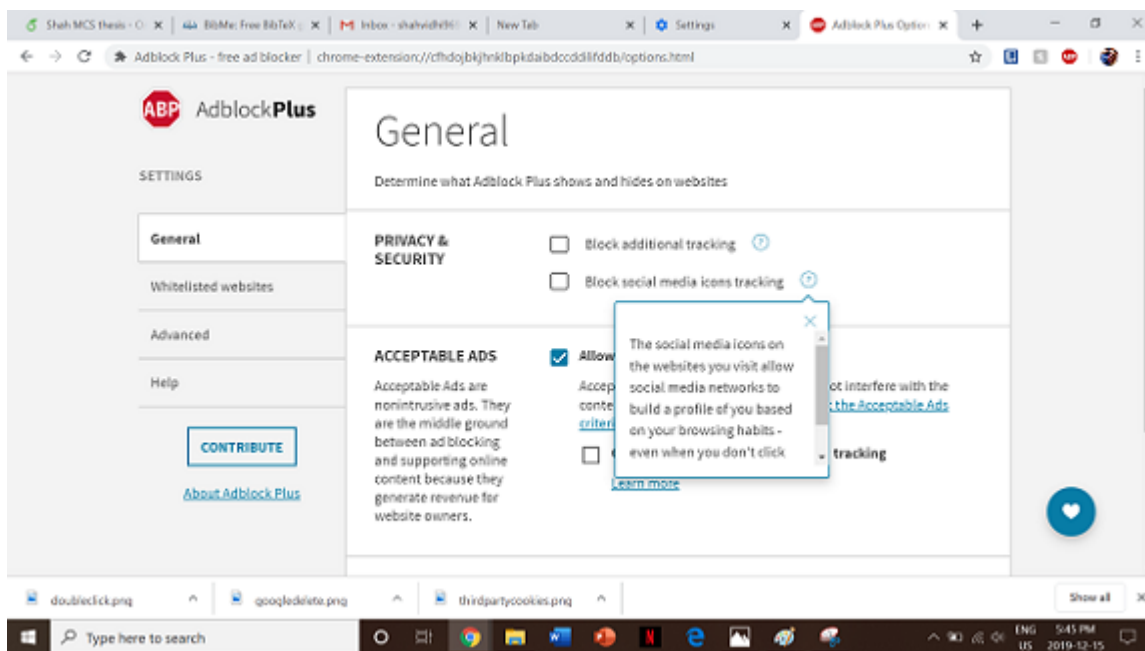


Figure 2.2: Ad-block plus talking about social media icons(Ref. [79])

get in people's way as they browsed the web, but also clear enough so people would be able to easily control what they shared. We were excited about Beacon because we believe a lot of information people want to share isn't on Facebook, and if we found the right balance, Beacon would give people an easy and controlled way to share more of that information with their friends. But we missed the right balance. At first, we tried to make it very lightweight so people wouldn't have to touch it for it to work. The problem with our initial approach of making it an opt-out system instead of opt-in was that if someone forgot to decline to share something, Beacon still went ahead and shared it with their friends.” [101]

Although Facebook had eliminated beacon mechanism but it seem to be using the like buttons to track online users on third-party sites independent of whether the user has a Facebook account or not.

2.5 Anti-tracking tools

This section focuses on the existing solutions we have to avoid web tracking, ads, and web beacons. Private modes, disabling/deleting cookies, Do Not Track, ad blockers, and disabling all HTML in the email is the primary existing solutions for users to maintain online privacy. While each

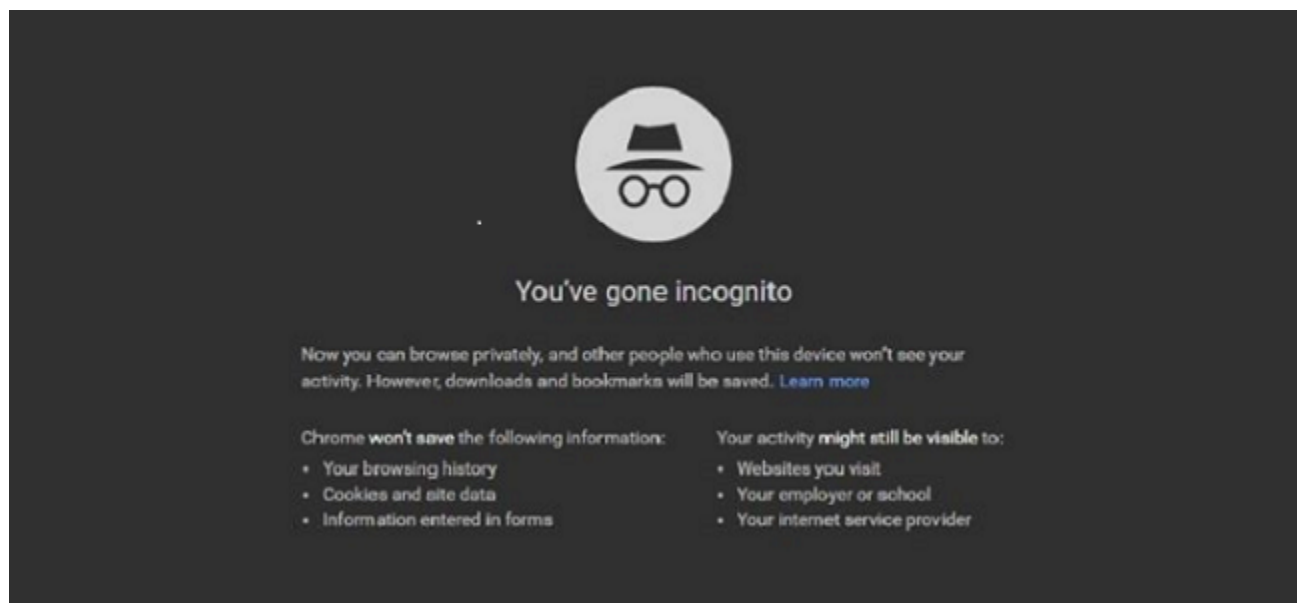


Figure 2.3: Google Chrome Incognito mode(Ref. Google)

provides some privacy protection, each has a significant impact on usability and/or functionality, making them, even collectively, partial solutions at best.

We further discuss a few anti-tracking tools which can be used by online users to avoid tracking and targeted advertising.

2.5.1 Private browsing or Incognito mode

Private browser modes—private browsing or incognito mode—are special browser windows or tabs that delete all state associated with them when closed. They do not record history, do not cache content, and they use a separate ephemeral cookie store. Because they do not allow the use of regular cookies, they mitigate most regular tracking. Although private modes can be helpful for users they are not a perfectly reasonable solution for users to use. Sambasivan et al. [88] observed in her study that the majority of users were not aware of what the private modes in the web browser did or where to find them. Another issue was the terms been referred to in the private modes were difficult for a user to understand. Several users often linked private modes with secretive activities which ultimately threatened participant’s values. Apart from them, Bielova et al. [12] proposed that private browsing modes cannot disable third-party cookies. And even after using private modes a user might be tracked as some user online activity is still visible, as explained in Google chrome incognito mode as Figure 2.3.

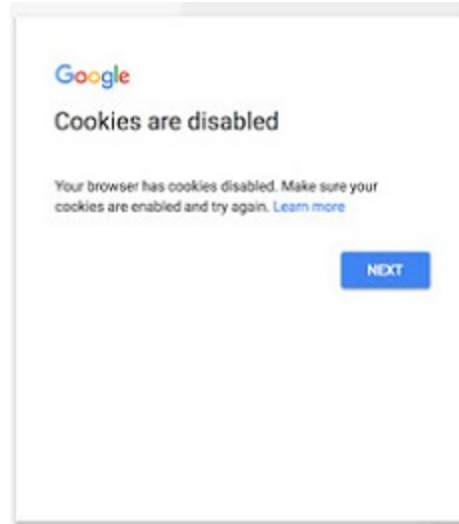


Figure 2.4: Message to enable cookies on Google’s login page.

2.5.2 Deleting Cookies

As we earlier discussed cookies are used by third-party advertising networks and web analytics services to track users’ online behavior. As people have become more aware of cookies tracking them many of them choose to delete cookies periodically or use tools or restrict the storing of cookies. However, deleting or disabling cookies cannot stop web beacons from tracking a user. Cookies can also be restored. Kamkar [59] talked about “ever cookies” which restore cookies after they have been removed. Ever cookies utilize diverse mechanisms, including HTML5 local storage, Flash cookies, and browser fingerprinting to determine how to re-create deleted or missing cookies. [59, 49]

Deleting or disabling cookies may prevent a user from having access to all of a site’s services and features. This can be seen in the statement given by one of the participants during a research study of McDonald et al.::

“We have no choices about cookies, because if you say no then you don’t get to go to the site. That’s not much of an option” [70]

This statement is grounded in fact, as can be seen in Figure 2.4, which shows that Google restricts access to web content when cookies are disabled.

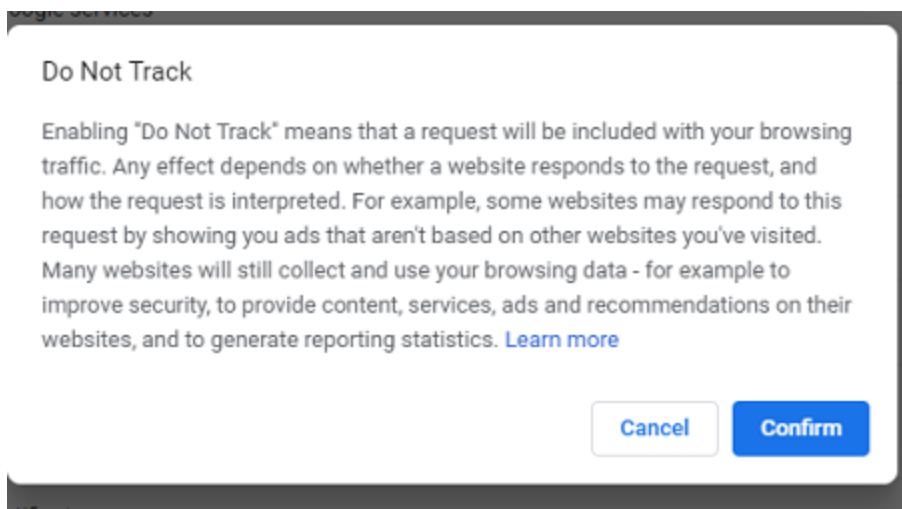


Figure 2.5: Do Not Track message from the browser(Ref. [43])

2.5.3 Do Not Track

The do not track header is the proposed HTTP header field DNT that request a website to disable its cross-site tracking of an individual user. DNT field accepts three inputs: 1 in case the user doesn't want to be tracked, 0 in case the user provides its consents to be tracked and null if the user has no preference. Mozilla Firefox became the first browser to implement DNT and later was implemented by Google Chrome, Apple Safari, and Internet Explorer [99, 8].

Microsoft enabled by default the Do Not Track feature in Internet Explorer 10 and Windows 8 and was highly criticized by advertising companies, who claimed that this decision should be purely of a user and must not be made by a company. The ad companies also said that these would result in the violation of Digital Advertising Alliance's agreement with the U.S government to honor a Do Not Track system because the coalition said that they would only honor the system if it were not enabled by default by web browsers. Due to the criticism, Microsoft announced to no longer enable the Do Not Track system by default, but they shall provide the user with clear information on how they can use it [99, 5].

Bacchus [8] in his study said that websites like google don't honor request set up by Do Not Track systems. Moreover, there are no legal or technological restrictions on the companies to use DNT. There is no penalty for a website if they dishonor the Do not Track claim. Hence we can say that the user has no control over whether Do Not Track is accepted by the ad companies or not.

This is evident from Figure 2.5 which shows a screen capture from Chrome DNT settings which says that even after enabling DNT some websites might still track users browsing history.

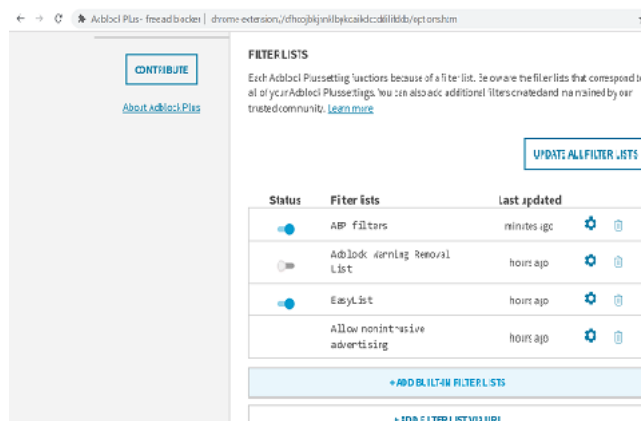


Figure 2.6: Ad-block plus filter rules (Ref. [79])

2.5.4 Ad Blockers

Ad blockers remove ads and beacons and restrict cookies from websites. Ad blockers thus improve user privacy. Also, because they remove many resource requests, they can reduce page load times; they can also minimize visual clutter on websites from distracting advertisements. Ad blockers can also help improve security by avoiding malicious advertisements [92].

Ad blockers rely on rule sets that define whether a given website is a tracker or not. But if a tracking website changes its domain, the rules will no longer match and ads won't be blocked from the new domain [12]. See Figure 2.6 for an example of ad-blocking filter rules.

2.6 Ad preference

Apart from anti-tracking tools, some advertising organizations allow online users to view their ad profiles. They also allow users to change their ad profile by adding or removing any interest that is within their ad profile. They can simply add or delete an interest or a subject on which they would like to see ads or not see ads respectively. This functionality is called an ad preference manager and is offered by ad platforms companies such as Google and Facebook [9].

However, in past studies, it has been seen that users are either unaware of this functionality or they find it difficult to navigate through. Hence below we discuss the issues users face while using ad preference managers as documented in past research.

2.6.1 Issues associated with ad preference managers

Facebook permits users to control the accumulation and use of information or to give feedback about the relevance of advertisements. Researchers have found that ad networks provide information about why a specific ad was shown to a demographic and not an individual user. Hence, it is still difficult for users to see how the advertisement is chosen specifically for them. Besides, numerous users still need the confidence to control the information, be it because of restricted comprehension or on the other hand because of awful user interfaces. In addition to this users are worried about tracking in general, yet current tools don't moderate their worries [71, 30, 3, 28, 55]. A few organizations give "privacy dashboards" where users can see a portion of the data the organization has gathered about them and modify the ads they receive [36, 44]. Shockingly, research experiments have discovered that the data on the dashboards is both incorrect and obscure [71, 30]. Wills and Tatar [111] additionally found that Google didn't uncover all interests in the Ad Settings page as some of the user interest was missing from the ad settings page.

2.7 Summary of background

We discussed what is tracking, the technologies used to track online users, user's perception about online tracking and targeted ads, and anti-tracking tools. We saw that tracking is not only limited to the web but it is also spread across communications channels through web beacons. We also focused on user's perception of web tracking and targeted ads where we found that users don't like targeted ads and they were embarrassed when certain ads based on sensitive topics were shown. They expressed their concern about not having control over web tracking and targeted ads. We further discussed existing technologies to avoid tracking and ads in which we talked about Do Not Track, private mode, and deleting/disabling cookies. We also covered the issues associated with ad preferences. We discussed issues users have with the existing anti-tracking tools. We highlighted the difficulties associated with each anti-tracking tool.

In the next chapter, we explore the psychological impact of online tracking, particularly the source of its perceived creepiness.

Chapter 3

Creepiness

In this chapter, we discuss what is creepiness as per human psychology and the factors leading to it. We also discuss if online tracking is creepy with some research evidence. Based on the analysis of the factors leading to creepiness we discuss ways we can disrupt the creepiness issue in the context of online tracking. We discuss if forgetting can be a solution to get rid of the feeling of creepiness. We further focus on challenges associated with forgetting data. And at last, we highlight the research question of this work.

3.1 What is creepiness?

Creepiness is the condition of being frightening or causing a sentiment of dread or unease. It is a feeling of fear which is aroused by a possible threat [74]. The web has been depicted as progressively frightening [52]. Creepiness has different definitions in a different context. It can be human behavior or human being itself. To support this we have McAndrew and Koehnke [69] who highlighted which classes of behavior can be creepy for a human being. If a person is listening to your conversation while not talking to you can be considered creepy. And a person watching an individual inappropriately can be creepy as well. A person either coming too close or the one who keeps talking about sex or something too personal can also be a creepy person. These all are based on the main theory which is when human beings are unsure of whether to fear or not and when they are not sure if the opposite party is going to cause them any harm in a situation they feel creepy [69].

Based on human psychology we can assume that creepiness is aroused when we are uncertain of possible harm. It is creepy when a stranger constantly observes us and pretends to remember us or our behavior to possibly harm us. It's similar to a situation when a person is observing you with an unclear motivation or intention. Anyone in these situations would feel creepy because they are not sure whether they need to worry about it or not. It causes a feeling of unease [74, 103, 69, 83]. Thus we can define creepiness as a feeling one can get when they are non-consensually observed while engaging in private behavior. For example, a stranger watching looking at an individual

through his window, while he is doing his own thing at home would normally be creepy.

We further discuss the factors causing creepiness.

3.1.1 Psychological factor leading to the creepiness

Based on the definition of creepiness if we had to hypothetically derive the most common factor leading to the feeling of creepiness than we can derive one major factor than it would be ‘Observation’: Being watched with unknown intentions.

As we discussed earlier when someone is observing us constantly with an unclear intention and we are not sure what is their purpose and if they are going to harm us than this situation can be creepy as per the definition. Usually, we are been watched whenever we are in public space. We may pass by thousands of people while we are walking on the streets or present in a public space. We don’t remember them because we never pay attention to all of them. But if someone is staring at you or paying keen attention to you than they are doing it intentionally to observe you and possibly remember your habits.

3.2 Online tracking is creepy

In this section, we understand if online tracking is creepy or not. We discuss research evidence where users described their online experience as ‘creepy’. Further, we compare the factors leading to creepiness we found in the previous section with the online tracking process to find out if online tracking is creepy or not. Online tracking is a process of recording user’s online behavior to show targeted ads based on the recorded data. Often users have said that the extent of personal information tracked by different tracking tools is very creepy for them [28]. We have several research papers who are evidence of online users claiming online tracking to be creepy.

3.2.1 Research evidence

In this section, we aim to highlight what is the users perception of tracking and targeted ads in past studies to understand what is their derivation of creepiness. We discuss the situations which make online users feel that tracking is creepy and what is their overall outlook towards the idea of creepiness in terms of web tracking. Here we show various research evidence to support the generalizations.

The web has been depicted as progressively frightening [52]. Dolin C et al. [30] also found that online users discovered targeting and personalization frightening or creepy. Tracking resulted in negative feelings for the data collector. Not only this Heller [52] studied that the feeling of creepiness increased due to lack of transparency of what data was collected, how it was collected, where it was used and to what extent it was shared. In support of that Yao et al. [114] also demonstrated that individuals are increasingly worried about the sorts of user data gathered than who was gathering it, while Agarwal et al. [3] discovered that users were more worried about the content of targeted advertisements being presented than the related tracking.

Not only this many users feel that web tracking is unethical too. It was found in a study conducted by cybersecurity firm RSA Security that 68% of customers express web tracking to make targeted ads is unethical [112]. Another study by the worldwide cybersecurity firm RSA Security found that there was expanding user reaction in light of various data breaches, uncovering a shrouded danger of advanced targeted advertisement: loss of user trust. The investigation over viewed more than 6,000 adults in France, Germany, the United Kingdom, and the United States and found that overall, online users progressively saw targeted advertisements as unethical and intrusive [6].

There are many sensitive topics which user don't want ad companies or web analytics to track. Often users are been tracked without even them knowing about it. Their data have been tracked and collected without even their consent as we saw it in Chapter 2. Bilton [13] described such incidents in his blog how an application named 'Girls Around Me' would track girls without their consent and provide their respective Facebook profile information without their consent and without even the girls knowing about it. It is similar to a stranger keeping an eye on you and sharing all your details without your consent to another individual. He described that this app named 'Girls Around Me' utilized 'Foursquare', the location-based service, to decide the user's area. It is at that point examines for ladies in the region who have as of late checked in on foursquare. When a user using the app 'Girls Around Me' sees a lady he'd like to converse with, one that has no clue that someone is snooping on her, he can interface with her through Facebook, see her complete name and profile photographs, and then send her a message [13].

There are also several other instances where users felt creepy when they were shown ads related to topics which they talked about but never searched for. Some users suspect that their conversations are been tracked and used to display targeted ads. One such incident is described by

Bjorn [15] where he was talking about going on a February ski excursion and afterwards was presented a ski-related advertisement without having done any other online activity related to skiing. To understand what is creepiness for online users two Brooklyn-based craftsmen made a site called The New Organs which endeavors to imagine the undeniably forceful targeted ads that online users are displayed. Some of the user stories on the sites resemblance towards the user's perception that cell phones are listening to their conversations [21].

“At a party, had good contact with a girl. Nothing happened. We both went our separate ways. Later, she was the first friend's suggestion on Facebook.”

“I was talking on the phone about a street where I was supposed to meet some friends. And the first suggestion in Google Maps was this street. I've never been to this street before.”

“A friend sent me a photo of the new keyboard he had bought (we had talked about it over voice, no typing) and from then on I would see advertisements for this very specific keyboard!”

“I bought these pineapple-shaped earrings in a small cafe in the Queensland countryside. I paid cash. A short time later I googled plastic jewelry on my phone and an image of these exact same rubber earrings appeared at the top of the search.”

“Coworker described a gift she got for a friend for Christmas. I responded enthusiastically and said I'd love one. BAM! I shit you not, I had an email from Amazon the next day with an exact link to THAT product. Creeped us both out [21].

Based on the above statements it seems that some users feel that their communications have been monitored and used for targeting ads. Mainstream advertising firms state that they do not use communications for targeted advertisements; nevertheless, it is clear that the mere possibility of such observation is very creepy to many individuals.

Apart from this, there have been several other instances where social media was blamed for tracking users without their consent. As Facebook once got in a difficult situation in March 2018 for recording each telephone call clients made on their Android devices. When updates on the Cambridge Analytica embarrassment broke and clients around the globe began vowing to blacklist Facebook, individuals began downloading and looking once again the information the internet based life mammoth had been putting away on them. Incredibly, they found that Facebook had tracked each telephone call they made. Facebook had full details on who they had called [109, 77].

Facebook users noticed an image posted by Mark Zuckerberg where he had dark tape over his workstation's webcam. Facebook has never admitted to watching users through their cameras, yet the way that Zuckerberg wants to cover his webcam suggests that Facebook knows something that we don't know. On the contrary, other organizations might be doing this as Oertel [76] mentioned that an organization called Realeyes has created programming called emotional analysis that uses the PC's webcam to watch and dissect individuals' facial responses to ads [76, 77].

Another similar situation is license plate scanners or cameras. Marnie [33] described how various partnerships have set up scanners for license plates around the globe and are using them to gather information on wherever an individual goes and wherever they have been. She further added that the largest organization that was scanning number plates in the US previously had two billion records of tag filters on a document in January 2015. They would package the data they got from filtering licensed together with credit checks, buy history, data on where you live, and data on who you know and offer everything to advertisers. It's accepted that insurance agencies utilize this data to set their rates. If a user has been discovered passing through risky neighborhoods, their insurance agencies can wrench up their rates a tad.

Another aspect of creepiness for online users is health and religion-related ads. Several pieces of research have found that health issues are considered sensitive for most of the users. The argument can be justified by the following research findings. Dolin et al. [30] found that participants were less accepting of inferences and personalization on health themes, though they were progressively okay with points like travel. They were likewise generally awkward with subjects identified with religion. For instance, "Christianity", "Christian and Gospel Music", and "Islamic Holidays" were all subjects for which participants felt awkward with web personalization. Leon et al. [62] also discovered that users were happy with sharing certain classes of data with ad networks while being reluctant to share other classes.

Based on user shared experiences what we understand is that collecting data about users without having their consent and the lack of transparency is considered creepy for them. Also targeting ads based on health and religion is creepy for users. So, now we further compare psychological factors leading to creepiness with user perception of creepiness to find out if online tracking is creepy.

3.2.2 Analysing Factors leading creepiness in the context of online tracking & targeted ads

In this section, we analyse the psychological factor leading to creepiness we discussed earlier in this chapter with the user's perception of creepiness to see if online tracking is creepy or not.

The factor we derived in Section 3.11 of this chapter was 'Observation' without an unclear intention. Now if we examine the process of online tracking, the way it works and the way it handles the data we might be able to connect the gap between creepiness psychologically and creepiness in terms of online tracking.

Generally, online tracking is a process of observing online user behaviour to assess their likes and dislikes, their demographics, their community, and a lot more. As we discussed in Section 2.2 there are various tracking tools available to collect user data such as cookies, web beacons, and browser fingerprinting. We also know that online services are observing us all the time: they observe how we interact with different sites, what activities we do while we are online and all. They observe us to understand our preferences and to influence our behavior. In a way, we know that they are observing us but we are not sure of the intent behind it.

If we analyze the process of online tracking then we can divide the whole process into three parts: recording/observing our online behavior, remembering that behavior, and at last analyzing that behavior. In other words, online tracking is a process of observing, attention, and data collection. Tracking tools like cookies and beacons observe all our activities online: they collect our browsing history, IP address, websites we visited, products we bought, search keywords and lot more. These data are transferred to ad servers to remember users. These data are stored in the form of big tables in the databases of ad networks. Ad servers are used to display ads on websites. They decide which ad should be shown at which place based on the analyzed data. Ad servers interact with both the advertisers and publishers. They also collect data about how many times an ad was clicked and who watched the ad to improve ad marketing. Based on these data an ad profile is made for each online user. These profiles are used to tailor ads for users [57].

The process of online tracking matches the psychological factors leading to the creepiness, as the process of tracking involves observation with an unknown intention. But to come to a strong conclusion we further discuss some further evidence on the creepiness of online tracking.

We saw earlier in Chapter 2 that Facebook observes users on third party sites through like buttons so that it can collect user preferences and can use them to show targeted ads on Facebook. Not only does Facebook observe its users but it also observes users who don't even have a Facebook account to collect data about them and encourage them to join Facebook. Simonite [93] said that

regardless of whether a person doesn't utilize Facebook, despite everything they're observing all that you do. On about each site, there's a tricky little catch that is watching an online user: the "like" button. Facebook has added code to the "like" and "offer" ties that showdown on pretty much every article online that lets it subtly record your actions. Oliver [77] talked about how Facebook used like buttons too track online users in his blog:

"You don't have to click on the buttons for them to watch what you're doing—if you're on a website that invites you to share their post on Facebook, the company is watching what you're doing. They're watching your comments, they're watching where you go next, and they're selling everything they learn. Your behavior on almost every website gets used to make targeted ads for Facebook, Instagram, and any company that pays Facebook for its ad services. So it doesn't matter if you opt out. It doesn't change anything if you boycott Facebook. They're still watching everything you do."

Jennings [54] described why it is the least difficult clarification for targeted ads to be creepy because they know about the telephone one uses, computer, and the internet itself where all is said and done which collects a gigantic amount of data about a user. Google, for example, knows every site the user has ever gone to in their life, and Geolocation can tell where they live, where they work, and where they've voyage [51]. While credit card organizations know what they had purchased, the brands that sell those things can utilize that information to foresee the things they'll purchase later on. They can even that someone is pregnant before even her family knows [31].

Based on the above discussion, user's perception about online tracking and targeted ads seem to be reasonable. Online tracking is creepy because companies observe the online behavior of each individual with an unknown intention of where the data will be used. The lack of transparency leads to creepiness for online users. The process of online tracking resembles the factors leading to creepiness as companies observe the online behavior of users and collect data about them to show them targeted ads. However, based on the discussion in section 3.2 it seems that that the main concern for online users was data retention rather than them been tracked [3, 30, 114, 52]. Users felt creepy because they don't know the intention of ad companies behind collecting their data and also they seemed to be unaware about the amount of data been tracked. They might consider this as a potential threat to their online privacy hence causing the feeling of creepiness as many participants felt that the process of online tracking was invasion in the privacy [29]. They also expressed that the lack of control over targeted ads felt creepy. Based on this, we hypothesize

that the process of data retention in online tracking is creepy for online users and hence we propose forget to disrupt the data retention part of the creepiness feeling.

3.3 Forgetting and creepiness

We have seen that online tracking is creepy because it tracks and collects user data without even knowing them. And most importantly the intentions are unknown behind collecting this amount of data. In this section, we discuss whether or not having control over tracking data can overcome this feeling of creepiness.

As it is observed in multiple users studies that lack of transparency and control have led to the loss of user trust in ad-networks. Consumers develop a negative feeling for advertisers. This negativity can also impact the productivity of a product brand and the advertisers and publishers [74, 103]. As a result, ad companies have decided to provide some control to users in the form of various tools and avoid tracking. Some of the tools include private mode or Incognito mode, Do Not Track and Ad-blockers. These tools can help the user to avoid online tracking and get rid of annoying and creepy ads. But each tool has its own issue or limitation. We have already discussed the tools used to avoid tracking and their limitations in Section 2.5 and we observed that the existing anti-tracking tools are not efficient enough to provide the user with required control and transparency to deal with the creepiness issue. Also, Cranor [28] discovered that the existing tools didn't focus completely on user's requirements of control [88].

Based on our research background, current anti-tracking tools are not sufficient enough to address the issue of creepiness as we discussed their limitations. Hence the issue of creepiness related to data retention is not been addressed yet. Thus we came up with an idea of understanding user perception about forgetting tracked data. Our hypothesis behind solving the issue is that we think that the actual problem isn't tracking but the retention of so much information about the user in an opaque fashion. Based on previous researches it seems that the feeling of creepiness is originated from the process of data retention rather than tracking. That's why we think that forgetting can disrupt the retention portion of creepiness.

Forgetting means allowing the user to remove and delete the stored data about them. There are several opt out options yet it seems that there is not a single one which can solve the issue of data retention, but forgetting can potentially help to do it. It is very difficult for us to make a human brain forget anything we don't like. If we compare the ability of humans and computers in terms of forgetting things than we can assume that we can make computers forget/delete things more easily

than humans. Humans have a network of neurons in their brain in which each cell is connected to thousands of other cells and they are connected to millions of cells. Likewise, computers also have big data where there are different files connected to multiple other files and so on. Deleting a file from a single computer is very easy, as it is just one step process.

However, as the scale and complexity of data increase the challenges of deletion increase. To understand how the data deletion process works in computer data storage systems we first understand how data is stored in the next section.

3.4 Storing User Data

As we discussed in the last section, challenges with deletion of data increase as the scale and complexity increase because big data has lots of data and each piece of data is entangled to other pieces of data. Especially for advertising platforms and social networking sites produce almost terabytes of data almost every day. So it becomes more complex when it comes to deleting data from big data [60]. Hence in this section, we first learn about different storage facilities used to store and control data. We discuss File system, relational database, distributed systems, and machine learning. We discuss how this system stores and control data to better understand the deletion process and issues related to it.

3.4.1 File systems

It is a data structure used to store and retrieve user data. Every piece of data is divided into blocks which are called files so that is it easy to manage the data. Every file has associated metadata which specifies information about the file like size, name, owner, permissions, location. There is various file system which includes a distributed file system, database file system, tape file system, transnational file system, etc. The file system is operated onset of logical rules. Linux has a virtual file system. In Linux, all files have an associated inode. An inode stores the metadata of the file and the location where the data is [35, 102, 65].

To retrieve, delete or update data one needs inode of a file or user to have to specify the block location to delete the data. An inode can also be called an index number [35, 102, 65]. In the later section, we explain how this inode is used for the data deletion process.

3.4.2 Relational database

In a relational database, data is stored in the form of tuples and associated relations. Tuples are just a sequence of elements and relation is a domain set that specifies which value can a tuple contains. Data here is stored in the form of tables unlike files and directories [26, 25].

Tables are organized as rows and columns and to retrieve the data SQL language is used. In the case where there are millions of data stores, it becomes hard to retrieve a single data or delete a single data in that case index comes in place. An index is used to minimize the search efforts for data and to control the data effectively. The index will store the data in a column in a data structure. Whenever a user requests a data deletion from a relational database the data is not deleted instead it is marked as deleted and space is made available similar to file systems [26, 25].

3.4.3 Machine learning

The machine learning process is based on algorithms to perform a certain task. The algorithms are trained on sample data without external instructions to perform a specific task. The training data is used to make predictions and decisions in the future. Machine learning is a core part of artificial intelligence. It is used in recommender systems as well. The inputs are user preferences like products bought, similar items, browsing history to provide personalized ads. The advantage of machine learning models is that we can train different models based on our requirement [85, 64].

3.4.4 Distributed system

Companies like Google, Facebook, Amazon, eBay, and many other big companies track online data whether it is user-specific of ad and website specific. They all have different storage facilities to keep the user data securely. They have chunks of data that need big storage. They store data in petabytes. Around 20-60 petabytes of data are been stored each day on Facebook and Google [10]. This amount of data is stored in a distributed storage systems either distributed file systems, distributed database, and others. Hence, distributed data makes the problem even worse. Moreover, such a big amount of data needs high power and capacity storage servers which can process the data efficiently.

We further discuss examples of storage structures: Google cloud storage and Facebook F4 to understand the theoretical approach of deletion. We also talk about the recommender system and machine learning in this section, however, we don't have any theoretical explanation for the same.

3.5 Data Deletion Process in different data structures

In this section we discuss google cloud storage and Facebook f4 storage systems data deletion process and we also highlight the data deletion process in the recommender system and machine learning system using AI. But we don't have a practical example of a recommender system and machine learning that allows for fine-grained data deletion.

3.5.1 Google cloud storage

Google Cloud storage is an online service where users can access data through the web. Google has four stages of the data deletion process. According to Google initially, when there is a request to delete data, the very first stage is to mark the data to be deleted. Once the data is marked to be deleted the second stage is often called a soft deletion in which the recovery period of the marked data is confirmed to see the data has enough recovery time before getting deleted. Usually, each data has a backup so there might be two or three replicas of the same data. So they make sure if the data can be recovered on time which needs to be deleted. The third stage is deletion from the logical system. There are two ways to do this: Overwriting the deleted data or using cryptography. This technique deletes all the keys associated with the encrypted data to decrypt it which makes the data unreadable. The last stage is expiration from a backup system which also uses the same technique:- overwriting or cryptography erasure. So actually the data is not been deleted but is either overwritten by marking the space of data as available or by deleting the keys associated with data for decryption making data unreadable [45, 46].

3.5.2 Facebook F4

Facebook's haystack stores BLOB(Binary Large Objects). It makes 3 replicate of each data to maintain high fault tolerance. There are some data on Facebook which are old and not very frequently accessed. With the increasing growth of users, the requirement for space also increases. To deal with unnecessary space required by not frequently data Facebook divided the BLOB storage into two parts: Hot BLOB which is frequently accessible and Warm BLOB which is less frequently accessed. Warm BLOB is managed by F4 which has a low replication factor with the same fault tolerance but this saves a lot of space [10]. Birk et al. [14] mentioned that in the existing cloud storage system there is no way one can verify that the requested data was deleted or not.

Deletion in F4 storage is handled by cryptography erasure. Every data has a separate key stored

in a separate database. The data is encrypted and the key is used to decrypt the data. If you delete the key the data is gone. Technically speaking then the data is not deleted but deleting the key will no longer provide access to the encrypted data so they define it as been deleted [73].

3.5.3 Recommender system

E-commerce sites have been using a recommender system to improve the user experience of shopping and to improve their sales. The recommender system works on different user inputs. The input data for such systems include user's purchase history, Likert scale about various products, comments on any products, editor's choice of products, demographics, products bought by common friends, etc. The system takes all this as input and analyses it to recommend either a similar product to what user what searching for, a similar product which user bought, products bought by users friends or families and products matching the users Likert scale for other products. When it comes to deleting data from the recommender system it is even harder because of the intermingling of data [90].

3.5.4 Machine learning and AI

Moreover, Ginart [41] explained in his paper that cryptography is not actually deleting the data but it is instead making it unidentifiable [18, 20, 17]. Ginart [41] further discussed the challenges associated with deleting data from a machine learning model. He said that most of the machine learning model needs to retrain the model from scratch when they have to delete a single entity from the model. To better describe the problem of data deletion in machine learning he considers an example of a patient database that is used to diagnose diseases based on different symptoms. If we need to delete a single patient record let's say it i th patient from n number of entries(patient) then basically we expect to update the model with $n-1$ entries of patients because we want to delete an i th patient record. But in reality, to achieve this most of the machine learning models have to retrain the model from scratch with $n-1$ entries which is time-consuming and costly. And it is not feasible to train the whole machine learning model when it is comparatively large as it would take weeks for data to update and it would take up so much of energy and cost.

It is seen that deletion is always an issue and data distribution makes is more difficult. We checked the data deletion process, so further we highlight the technical challenges associated with the techniques: cryptography, data overwriting and AI used to delete data in all the above data storage systems.

3.6 Challenges in Data Deletion Process

We learned how data is actually not deleted by either encrypted or overwritten deletion techniques but instead it is made inaccessible. In this section we specify what are the issues with these techniques.

3.6.1 Cryptography Erasure

Data are encrypted as they can be saved from malicious access or modifications by attackers. Most of the data structures encrypt their data to maintain the confidentiality of the user's data. When the data is encrypted a key is generated which can further be used to decrypt the data [72]. This process is also called cryptography. In companies maintaining public data, health information or any confidential information it is mandatory to encrypt the data as per government regulation to secure the data [89, 81]. When one needs to delete data in such a system they delete the key associated with the data which is called cryptography erasure. In short, data can be deleted by 'forgetting' the encryption key for the respective data [19]. There are various limitations of the cryptography erasure technique of data deletion. While this process makes data inaccessible but it also adds overheads as it needs more efforts for key management. Managing keys for a big database system could be costly. Moreover, some small business does not use encryption and hence it is not economically feasible for small companies to do cryptography erasure. Also, the process of encryption can increase CPU overhead. As cryptography erasure method deals with keys it has some drawbacks related to keys as keys can be corrupted or compromised and can be used to recover deleted data or existing data and keys can be easily lost or stolen [81, 113, 58].

Apart from all these limitations, there might be other issues as well if a system has data with various versions then deleting the key for data would not be feasible as it would be needed in the decryption of the data in future versions. Thus in a shared data structure, it might need to have a key for each shared block creating a burdensome amount of keys that can become unmanageable and at the same time costly [82]. At the time of writing a file with a new key, there are chances that the old data can be recovered even with the new key using special hardware tools [81, 113, 82, 58].

Moving further we enlist limitations with data overwriting.

3.6.2 Data overwriting

Another method used to delete data is to overwrite the data that needs to be deleted so that no one can access it but it also comes with certain limitations as it would not be reliable in cases where the size of data is large because overwriting large data files or overwriting data several times can increase CPU overhead and can ultimately affect the performance of the system [82]. Secondly, some system tools require users to overwrite data of the same size. Users should add new data which is equal to the free space available. So for example, if a deleted data has made a free space of 1 GB in a system then to do an overwrite of these data users should overwrite the entire 1 GB. Such a requirement can be time-consuming and also cannot guarantee the overwrite of data. Thus, it can make the data still accessible even after deletion [11, 37]. We should also consider that while deleting data, usually the system neglects the metadata associated with it especially in the file systems. Hence even after secure deletion, an attacker can get some important information from the metadata [58]. Moreover, in a system where data is shared between multiple blocks, overwriting data in the previous version ultimately changes the data in future versions. Hence to delete data securely without affecting other data first the system needs to find data dependencies of particular data which is time-consuming and takes up a lot of effort [82].

3.6.3 Artificial Intelligence

Artificial intelligence has several advantages as we can train different models as per our requirement. But there are some challenges with artificial intelligence when it comes to data deletion. As we discussed earlier, data is entangled in AI models based on patterns in data and how they relate to input and output. As discussed earlier to delete data in such a storage system, users have to retrain the entire system after removing a single data which needs to be deleted. We discussed an example of deleting a single patient entry above. Hence it is not cost-efficient. It is also time-consuming as it might take several days to complete the retraining of big models [41].

Hence, we can conclude that there are many challenges associated with the data deletion process but at the same time data deletion is not infeasible. As we saw that Google and Facebook used these techniques to delete data. Our only aim to highlight these challenges is to provide a sustainable research background in terms of forgetting data from a data structure. So, all in all, it is possible to forget/delete data as we discussed in section 3.5 but there are just some challenges which we believe can be overcome in future research work as we provide a strong background.

3.7 Inspiration behind the research study

Although there are several challenges associated with the implementation of forgetting or deleting data from such big data storage our aim is to understand whether the user would need something which can 'forget' their tracked data at first place. Hence our research question is to find if giving users fine-grained control over retention of online tracking data (related to specific targeted ads) would change the acceptability of online tracking. Since now no one has thought of understanding user perception about 'forget' to reduce the creepiness related to data retention so we are curious to know what users think about forgetting tracked data. If the user feels that forgetting data can help them reduce creepiness than future research can be made on this basis to implement it technically. But as we don't know about the user's perception of 'forget' we feel that this research study is necessary. We also think that this study can be a basis for future study on understanding creepiness and data retention.

Chapter 4

Methodology

In this chapter, we discuss the methodology of our research. We have three sections: research question and approach, the methodology of the first study, the methodology of the second study and qualitative analysis. We discuss our research aim and approaches to fulfill our aim. We also discuss the methodology used in both studies in separate sections.

4.1 Research Question and Approaches

Our research questions revolve around the user's acceptance of online tracking and targeted ads if they are provided more control over it. We aimed to understand user perception about 'forget' as an opt out option. To better understand that we also had some questions which were focused on user's knowledge about online tracking, targeted ads, and anti-tracking tools. However, our main research question is:

Will users be more comfortable if they were given the ability to control tracking data? Will the ability to delete tracking data make the user more comfortable about online tracking and targeted ads?

To address this research question, we proposed 'forget' as an opt out option from creepy targeted ads. Specifically, we included forgetting in the existing YouTube ad feedback system [115] which we have shown in the mock-ups. The purpose of 'forget' is that it allows the user to make ad companies forget the data which leads to a particular ad. To be the more specific user can make the ad companies forget the searches related to the ad. For example, is used is shown weight loss ad based on user searches and websites visited than choosing to forget should remove these data from the ad companies.

To understand the user's perception of our proposal, we conducted two user studies with an overall of 22 participants all adults over 18 years who can speak and read English. The study design for the first study included four rounds of questions, watching an informative video that was used by Blase et al. [105] study on perceptions of online behavioral advertising, scale questions and open-ended questions. In both the studies we defined creepiness in general to the participants

before we started the first round of open ended questions. The scale questions were provided on paper to the participants and the responses for the scale questions were also collected on the same paper while the open ended questions were audio recorded. We used mock ups to represent the idea of forget to the participants. The mock ups were in the form of paper. Our first study included screenshots taken from YouTube which displayed YouTube ad and ad-feedback system. The purpose of these screenshots was to understand users awareness about ad opt out options and what they think about forgetting their tracked information from ad networks. But the first study was inconclusive on the question of forget, and we think it was because we were vague about forget about how forget would work, so we designed a second study with a specific forget interface design. We revised our mock ups and added screenshots from the Google Ad Setting page [47].

4.2 First study

In this section, we describe the process of recruitment, the study design and we also talk about the information video. And hence we have divided this section into three parts: participants recruitment, study design, and information video.

4.2.1 Participants recruitment

The study was reviewed and cleared by the Carleton University Research Ethics Board Committee. We recruited participants using Poster, online recruitment and email recruitment. The posters were posted on Carleton Universities billboard as per Carleton's posting policy. I contacted the Carleton Research Participants page for online recruitment. After the participant contacted the researcher, we sent them the email invitation and the research consent form.

4.2.2 Study design

The study was conducted in the Carleton Computer Security Lab in a quiet room. Before starting the study, the participants were asked to read and sign the research consent form (See Appendix A). They were also asked if they wanted to provide consent for audio recording. All the participants but one agreed for audio-recording. The researcher explained the design layout to the participants and asked them if they had any questions before proceeding. The researcher made it clear that if the participant feels uncomfortable providing the answer to any study questions, they can say no. Participants were asked about their age and recent education qualification and how much time they

Representation of 'Forget' in Study 1

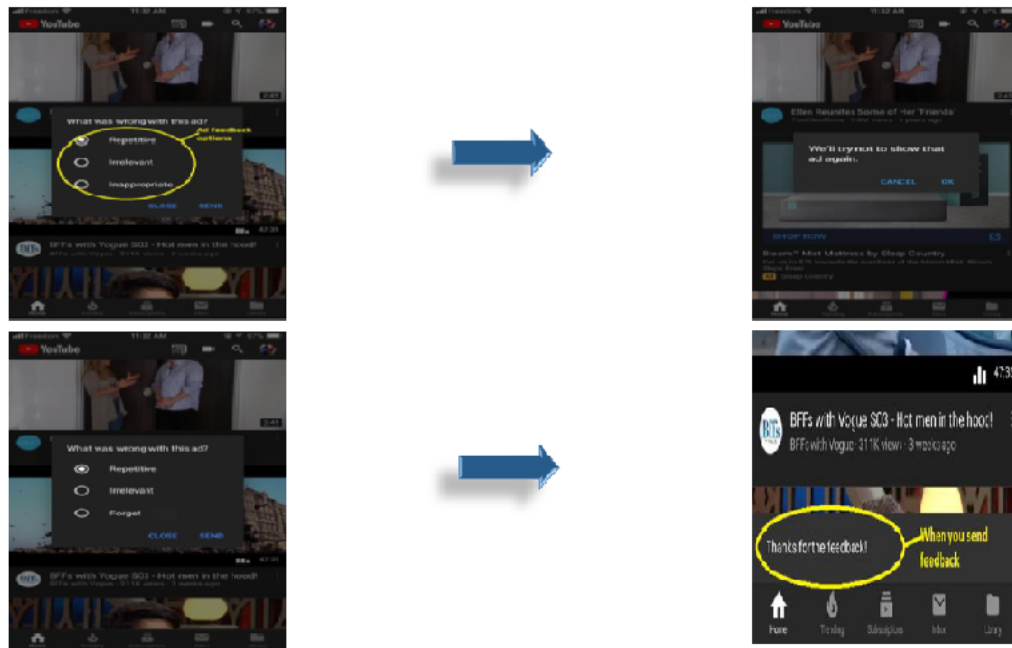


Figure 4.1: Representation of Forget in Study 1

usually spent online.

The first round of questions mainly focused on understanding users knowledge and awareness about online tracking, targeted ads and what was creepiness for them (See questionnaire 1 and 2 of Appendix A). We explained to the participants what was creepiness in general so that they would express themselves when they were asked questions about it. First, they were asked to provide their responses to scale questions in a written form. After each round of scale questions, we asked open-ended questions. After the first round of questions, we showed them an information video based on cookies and tracking from Wall street of Journal to give the participants better insight into how online tracking and cookies work. The video was approximately 7 minutes long which participants watched on an Apple iPad 9.7 inch. This video was also been used by Blase et al. [105] in his study of Online Behavioural Advertising. After the video was shown the participants were asked to respond to scale question which were based on the video and following that they answered open-ended questions which were also about the video (See questionnaire 3 and 4 of Appendix A). Basically, the second round of questions aimed to understand user perception about tracking after watching the video. We wanted to know if their views were changed after getting the information in the video.

Post second round we showed them a set of mock ups, a summary of which is shown in Figure 4.1 (full mock ups are in Appendix A). Each mock up is a set of screenshots taken on YouTube on an iPhone 8 plus. The aim of mock up was to show them how the YouTube ad feedback system looks like, how google ads were personalized, which data the ads were based on. After explaining them the mock up we repeated the task of questions (See questionnaire 5 and 6 of appendix A). After the third round of questions, we wanted to understand user's awareness about opt out options, ad-feedback systems, and ad personalization. So the second mock up had a slight change in the design of an existing ad feedback screenshot. We removed the third option from the ad-feedback option and added forget as an option by editing the screenshot. We didn't intend for them to understand the difference between the ad-feedback option but we intended to explain to them the purpose of 'forget'. Hence the last round of questions was based on understanding user's perception about forget as an opt out.

We further explain in detail why we chose the information video in this study.

4.2.3 Information Video

We used this video from Wall street journal [104]. The video explained cookies, web tracking, third-party tracking, a targeted advertisement which is all about our research study. It explains what was the main idea behind cookies. It explains the function of cookies with few examples and how big the ad network can be. This video was the right choice as it had all the information we wanted the participants to be aware of. So we asked them to watch this video even though they knew about tracking and cookies, because if some of them were not aware of cookies and online tracking they could understand it through the video as further questions were based on this. We wanted to provide a baseline of our study to the participants and this video had all the information. It was easy to understand and it gave examples to explain each of the above makings it a perfect choice. This video was also been used by Blase et al. [105] in his study of Online Behavioural Advertising. Getting inspired from there we selected this video as it best matched our study requirement. It was about seven minutes long.

4.3 Second Study

The results of the first study were not inconclusive because we think that we could not explain the purpose of forget to participants. Hence we made some changes to the mock up design so that

forget was well understood and conducted the second study. So in this section, we discuss how second study was conducted. We have two subsections: participant's recruitment and study design.

4.3.1 Participants Recruitment

We revised the recruitment documents used in Study 1 as you can see in Appendix B. We used the same method of recruitment as we used in Study 1. For the second study, we also submitted a change to the protocol form to the ethics board at Carleton. This study had three rounds of the questionnaire which only included open-ended questions. We didn't have scale questions for this study as they were inconclusive in the first study. We asked users to examine mock ups and screenshots. We removed the introduction video. We took screenshots from the Google Ad Setting page to show them what data exactly is stored within the Google Ad page and how they can opt out of ad personalization. Instead of placing forget in the ad-feedback system we placed a link at the bottom of the YouTube ad which says 'Click to forget searches related to this ad' aiming that it will make accessible for users. The revised questionnaire, mock ups, and screenshots are included in Appendix B.

4.3.2 Study Design

Same as the first study, this study also began with asking user questions about their age, their recent education qualification and approximately how much time they spend online. For the first round, we discussed open-ended questions about online tracking, targeted ads, creepiness and any opt out options the user knew of (See Questionnaire 1 of Appendix B). Following that, we showed them mock ups, summarized in Figure 4.2. First we showed them mock up 1 which included screenshots taken from YouTube from an iPhone 8+ (See Mock up 1 of Appendix B). The mock up showed them a YouTube ad and Google Ad Settings page in detail. The mock up also had an explanation on each screenshot. The researcher also answered questions of participants during the explanation of mock ups. We had a screenshot of two YouTube ads: credit card ad and nursery rhymes and baby song ad (See Questionnaire 1 of Appendix B). We also displayed the information on the Google ad settings page about the respective ads. We showed them how Google has creates an ad profile and the possible data collected about an online user. We further showed them how credit card and parenting was part of the ad profile. Based on these mock up 1 we had followed up questions. After that, we showed them mock up 2 (See mock up 2 of Appendix B) which included the 'forgot' link beneath the YouTube ad. We edited the screenshot used in the previous mock up to make the

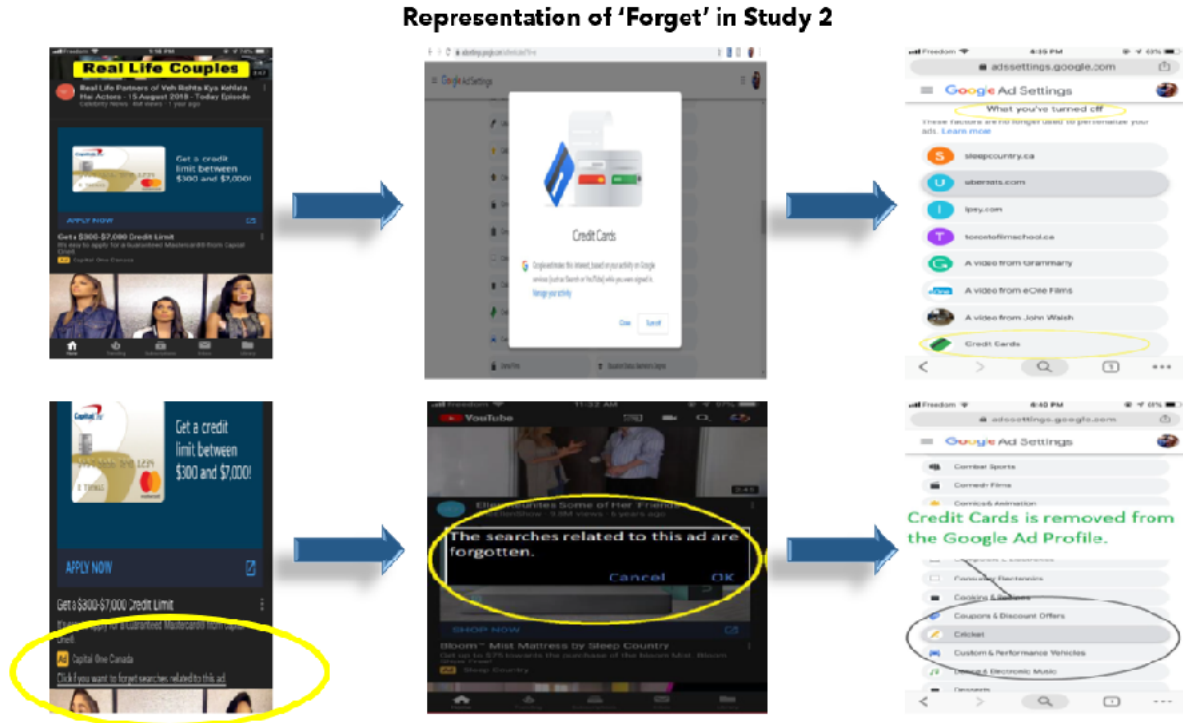


Figure 4.2: Representation of Forget in Study 2

difference more visible for users. Through this mock up, we explained to them how by clicking on the forget link the information from their ad profile which is stored in Google Ad Settings is removed. We also showed them how the stored information about credit cards and parenting was removed from the ad-profile within Google Ad settings after selecting the forget link. Finally, we asked participants a few questions about how they felt about the concept of forgetting in terms of controlling tracked information and creepy targeted ads.

4.4 Thematic Qualitative Analysis

We used thematic qualitative analysis for both the studies to evaluate our data sets. Thematic analysis includes identifying themes based on the qualitative data [66]. The reason of choosing qualitative analysis was that we had open ended questions for both the studies and thus most of our data was qualitative rather than quantitative. Also, we felt that thematic qualitative analyze will help us efficiently portray user's perception. The source of all of our data were the written responses of scale questionnaire and the transcribed audio recording for open-ended questionnaire. We almost transcribed around 10.5 hours of data. We printed the transcripts to analyze them. We

first tried to find codes out of the transcribed data for key points such as online ads, online tracking, creepiness, forget and others. We iterated through the transcripts to find themes for each key point. We iterated the entire transcript for almost 8-10 times. We also did a final iteration to compare our obtained themes with the participant's responses to find similarities. This process was conducted by the researcher on their own. The entire process took approximately two weeks. In Chapter 5, we discuss the analyzed data.

Chapter 5

Results

In this chapter, we discuss the results we obtained from both of the user studies. Our aim was to understand users' perceptions about online tracking and targeted ads and to see if having control over it can change their acceptance. We divide the results into different themes to explain users' perceptions about online tracking, targeted ads, creepiness, existing opt-out options and forget mechanism as below.

We have divided this chapter into two sections: results from the first study and results from the second study. Within each section, we have different themes based on the participant's responses. Each theme aims to explain the findings of our study. At the end of this chapter, we compare the results of both the studies in terms of 'forget'.

5.1 Demographics

We recruited a total of 22 participants for both our studies. We had 13 participants for Study 1. The age group of the participants in Study 1 was 20-28 with an average age of 23 years. Majority of our participants had a computer science background (5 were from computer science background), while we also had one participant from economics, finance, health care, chemistry, literature, law, public affairs, and management each. Their average time spent online was recorded to be 5.41 hours a day with a minimum of 1.5 hours and a maximum of 12 hours. We also noticed that eight of our participants were undergraduate students.

While during the second study we had 9 participants with the lowest age of 20 and the highest age of 45. The average age of the participants was recorded as 27 years. Six of the participants were pursuing a job while 2 of them were undergraduate students. Their average time spent online was recorded as 5.8 hours with a minimum of 1.5 and a maximum of 12 hours.

The participants in both the studies had minimal understanding of online tracking and targeted ads as it was one of the requirements for participation in the study.

The demographics for both the studies are listed in Table 5.1 and Table 5.2, respectively. We further discuss the results we obtained from both studies in the rest of this chapter.

Participants Pseudo Code	Age	Time spent online (Hours)
P101	20	5
P102	22	6
P103	22	5
P104	22	10
P105	20	3
P106	22	4
P107	27	4
P108	23	4
P109	26	3.5
P110	22	7.5
P111	27	3.5
P112	28	5
P113	23	5

Table 5.1: Demographics of Study 1

Participants Pseudo Code	Age	Time spent online (Hours)
Q101	20	2.5
Q102	30	5
Q103	28	1.5
Q104	23	12
Q105	25	2
Q106	29	12
Q107	45	2
Q108	24	5.5
Q109	27	10

Table 5.2: Demographics of Study 2

5.2 Results from the first study

In this section, we discuss results we obtained by analyzing open-ended and scale questions. We chose thematic qualitative analysis methods for analyzing our results. We categorize the results in various themes and at the end of this section we also discuss scale responses.

5.2.1 Are targeted ads really of any help?

Participants were asked what comes in their mind when they hear online ads. Most of our participants replied that they found ads annoying or frustrating. P107 said: *“Online ads oh my gosh...The first thing that comes to my mind is annoying. Very annoying.”* P103 also stated his frustration for online ads: *“Annoying. Just because most of the time it’s something that’s not related...most of the time. Oh, sometimes it’s related to what you kind of search. But that’s kind of annoying. Also sometimes it’s very random but it’s annoying.”* P105 also felt the same as it came in her way while she wanted to access data on a site: *“The first thing that comes in my mind when I hear online ads is very annoying. Each time you need to wait for a few seconds and minutes and then skip the ad. And sometimes its also appearing many times on the screen and sometimes you have half of your screen occupied with ads so a bit annoying sometimes.”* P111 was really angry as no matter what he used to block ads they would still display and animate: *“First thing that comes to my mind...I would say is graphical bloat. If I pull a web page and if I am not using an ad blocker some of these web pages do their best just to cram as many ads on the page possible. They are quite often either moving or flashing or doing something just to bloat that page and make it like...put in the way of content...and they are just ugly.”*

Other participants explained how ads would occupy the whole page and would come in their way to access web content. P101 said: *“The first thing in my mind comes when I hear online ads would be Banner ads on the web pages. The one takes the places on the entire page like in the headers and bars that sort of thing.”* On the other hand, P109 was frustrated because the ads she would see would not be relevant: *“The ads on Facebook so the ads on social media, the ads in email, the ads when you go to a shopping website and the advertisement in the sidebars or pop-ups and stuff like that. They are pretty annoying because most of them don’t relate to me and I am like why am I seeing this it doesn’t apply to me or relate to me.”*

Participants were furious as they would see ads everywhere. They explained where they would see most of the ads. Some participants explained that they would see more ads on social media

while others said that they would find more ads on random sites. Participants P105 expressed how even buses are not left for displaying ads: *“Everywhere. Everywhere you go, any websites. Sometimes ads with money they do advertisement on YouTube and even on the bus so everywhere.”* While many other participants felt that they see most ads on their social media accounts such as Facebook, Instagram, YouTube, and Twitter. P10 told: *“I use twitter a lot so there are lots of ads on twitter but that one is promoted or sponsored. And they say promoted up the bottom. Those are really annoying. And then some ads for online shopping too. You see them on other sites.”* P109 supported the same sentiment: *“Social media and much pretty much social media is the big one so like Facebook, Twitter, Instagram.”* On the other hand, P111 observed that he gets more ads on random sites rather than the one he knows: *“Not the sites that I would visit the most. I probably see the most ads when I am just trying to find certain information on something. Like for example if I am on Facebook for better site pick I don’t see a ton of ads but there are ads on Facebook but I really see the ads that really bother me when I go to like...Let’s see if I am looking for a temperature on which I need to cook chicken thigh...I google what temperature to cook chicken thighs at and I go some cookingabc.com they will cram like a ton of ads down my throat and put them all over the place. Not the places that are too frequent but places where I go to for information. The places where I don’t necessarily go very often.”*

All in all, online ads are annoying and frustrating for participants as they block a lot of space on-site, they would see ad everywhere they were not related, the would come in a form of animated, often pop-up ads.

5.2.2 What is online tracking: Companies use cookies to track

Participants seemed to have knowledge about cookies when asked about online tracking. They were aware that cookies track their data. When we asked participants if they knew about online tracking and what they knew P106: *“Oh yes. So companies use cookies so you know if you are on the websites there are cookies and they track what you have seen and what you have clicked at. And then, later on, they use that to show you ads that you will be interested in.”* P101 also explained his concept of cookies however he wasn’t sure he explained it very well: *“I don’t know too much about online tracking but I guess some sites you do ask or rather when you do visit them, they save cookies in your browsers like temporary cached information. One thing I could think of would be some shopping or commerce sites have referral programs so if you sent a referral link to a friend then that information will be cached into their browser even if they revisit the site later.”*

Let's say within a day or week after they clicked the referral link, they will still get the credit even if they don't have the same URL kind of deal. Um, and the other more common one you hear about would be with share button or like Facebook button." P113 was also aware of the cookie concept and interpreted correctly as he mentioned: *"I am vaguely aware of online tracking. I think that the web pages keep cookies on your computer and they use it to track you and then not sure if it's related but websites can find your IP address and also pages you visit."*

While on the other hand, some participants thought that an algorithm tracks them as P110 said: *"I don't really understand it. I just know that some algorithm is tracking me but I am not too sure of how it works."* And P104 talked about artificial intelligence being behind all the tracking. He added: *"I think the best example for this would be Google so whenever I use google I am signed into my account and Google wanted all of my search requests. They have of course have my location data through my phone and everything. And they basically kind of connect them and then use AI and machine learning and all that to see what does this guy likes and maybe create a profile. And then finally they target ads based on the profile."*

It seems that half of the participants were aware of how online tracking worked with cookies. But at the same time, they were unaware of other tools used for tracking like beacons, fingerprinting, etc. Although, participants knew about tracking, only a few could explain how cookies work.

5.2.3 It is Creepy when I realize the amount of data they track without my consent

8 out of 13 participants informed that they felt creepy when they saw an ad that was based on their communication. They were surprised that their communication was tracked without their consent and hence they felt it was creepy when they saw ads based on their communication and thought about the amount of data recorded about them. P106 shared an incident that happened with her based on her communication with her friends: *"It's so creepy when I have been talking about something or thinking about something and I haven't looked it up and it showed up somewhere. It's like high level creepy. I can't remember but I was talking about something with my friend's evening and it showed up the very same evening...It was something from Amazon. I can't remember what it was though."* A similar incident was mentioned by another participant: *"I am aware of online tracking. In my words, what happened was that I and my friends we were four friends and we were talking about yellow dresses and suddenly my friend who was just scrolling down on Instagram found a sponsored ad which literally said buy yellow dresses. We were all spooked. I mean it's so*

weird that if you and I were talking about the water bottle and suddenly if you go to insta and find a water bottle ads it's so weird. It happened 3 years ago but we still talk about it. I mean it was that we were shocked that is our phone listening to us or what." P104 also quoted a similar incidence where he mentioned: *"Creepiness for me...A good definition would be me and my sister would talk about something or a product so and we were just talking about it and then she put her YouTube on the Roku TV which doesn't ad blockers so you need to see all the ads and so that product ad comes in. I mean it could be a strange coincidence because there is no proof of google recording us but still that's very creepy. That's the most intimate definition of that."*

Apart from users perception about companies listening to them, participants were also worried about the fact that they don't even realize how much of their behavior is been tracked without their consent. Many participants expressed their concern by describing past incidents. They felt that it was unethical for companies to track them without their consent. And they are shocked when they see an ad based on data they didn't intend to share.

P112 felt tracking was unethical as she spoke: *"Online tracking...it is creepy. Because I don't know I don't want them to track me and also they don't have my consent. I didn't accept or gave them permission to follow me like tracking me. So that's just not even creepy but it's unethical."* She further shared a true story to explain why she thinks that tracking is becoming creepy and unethical. Do you feel tracking is creepy? To this, she replied: *"Yes because now they know what I like or don't like. I have example of why is it creepy because I am not sure if you heard of it or not but it was in the news here in Canada one secondary school I think she was 16 or 17 something and she received like the pregnancy stuff an advertisement at her parents home because she was still living with her parents right... And her dad so pissed and he got so mad that he went to Walmart and be like hey my daughter is just like 16 so why would you send her the pregnancy stuff, vitamin, and stuff, why would you do that? And after a month he went back and apologize because she was pregnant and because Walmart was tracking her they knew that she was pregnant before she knows. That's very scary."* Another participant also mentioned his concern: *"Online tracking is creepy. It's data that I don't intend to give but I have given. Instead taken without my consent now that said I mean if there is a website they need to have some level of recording for example if I click on the link they need to know where this link is taking me to or which product is that so that they can show me the product I request but that's my intent. If they collect data and associate data based on history and things like that and that's where I draw a line of creepiness."* P102 believed tracking is creepy because it is an invasion of her privacy: *"I do feel it's creepy. Because it seems*

like companies, ads and those running ads know about you, what you are watching and so it's weird. It's a kind of invasion in your privacy. Someone out there knows about you what you do online, websites you visit I mean it can be anything."

Participants also felt tracking was creepy because they were shown ads based on personal information they did not think would be. P107 said: *"Really recently actually this is something very personal. I got an ad recently just mentioning tampons and pads that were like really creepy and very personal. I don't know about tracking much but they knew about my gender too so that was very creepy I must say."* P113 talked about his exact location been tracked: *"Sometimes when they have the proper town like I get this ad saying that singles in your tiny little town because I am not originally from Ottawa. My parents and family stay in a small town that usually [nobody has] heard of. So ads like find singles in [redacted] then that's creepy."* P108 also shared a similar opinion when he said: *"When they know too much information about me. Let's say like even home address or my age exactly or my name that's creepy. So I have this google survey app which gives me money for question and so sometimes the questions are very specific to me like "oh, what do you do with this car" and then when it's the car I have so it's a little strange that they know exactly what car I have and whether I am living alone or not. And so that's a little creepy."*

Overall, we understood that our subjects felt that tracking was creepy and unethical because ad companies collect data without user concern, they have access to too much personal data, users suspect that companies listen to their conversation based on their ad experience and they display ads based on data which are is personal.

5.2.4 To avoid ads: Ad blocker is universal

In this section, we explain what the user knew about ad-blocking tools. When they were asked if they used any ad-blocking tools, half of them replied that they used ad blockers while other half were aware of ad blockers but never used it. One of the popular ad-blocker product which was mentioned by many participants was ad-blocker [1]. While some of them also mentioned u-block origin [53], ad-block plus [79]. P107 said that she didn't use any ad blocker but she wishes she could have learned about it as she spoke: *"I have never used any ad blockers. But I have a friend actually who uses it I should have asked her more about it and she is a computer science major as well but I just didn't ask her."* P102 also belonged to the one who didn't use it but knew about it as she said: *"I have heard about them but never used them."* While P103 had difficulties finding the right one as he quoted: *"I tried to download one online but that wasn't too good because apparently*

there is like YouTube ad-blocking tool which one of my friends told and I tried to research it but I had trouble with finding the right one which blocks it so I didn't really go. So no experience not really."

We also had other participants who use ad blockers. They use an ad blocker on daily basis but at the same, they also highlighted the issues with it. P111 gave a statement highlighting the issue with ad blocker as: *"Again U-block [53] one I use. Issues nowadays are that companies or websites are able to see if things are been blocked and they say that you cannot use these websites unless they are disabled. So at that point, I go to different websites."* P104 also had a very similar response when asked about ad blocking tools, he said: *"Ad blocker that's the standard one. I mean, I love that software and I make sure that I install it on everything. I haven't had many problems at all. Actually, there are a lot of news sites that say that oh you have an ad blocker please disable it. and what I usually do is that I just go to another site."* Along with these participants we had P106 who shared similar views on ad blockers and she quoted: *"So I use ad-block plus and then ad block plus for YouTube but every once in awhile I will come upon a website which says that we notice that you are using an ad blocker we need you to turn off to read this and that's a little frustrating. But I also understand that their revenue but there is a fine line."* While P108 discussed how ad-blocker [1] still allowed some ads: *"I have used a few. There is one called Ad-Blocker [1] and that one got deleted and became I think U-origin [53] and I use that now but it still allows certain ads to come up online but overall it does block [most] ads but not all of them I would say."*

So overall we can say that everyone knew about ad blocker even though they didn't use it. The ones who used it also explained the limitation associated with ad blockers which included access to web content on some sites.

5.2.5 Flash Cookies: I didn't know those

After the participants were shown the video, they were asked if it influences them to change their perception about online tracking and targeted ads and we couldn't believe that most of them knew about what was explained in the video. P112 mentioned that: *"Not really because I knew that they are tracking me but I didn't know about cookies but now I know more. It was shocking I didn't actually know that they are saving whatever like every single page I am going. I knew that but I didn't know that they are saving and sending to each and every page I go."* P111 also said that he had no influence from the video because he already knew what was explained there he responded: *"No not after watching the video. I don't like them as much."*

There were other participants who were unaware of respawning of cookies explained in the video. P105 realized why was she getting the same ads after deleting cookies: *“I already tried deleting my cookies but it didn’t work well. So I think I know now and also I feel that people are not very well informed. Even if we delete cookies there is not much we can do.”* P108 also thought the same: *“Not personally just because I knew to some extent that’s what they did and I kind of like that the specific ad is related to what I like because if it was random then it would be more annoying than anything else. So not really. But I wasn’t too sure of the cookies that they can track you even if you delete the cookies. So I thought that once you delete the cookie they really don’t know what preference you had.”* Another participant P113 quoted: *“I don’t think my view as an overall opinion on them has changed but there was interesting information in there that I don’t know about like how they worked but I don’t think that is significant enough to change how I felt towards the targeted ads. I found interesting the difference between the original cookies and the third party cookies how the ad networks tag their cookies on top of the other one. And that when the ad pops up on different websites that’s because they are from the same ad network using that third party cookie which they tagged on. Also, flash cookies I didn’t know those were things and that can spawn like that.”*

All in all, we found that the user knew about cookies and how it worked as we derived from our first round as well. But they were not aware of the respawning of cookies. After watching the video user felt that the only way they knew was deleting cookies which also don’t work. They realized that even after deleting the cookies they are re-spawned as shown in the video so they felt that they can do nothing much to control the tracking. They felt helpless.

5.2.6 Google opt out: No I wasn’t actually aware of it

We showed the participants screenshots of an ad feedback system of YouTube which had several options including ‘stop seeing this ad’, ‘why this ad?’ and further option to navigate to Google ad personalization page to control the ad. When we asked questions to know if participants were aware of ad feedback or they did they ever click on these options we found out that they were aware of the options stop seeing this ad but they never knew that they can navigate to the Google ad personalized page and change the settings. P107 was shocked to learn about google ad personalization option and she responded: *“Never. I was very shocked that I can actually go into it and edit that setting. Oh my gosh, I still can’t believe it.”* P112 also had a similar reaction and replied: *“No. I had no idea that this exists. I never heard of it and I never checked anything.”* P108 had some information

about adoptions but never went to the page as he spoke: *“I went into it to a certain extent but I didn’t really dive into it to know why it was there. I didn’t know that Google has this page where I can see the ad personalization or I didn’t look into it.”* P103 talked about the drop-down menu for YouTube advertisements. He quoted: *“Probably I have seen the three dots but I have never clicked on it. I never knew about it.”*

Further participants were asked if ad feedback shown in the screenshots was useful. And to this most of them partially agreed. They thought that this was a good initiative; however, it is not going to help them much in controlling creepy ads, because they are going to see other ads. P106 laughed when she was asked if ad feedback will help make ads less creepy and she added: *“Less creepy...Laughing...Probably not but it might make some easier to deal with.”* P104 thought that it will make it more bearable: *“It’s not gonna make it less creepy but I think it will make it more tolerable for you. Like I mean if you don’t wanna hear the O-train opening on the 14th, 50th time then you can reduce those kinds of ads then ya you know it [works] that way. But at the same time, it’s equally creepy because you are going to specify your location and they know that you are in Ottawa and so they send you this o-train ad so yeah.”* P113 shared that it will just refine the tracking algorithms: *“Hopefully. Um...yeah I would like to think that it would. Realistic mostly it just means that they are refining their algorithms. But it would appear less creepy from looking at the ads I suppose.”*

To summarize, subjects were not very aware of existing YouTube ad feedback and the Google ad personalization page. They also believed that these resources wouldn’t significantly reduce the creepiness of ads. It is hard for them to guess where the opt out options are located and how they can navigate there. Even after giving feedback or changing settings they are not going to stop tracking.

5.2.7 Some level of control helps

The participants expressed their desire to control the amount of online tracking and targeted ads. When we asked them if they would be more comfortable if they were given control over tracking and targeted ads, to this P113 expressed his desire to control his data used for targeting ads and added: *“Yes definitely. Mostly just because I do see the benefit of them having that information. Like targeted ads when they work they are useful so it’s not feasible for them to do with no information but then there are certain things which I don’t want them to know about with things which shouldn’t be commercialized that’s really not the right word but that’s like the idea that’s*

something I don't want them to try to make a profit of it." Another participant P104 expressed his concern about how users are unaware of the amount of tracking is done and the amount of data that is being collected. He shared his thoughts on the same: *"I think the creepiness with targeted ads and online tracking has to do with you don't really know to what extent they have your information. I am sure Google has publicly that we are just storing your age and like things to just make conformable about things but who knows what they are actually storing and there is this idea of big data and bulk storage information and things like that so we don't really know the extent of what they are tracking so that's something they are contributing to the creepiness factor. So personally obviously having some level of control helps."* Other participants including P107 and P110 agreed that they would be more comfortable even with a little control. P107 quoted: *"Oh yeah absolutely. And now that I understand like how ad tracking works now I will be a little bit more comfortable if I have more control over what information they will take regarding me."* While P110 said: *"It would make me more comfortable for sure if you can pick and choose. Like if they couldn't have my location or my age but they know what I am shopping for maybe so that it can stay in my cart that's ok."*

5.2.8 Views on Forget

Based on participants responses about getting control over tracking and targeted ads with the help of 'forget', we found that there were two classes of people, one of which thought that forget was helpful in getting the control they need but other class of people thought that it was a little vague in terms of how it worked and which data will be forgotten. Hence we have two groups below and we discuss what each group of people thought.

Forget helps getting control

The participants were shown mock up 2 of study 1 (See appendix A) to explain our proposal of forget. We explained to them how forget was one of the option in YouTube ad feedback system and if they chose forget than it would delete all the data that lead to an ad. We gave them an example where if they don't want to see any ads about diabetes advertisement than they can select forget options which will then delete all the data that lead to this ad.

After showing them the mock-up when we asked them about forget whether it will help them get the control they need, to this P113 replied: *"Ya probably if there are specific ads which you find creepy and they are based on certain information and if the companies can forget that information*

then theoretically they won't be sort of ads." P107 also agreed and spoke: "That's a good question. I think that if the forget option was there it would just make the ad less daunting and less creepy too. It would be nice to have that option for example if I see an ad that doesn't sit well with me then I could just click forget and then I would have peace of mind knowing that whether for momentarily but my information is forgotten."

We further discuss the second group of people who had different views on forget and though that it was vague as it didn't explain how it works.

Forget is little vague

When asked about forget as an opt out option they mentioned that it was very vague and we think we failed to make them understand the purpose of 'forget'. Some participants thought that it wouldn't be feasible for companies to actually delete data and hence they were not able to trust the forget mechanism completely. One of them was P103 who expressed his distrust and said: *"No there's always gonna be next ad that's gonna creep you out. Unless there is something like legislation that says that if you do click on forget the company will actually forget your information. Because that's always my question that will company actually erase my information. So I think it can help a little but not too much."* P110 also shared the same opinion about forget as expressed his concern: *"If I know that there is an option like these then I would definitely choose it but still I feel that they shouldn't be doing this in the first place. Also, I feel like even though they give you that option, I doubt that they actually delete the data. Like I don't trust them in general. I am sure that they are selling them to someone like some network or company."*

While P111 explained his ideas about tracking very well. I believe that he had core knowledge of how tracking works. According to him, the forget proposal should have a little more information about how it works and what data it will forget if chosen by the user. He felt that the forget mechanism shown in the mock-up had little or no explanation and hence he found it a little vague. He said: *"In the long run maybe but in the short term every ad that they show me while they are learning or forgetting would potentially still be creepy. So it's all a bunch of creepiness that I had to get through before I get to the non-creepiness. The fact that they target ads to me in the first place tells me that they are watching me so it's creepy. Also when I see forget though I need to know what information would be forgotten because I am willing to bet that any piece of advertising has a number of a different set of vectors that have all gone into it to produce this one result. So probably I feel like it's a tree and if you look at it I will show you the ad but what is it that makes a*

Scale Questions	Participants response				
	Always	Very Often	Sometimes	Rarely	Never
Q1	4	5	3	0	1
Q2	1	2	5	4	1
Q3	6	1	3	0	3
Q4	1	0	5	2	5

Q1: How often do you see online ads?

Q2: How often do you see ads related to your interest?

Q3: How often do you use ad-blocking tool to avoid online ads and tracking?

Q4: How often do you provide ad feedback?

Table 5.3: Responses for Scale Question 1 to Question 4.

decision. That's probably a bunch of data so when I say forget what data is actually been removed or what data has actually been forgotten. So I would rather like to see like say forget the data which caused the ad or forget the related ad or something like that so that I know what exactly they mean. So it's more ambiguous."

All in all, due to a lack of explanation, we conclude that participants might have expressed their distrust of whether it will actually delete data for the user. Hence forget would be more appreciated if it was kept in a more clear way explaining how it will delete the data and what data will be deleted. We can see that we had mixed reactions to our proposal of 'forget'. Some participants felt that it could help them get the control they were looking for while other participants thought that it was vague and they couldn't understand what it did. Some of them also expressed trust issues on ad companies as they thought the ad companies could not forget data. Thus, the results for forget were inconclusive. And hence we decided to conduct a second study.

5.2.9 Scale Questions

Based on the analysis of scale questions for study 1 we observed that most of our participants reported seeing online ads very often. Most of the ads were random ads. Participants do not find online ads as useful as 5 of them chose scale response 'not at all useful' and 7 chose 'slightly useful', and none of them chose 'very useful' or 'extremely useful'. It is evident from scale responses of Q1, Q2, Q5, and Q16 in Table 5.3, 5.4, 5.8. Moreover, they think that tracking nowadays is more than they would like. We can say this based on the scale responses we received for Q17 and Q18 in Table 5.9 and Table 5.10 respectively.

Scale Questions	Participant responses				
	Definitely	Very probably	Probably	Probably not	Definitely not
Q5	2	4	5	1	1
Q6	1	1	8	3	0
Q7	3	3	3	2	2

Q5:- Do you think interest related ads are better than random ads?

Q6:- Do you feel that the ad feedback changes the ads you see in future?

Q7:- If ad companies delete information that you don't want them to save, will that change your acceptance of online tracking and targeted ads?

Table 5.4: Responses for Scale Question 5 to Question 7.

Scale Questions	Participant responses				
	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
Q8	1	7	1	3	1
Q9	1	4	3	2	3
Q10	1	9	1	1	1

Q8:- Did the video contain significant information that you didn't know previously about online tracking and ads?

Q9:- Does the ability to provide ad feedback makes the online ads more acceptable?

Q10:- If you had better control on tracking will you be more open towards accepting online tracking and online ads?

Table 5.5: Responses for Scale Question 8 to Question 10.

There was very little or somewhat influence on the participants of the information video because they were familiar with what was shown in the video. This is based on their responses we received from Q8 and Q13 in Table 5.5 and Table 5.7 respectively.

Half of our participants used ad-blocking tools to avoid online ads. However, we had another half who used it sometimes or never used it. This analysis is based on responses for Q3 and Q11 in Table 5.3 and Table 5.6 respectively.

As per participants response Q4, Q6, Q9, and Q15 in Tables 5.3, 5.4, 5.5, and 5.7 we found that most of them were not in the favor of the ad feedback system we showed them through mock up in the first study and felt that it would not make any difference in the way tracking and targeted ads works. Also 4 out of 13 participants never clicked on 'why this ad is displayed?' information on YouTube advance options while 4 of them rarely clicked on it and three of them didn't know that they could actually access this kind of information as per Table 5.11.

Scale Questions	Participant responses		
	Yes	No	I don't think if it's possible
Q11	9	2	2
Q12	8	2	3

Q11:- Do you use any technological means to avoid online ads or tracking?

Q12:- Do you think it is feasible for ad companies to forget(delete) information that they had previously gathered about individuals?

Table 5.6: Responses for Scale Question 11 and Question 12.

Scale Questions	Participant responses			
	To a Great Extent	Somewhat	Very Little	Not at all
Q13	3	5	4	1
Q14	6	5	1	1
Q15	0	5	5	3

Q13:- Were you surprised by the content of this video?

Q14:- After watching this video, are you more or less interested in using tools to block ads and limit tracking?

Q15:- Do the feedback options influence your decision in providing the ad-feedback?

Table 5.7: Responses for Scale Question 13 to Question 15.

Q16:How useful do you find online ads?				
Participant response				
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful
5	7	1	0	0

Table 5.8: Response for Scale Question 16.

Q17:How do you find online tracking nowadays?		
More than I would like	About right	Less than I would like
8	4	1

Table 5.9: Response for Scale Question 17.

Q18:How comfortable are you with the amount of tracking done by ad companies?			
Extremely	Very Moderately	Slightly	Not at all
1	3	5	4

Table 5.10: Response for Scale Question 18.

Q19:How often do you try to get more information about why ad is displayed?				
Always	Very often	Rarely	Never	I didn't know this was possible
0	1	5	4	3

Table 5.11: Response for Scale Question 19.

As expected, 10 out of 13 agreed that it would make them more comfortable if they would get some sort of control on tracking and would change their acceptance for online tracking and targeted ads. They also felt that it is feasible for the companies to forget (delete) data on users they have tracked as per Table 5.5.

Based on the scale responses we can say that overall tracking has increased to a level that the user would not like based on the scale response for Q5. We assume this can be also because user seemed to be aware of online tracking and cookies.

After watching the video, participants were more inclined towards using the ad-blocking tools. We noticed that they were unaware of YouTube ad-feedback and google ad personalization opt out setting. As a result the user strongly desired to get control over tracking and their tracked data.

When we asked participants about getting control they replied that their acceptance of online tracking and targeted ads could be changed with even a small amount of control provided to them. Our scale responses also had a positive response when participants were asked if they would like more control over tracking as in Q10 and Q12. They believed that deleting was feasible for companies. But at the same time, they seemed to have doubts about an existing ad feedback system helping them achieve the control they are looking for.

We did not have scale questions for the second round as the scale responses were inconclusive and they were very much similar to the responses we received from open-ended questions. Hence in order to remove the redundancy of responses we decided to avoid scale questions for second round of study.

5.3 Results from the second study

The results from the first study were inconclusive as the views on forget were mixed, so we decided to perform a second study to understand user perception about forget.

We didn't use the information video as we learnt from the first study that the participants knew about what was explained in the video and it had a little or no influence on our study. We also avoided the scale questions as they didn't provide any significant additional information versus the open ended questions in the first study.

5.3.1 Online Tracking: I think they know pretty much everything

When participants were asked about online tracking and targeted ads, most of the users expressed their concern as advertisers track a lot of information about users. Participants mentioned how they would search something and it would immediately come up in targeted ads. One of our participants who worked for an e-commerce company earlier discussed her experience, Q103 said: “*In my previous work, I was working in an e-commerce company so I was working in the marketing team so I knew people that were [in] business analytic. They used to track user engagement online and see which products they are looking at and they kind of see what category they fall into like age groups and they basically filter the impressions they get and based on all this they target ads. So for instance, if someone is looking for a six-seater table and if they see that the user has come on the product very often so now they just put ads which have all the six-seater tables for them on websites.*” She further added: “*I think it is lots of data. It's creepy at times because I know I work in that field and I know what's going on behind. But I still the same thing because I don't have any control over it. They know our IP addresses, they know our email address because you are always logged into your email, they have our age, location, country, they also have a gender and the websites I am going. I think they know [they] can track pretty much [what] they want.*” Another participant also expressed her views on how everything is tracked when she was asked what data she thinks companies collect for targeting ads: “*Oh everything. Location, age, job type, whether you have kids or not, what type of friends you have, what type of activities your friends are involved in like if they play football or not because they might target you with ads based on football if your friends play football. They might target you with political modes if you are involved in politics. So what you do, what type of friends you have based on that they build your profile.*” On the contrary, some of the participants thought that the companies also track their audio as Q108 responded:

“They track my browsing history, location, also some applications ask for audio access so they can hear us.” In support of this Q104 described similar story when she was asked if she recalls any creepy incidence in terms of tracking and targeted ads: *“Sure there is one time I specifically remember, I was on VIA Rail and I just saw a field of lavender and I thought it was really nice and the next thing I know is that I am scrolling on Instagram and I saw an ad of lavender like a farm where lots of people go and take a picture so I found that was kind of creepy because I didn’t vocalize it or said anything but I just thought it and saw it. So in my experience, that was creepy.”* Q102 also quoted: *“I heard about some of the devices like Alexa, Google home that at some point they record our sound to provide suggestions but I haven’t used these devices. It’s good to have some shopping ads but sometimes it’s too much I don’t want everyone to know about my interest.”* We also had some participants who mentioned cookies in their response when asked about tracking and targeted ads. Q101 described cookies when asked about tracking: *“They probably track my browsing habits and I don’t know how exactly the cookies work but they will follow you on the web sites and where you go around.”*

We also had few participants who knew what was targeted ads but wasn’t sure how it works, as Q102 said: *“I am not very familiar with tracking and targeted ads but I remember when I was searching for furniture on Wayfair it would come up on my Facebook account as an ad afterward.”* We had Q106 who found targeted ads useful as it showed him offers. He said: *“I am not very much ware but I know it. For me, I feel it’s useful because If I want to buy anything and they show me ads for that then it’s useful to buy. I can also get some deals and good offers.”* All in all, participants knew about what information about them is tracked and most of them believed that the ad companies knew everything about them. They listen to what they speak about and they know more than users would want.

5.3.2 Anti-Tracking: You really don’t have a choice

When participants were asked what they would do when they saw a creepy ad and if they were aware of any opt out options, we got mixed results. Some of the participants were not aware of any opt out options while other sets of participants knew one or two tools and spoke about private browsing mode and ad blockers. But at the same time, they expressed their inability to control them despite using tools to avoid tracking. Q107 explained issues she has with ad blockers: *“I close the window or sometimes I clear all the cookies and browsing history. Sometimes I put ad-blocker [1] on the browsers but many sites when you want to view their content they won’t show it to you until*

you disable the blocker so that's another way to track you. So sometimes even if you try you really don't have a choice." She further added when asked about the ad feedback system of YouTube: *"The options are fine but maybe less than 1 percent of people know that there are these options because it's not very visible and they don't tell you where the options are so visibility is definitely one of the issues."* Q104 also stated her issues with Facebook ad profile settings: *"The last time I saw a creepy ad, I went through my Facebook setting and kind of looked at privacy settings to know what exactly is tracked and I couldn't exactly figure that out as it was very ambiguous. But I just tried to make it the most private setting possible. It's difficult to navigate. I do know Facebook allows you to check about your interest but I am not entirely sure how it works."* Q109 also gave his statement in favor of not getting enough control over ads as he said: *"I erase my browser and clear the cookies. I use incognito mode but still, I get ads at times."* Apart from these participant's all other participants were not aware of opt out options. Q108 answered: *"I am a normal person so I don't think I can do much so I just ignore and I am not aware of any opt out option."* When Q101 saw a creepy ad he said he just scrolls away: *"Nothing I just scroll away."* And he further added that: *"I am not aware of any options."* When we asked Q103 if she knew any opt out options she stated: *"No I wish I had known any of them. But I have the ad blocker."*

Overall the participants who were aware of the blocking tools had issues using it and most of them were completely unaware of available opt out options and they considered there is nothing that they can do to control tracking and ads.

5.3.3 Do not track: Chats and health-related ads should be off the table

Participants were when asked about what kind of data they don't want to be tracked. To this, we had most of our participants responding to health-related ads. They found ads that were too personal to them and were based on health topics that were too creepy. But also they had this perception that their communication, emails, chats were also been tracked and they don't want those to be tracked.

Q102 expressed her concern on health-related ads: *"It may be many things. For example, doctors say that it's not good to google search for symptoms online but we often search for that if we have to say cough or a headache and this is very private to us and we don't want anyone to know about it. So that's something that I would say health-related stuff I don't want them to track."* She further added: *"I think if it's online shopping then it's fine even though it might be distracting but that's fine but if it's related to health than I would not like those ads."* While others had concerned about their personal chats. When Q104 was asked about activities she wanted to not track she

replied: *“I will ask my personal conversations online like I am on Facebook or messenger that kind of thinking. Usually, you expect that you and your person know about what you are talking about.”* Q107 also made it clear that personal chats should not be tracked: *“My communication like emails, chats, text messages I think those should be off the table and I don’t think that they should use this content for contributing to building profiles or advertisement. Except that the websites we visit I understand that because we accept the cookies and all but for sure the communication should be always out of all.”* Q108 also agreed and said: *“My private emails, my browsers, my status I don’t want anyone to track this about me.”* While Q104 talked about the vocal conversation to be not tracked: *“For example when I am talking about something in a conversation and that happens to be an ad later on, so that’s creepy and that can be anything from a product to places anything.”* Q103 replied that she doesn’t have a workaround but to accept the data which is tracked but she definitely didn’t want her IP and email to be tracked: *“I don’t want them to track anything about me but I think if they won’t do that I won’t be able to use the internet so I don’t have a workaround. But if they don’t have my email address and IP address than it would be good.”*

So we can say that participants are highly concerned about their health details been tracked and moreover they don’t want their personal emails, chats, and conversations to be tracked.

5.3.4 Forget: It gives you a sense of comfort

For the second study, we changed the presentation of forget in the mock ups (See mock up 2 in appendix B). Unlike the first study, we added a link which read: *“Click if you want to forget searches related to this ad”* beneath a YouTube ad. In the first mock up we showed them ad of credit card and nursery rhymes. We added a link beneath this ad. We aimed to make it more visible and understandable for participants. We explained to them how any specific ad is shown to them based on google ad personalization. Further, we explained how clicking on the link will delete all the searches related to the specific ad. We also showed them screenshots after they would select the link. The data from google ad-personalization will be gone after the link was selected.

In our mock up we took an example of a credit card. So we explained to them that if they wanted to remove data behind this ad then they can select forget.

Most of the participants accepted that they would be more comfortable to have ‘forget’ as an option because it will give them at least some control and peace of mind. When Q103 was asked about forget she said: *“Yeah. I like the idea of having the idea to erase everything about an ad. If not anything this is the best we can have as compared to what we have now. I would rather prefer*

no ads until and unless I asked for it. But if that's not my options then definitely I will go with his." Q104 also replied: *"Yes I think it would make me feel a little bit more in control of what's been tracked because we don't have any idea the amount of information these companies have so I feel like I will have a little bit more sense of comfort knowing that I can click that option."*

When they were asked if they would want forget as an option, Q106 responded: *"Yes definitely because you don't need to navigate through google settings so it will be much easier."* Q107 described that it will give her peace of mind: *"Yes it will give you more control and it will give consumers a little peace of mind as long as it's working the way it is shown."* While Q109 thought that it is more visible so it's useful: *"Compared to three dots I would say it is more visible and understandable. And it is useful because if you click it searches are gone and are not stored anywhere so it is more comfortable for any user."* Q108 said that he will be able to get rid of annoying ads: *"Yes of course because it will help me get rid of some annoying ads."* We asked the user verbally to choose from 1 to 5 in terms of how useful they thought forget was in controlling tracking and targeted ads. 1 was the lowest and 5 was the most. 5 out of 9 participants thought that forget was most useful and chose the highest scale and 3 participants chose scale 4 while there was only one participant who chose 2 because he believed that he had to select these options for every creepy ad.

Overall participants seem satisfied with the idea of forgetting and they believed that it is a good first step.

5.4 Overall analysis of both the studies

We summarize the results of both studies in Table 5.12. We observed similarities in the results obtained from both of the studies. If we consider the user's knowledge for online tracking, targeted ads, and ad blocking options, we observed that both the groups shared the same opinions except for the forget proposal. We noticed a change in their views for 'forget' because obviously it was explained in a different way in each study. As we explained above in the themes, we observed that the participants had an understanding of tracking and cookies but at the same time were unaware of other technical means used for online tracking. They had similar views on creepiness and existing tools which are clearly evident from the above themes hence we talk about forget here.

In the first round, we had mixed responses for 'forget' as we said, half of them felt it was useful and remaining didn't agree with it completely due to reasons we addressed in the above section. While on the contrary, we saw that after adding a clear explanation and an example of how 'forget'

Subject	Results themes	Interpretation
Online ads	Are targeted ads of any help? (Study 1)	Participants felt that the online ads were annoying and frustrating as they would see them almost everywhere. It hinders them from accessing web content.
Online Tracking	Companies use Cookies (Study 1)	Participants are aware that companies used cookies to track them on different websites. Cookies are used to track the items you click, IP address, site visited, etc. Based on this they create an ad profile.
	They know everything (Study 2)	Participants think that almost everything is tracked about them including their address, email, browsing history, chats and even their audio.
Creepiness	Data is collected without my consent/ They are listening (Study 1)	Participants expressed their concern when they realized that they saw ads based on their conversation rather than their searches. They feel creepy when they realized the amount of data been collected without their consent.
Flash Cookies	I don't know those (Study 1)	Participants were shocked when they learned about flash cookies through the information video. They didn't know that cookies can be re-spawned even after deleting them.
Ad blocking	It is universal (Study 1)	Almost everyone knew about it even though if they didn't use it and the most common one was ad-blocker.
Anti-tracking	I wasn't aware of it (Study 1)	Participants were not aware of opt out options including google ad personalization opt out.
	We don't have a choice (Study 2)	Participants find it difficult to reach the opt out settings and hence they felt that they have no choice.
Do Not Track	Health and personal chats should be off the table. (Study 2)	They felt that health-related data and their personal chats should be not tracked.
Forget	It is little vague (Study 1)	In the first round, participants felt that forget didn't explain it's the purpose and how it would work and hence they termed it as vague yet useful.
	It gives a sense of control (Study 2)	Participants really appreciated the idea of forgetting searches related to a creepy ad as it gave them a sense of control over tracking and creepy ads.

Table 5.12: Results Summary

would work and making it more accessible and visible we had a different response. We had a more positive response to 'forget' in the second study.

Based on the difference in the results for forget we learned that an online user looks for an opt out option which is visible to him, which is accessible (the one where they don't need to follow multiple steps or is hard to navigate), which is easily understandable (explains what is the purpose clearly) and doesn't confuse them, which is also quick and something which they can trust easily.

From this study, we learned that online users are aware of online tracking and targeted ads as they mention cookies in their responses and they know that ad networks and third parties know almost everything about them. They also feel that even their conversations have been accessed for targeted ads as most of them spoke about their experiences when they saw ads based on what they were talking about rather than what they searched for. However, this isn't supposed to happen as companies say they don't have access to their conversation [110]. But at the same time, we would say that they are not completely aware of how tracking takes place. They were surprised when they learned about flash cookies through the information video. It was obvious that they would not appreciate online tracking as they knew the extent of tracking. Online users expressed their disagreement with the amount of data tracked and collected about them. Some of them also considered it a privacy violation. We observed that in order to protect themselves from targeted ads only some have started using ad-blocking tools and the most common one we found among them was ad-blocker [1]. Almost everyone knew what was ad blocker whether they used it or not. But they still think that the anti-tracking tools or ad-blocking tools are not well adequate to meet their requirements in terms of control as they felt that companies will find ways to track them anyhow. Some sites would restrict a user from their content when they used ad blockers or ask them for a subscription. When we asked the participants what if they had more control, everyone agreed that it would be more comfortable for them to accept online tracking if they would have even a little bit of control because existing tools are not helping them much. We made them see ad feedback options and asked if they knew about any opt out options. The majority of them didn't know about any opt out options and the ones who knew were not aware of how they could reach them. Now, this points to improper design or implementation of opt out options. They were literally surprised by learning about Google's opt out from ad personalization.

We tried to understand how would they react if they could 'forget' or delete data that had been previously recorded. And as predicted, we got positive feedback about 'forget'. In the first study 'forget' was not clear and it was hard for the user to predict what it would do and how it functioned

but this opinion changed when we presented ‘forget’ with clear explanations. So from this, we understood that an opt out should be self-explanatory, visible, accessible and easy to use. Having an opt out which is hard for the user to find or hard for the user to understand if it is an opt out or not is by no means useful.

Overall we found that we reached the point of saturation for both the studies in terms of users perception about online tracking, targeted ads and anti-tracking tools as we received repetitive responses in both the studies. However, we did not reach saturation with respect to views on ‘forget’. Our results were inconclusive in the first study, and in the second study there were also mixed opinions regarding ‘forget’ with two out of nine having significant concerns. A larger study on ‘forget’ would likely find additional perspectives that may be relevant when considering the utility of fine-grained control over retained user tracking data. However, we believe the results are strong enough to indicate that many users would find some benefit in having an option like ‘forget’.

Chapter 6

Discussion

This thesis provides an insight into what is creepiness for an online user in terms of online tracking and targeted ads. It not only highlights our subject creepiness but also provides an understanding of user mental model about other different aspects like ad-blocking tools, opt out options, forgetting and deletion of data.

We noticed that participants felt annoying when they saw online ads. They were aware of how companies use cookies to track their data like location, IP address, age, browsing history, etc. But they were not sure of how the tracking worked in detail. When we analyzed their knowledge about ad-blocking tools we found that there were two groups of people. Each group talked about ad blockers. They would use different ad blockers such as Ad-Blocker [1], uBlock [53], Adblock Plus [79] to avoid annoying ads. While we had one group which found ad blockers useful to help them get rid of ads, others thought that those were not good solutions because some sites won't allow their content to be viewed if they have the ad blocker on. In that case, the user feels that there is nothing they can do even if they have ad-blocking tools.

It was interesting that the ones who were not from a computer science background knew about the concept explained in the information video (likely due to our recruitment methods). However, we realized that the majority of our participants were unaware of flash cookies and respawning of cookies even though they had a technical background. It came as a shock for them because earlier they believed that deleting cookies would erase all of their tracked data which was not the true story. Few of them expressed their concern due to lack of control and the fact that how big the ad network can be.

We also learned some interesting facts when we asked the participants about existing opt-out options available and how often they use them. We showed them screenshots of YouTube ad feedback options and google ad personalization page. We were astonished to know that except for a couple of participants no one knew about the Google ad personalization page where they can go and change the settings and even turn off ad personalization. Based on their responses we found that the reason behind them not knowing this is that it is not represented in the right way. It is hard

for them to navigate to the settings as it was not very clearly mentioned. They also criticized the three dots drop-down menu which lead to advanced settings for an ad as they were too small to be noticed.

We had similar responses for both the studies in terms of user's knowledge about tracking, targeted ads, and opt-out options except the forget proposal. In our first study, we implemented forget as part of the YouTube ad feedback system to which participants had a mixed response. Half of them felt that it was pretty useful as it would help them delete data they don't want to be tracked or stored. But the other half had a feeling of distrust about ad companies, whether they would actually delete data. They felt that forget could be better if it had some kind of explanation about what data is to be forgotten, thus they thought it is very ambiguous. We took the negative responses in the form of feedback and worked on representing forget with a clear explanation this time and conducted a follow-up study. Also, the results we obtained from the first study were inconclusive hence we decided to do a second study to understand user responses about forget. In the second study, we implemented forget in the form of a link which said: "Click to forget searches related to this ad" just beneath an online ad. As we hoped we received a positive outcome on forget as participants appreciated the ability to forget or delete data and said that it is a good first step to help them control the creepiness. They liked the idea of deleting data. They also appreciated the way forget was represented in terms of visibility, understanding, and accessibility. They compared this with YouTube feedback and said that they now don't need to navigate through the whole set of options. It is very clear and easy to click on just one link.

In the first study when forget was part of the ad feedback option in You-Tube, participants felt that forget was very vague because there was no explanation about forgetting. Moreover in the first study mock up if a user had to select forget they were supposed to go through multiple steps first to reach forget. Users felt that it shall be time-consuming if they wanted to choose forget for every ad they saw. Participants were not explained what shall happen after they select forget. They didn't know which data will be removed and how.

All these factors gave a conclusion that 'forget' was vague and hard for a user to choose as an option to get rid of creepy ads because it had no clarification or explanation about what it did, it was ambiguous for online users. Hence we changed the mock ups and added more explanation to the 'forget' feature. We also explained what would happen if they chose to forget and which data will be removed. In this study, we changed the presentation of 'forget'. Instead of keeping that as a part of an ad feedback system we added it in the form of a link with novice explanation just

beneath the ad itself.

The change in presentation and explanation of how ‘forget’ worked changed user’s view on it. Most of the participants were in favor of ‘forget’ as it gave them control over creepy targeted ads and was easy to understand. Participants also appreciated ‘forget’ as it was visible and easily accessible.

Moreover with increasing use of anti-tracking technologies ad companies would also appreciate the idea of providing control to users in terms of data retention. If companies would provide control on data retention to online users than they would be less liable for tracking too much of their data, users might decrease the use of anti-tracking technologies which indirectly affects ad companies revenue. They also might not need to follow or implement any strict regulations or law as they are already providing control to the users.

We further discuss our research contribution, limitations, recommendation, and future work.

6.1 Contribution

As we discussed in Chapter 2 there have been several types of research conducted to address users concerns about online tracking and targeted ads such as [80, 91, 95, 55] but this research is the first to propose ‘forget’ as an option for addressing user concerns regarding targeted ads.

There has been an extensive literature on user’s perception of online tracking and targeted ads as we discussed in Chapters 2 & 3. In past studies, researches have found that online users feel that tracking is creepy and scary and the targeted ads they are shown are embarrassing and creepy. But the problem of creepiness has been not yet addressed. But this work is the first attempt to understand user’s perception about ‘forgetting’ or deleting data associated with targeted advertisements in an aim to address the feeling of creepiness.

Based on our finding it seems that users desire to control the amount of tracking and targeted ads. Participants mentioned that even a little control can increase their acceptance and trust in tracking. Also because they do not have sufficient control even though there exist several anti-tracking tools. Our research results thus confirm the results obtained from past studies about user’s acceptance of online tracking and targeted ads, specifically that providing user control can increase user acceptance.

Apart from the contributions, our research has some limitations as well which we discuss in the next section.

6.2 Limitations

The demographics in our study are not representative of the entire population. It is just a small group of the entire population. Also our demographic knew about online tracking and targeted ads. Thus the results cannot be suggestive for the whole population. Although our participants belonged to a subgroup of the entire population, it still represented savvy people who used ad blockers and who were aware of tracking and cookies. Hence the results are still useful.

The number of participants we had was not enough to understand user perception about ‘forget’ and user acceptance of online tracking and targeted ads as we could not reach a saturation point in our study in terms of ‘forget’. We obtained a mixed response for ‘forget’ as an opt out option. We understand that if we had more participants than we might have reached the saturation point in terms of users perception about ‘forgetting’ tracked data.

The results are based on specific set up of study which includes watching an information video and answering open-ended and scale questions. The results may vary if the study set up was changed. Also, they were asked about one ‘forget’ proposal that may or may not be feasible to implement. Functionality and interface design may have a big impact on how much a ‘forget’ the mechanism reduces perceived creepiness.

The study results had limited ecological validity as the participants did this study in a lab, they were audio-recorded, they were observed and we verbally explained them the mock ups. The participants were told the definition of creepiness in the beginning of both the studies which might have influenced their views. In practice they may have reacted differently when they saw an online ad and the ‘forget’ opt out, ignoring it or not having such positive feelings. We explicitly explained them how ‘forget’ worked and which data it would delete, and this explanation itself might have made participants view the option more positively. Such explanations were necessary, however, as the first study showed that any ambiguity in the nature of the proposed mechanism led to participant confusion. Our results were based on self-reported opinions which are known to not necessarily correlate with behaviors in practice. Also if the participants could see ‘forget’ as an implemented and functional feature than their opinion and responses would have been changed compare to the one we obtained from showing them mock ups. Thus we conclude that our results, while having marginal ecologically validity, are still useful as early research into the impact of giving users fine-grained control over tracking data.

As we defined creepiness before we asked open-ended questions in both studies we feel that it might have biased the results on their perception about creepiness in terms of online tracking and

targeted ads.

The actual implementation of ‘forget’ can have feasibility issues as it is based on the assumptions we made. We assumed that ‘forget’ could delete all the searched related to a specific ad. But as we saw that tracking data is stored in distributed data structures where it is often entangled with other data to display a particular ad. In this case, it might be difficult for ‘forget’ to function precisely as we proposed. Thus ‘forget’ is best understood as a way to understand users acceptance for online tracking when given some direct control over data retention.

In situations where the user might search for a product or any other data related to a specific ad in future which the user has asked to be forgotten, ad preference manager will be a better choice rather than ‘forget’ as ‘forget’ will delete the data related to specific ad for a time frame till the user has not searched for it again. And hence it can show the same ad or related ad in future to the user if he does a related search. While on the contrary, an ad preference manager will remember to not display an ad on a particular subject which user has turned off until the user allows. Note that ad preference managers won’t allow ads related to the subject which user has turned off to be displayed even though they may do searches or other online activity directly related to the subject.

Further in next section, we discuss the recommendation and future research work based on our study.

6.3 Recommendations and Future work

It seems that users would be comfortable if they had an opt out which is understandable and easy to use however it is still a wide area of research to find out an opt out which will not be undetermined or ignored as the existing opt outs are. Future research can be focused on creating an opt out which is easy to use and understandable without any explicit explanation.

We found that forgetting tracked data can make the user comfortable and increase their acceptance but we still need to explore how we can design and implement a mechanism which can function similar to our proposal of forget. We recommended the mock ups used in the second study to be used while designing any future opt-out mechanism. However, we are not sure how feasible it would actually be to implement an actual mock up that would work similarly to what we proposed. Hence we feel that this could be another area for research.

Big data companies storing tracked data can consider decaying of old data collected for the user. Google already has an opt out for ad personalization where users can turn off ads for certain topics. After turning off they can see the topic in ‘what you have turned off’ feature as explained

in mock ups. This way Google tends to remember what needs to be forgotten about the user. So we assume that it is still storing the data. Instead of storing this we recommend if Google can forget that data completely and if they do not store that data anywhere. We feel that it is way more convenient as the ad network didn't have to remember the unnecessary information and occupy space. Technically, It is difficult to delete data from the ad servers as we discussed earlier in Chapter 3 but at the same time the data can be made inaccessible. Although there are many challenges associated with data deletion, we propose 'forget' as a reasonable research area because it can be implemented to challenge the technical issues with data deletion. We aimed to understand whether the user would appreciate the idea of 'forgetting' or not. We wanted to give a concrete base for future studies based on 'forget' and 'delete' as now future research can be based on the assumption that user would like to have 'forget' as an opt out for online tracking and targeted ads and overall it can also help future studies on developing opt out options.

In our study, we just focused on YouTube and Google ad settings but we consider that the scope is still too wide to be explored to understand the concept of creepiness.

Based on our results we can say that at least some online users highly appreciate their online privacy and have a great desire to control the amount of their data been tracked. They feel creepy when the amount of data been tracked is not transparent, they do not have enough control over tracking and creepy targeted ads even though there exist several anti-tracking tools. It seems that users want an opt out which provides them the required control, which is easy to understand and easy to use. However, we do not know how one can implement a user interface which can fulfill users requirement of opt out. But we understood that 'forget' can reduce the creepiness and increase the trust and acceptance of online users on online tracking and hence it can possibly lead to mechanisms that match user requirements.

Bibliography

- [1] Ad-Blocker. Ad-blocker. <http://www.ad-blocker.org/AdBlockerChrome>. Retrieved January 16, 2020.
- [2] Adobe. Flash player on adobe support community. <https://community.adobe.com/t5/flash-player/>, Dec 2019.
- [3] Lalit Agarwal, Nisheeth Shrivastava, Sharad Jaiswal, and Saurabh Panjwani. Do not embarrass. *Proceedings of the Ninth Symposium on Usable Privacy and Security - SOUPS 13*, 2013.
- [4] Istemi Ekin Akkus, Ruichuan Chen, Michaela Hardt, Paul Francis, and Johannes Gehrke. Non-tracking web analytics. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 687–698. ACM, 2012.
- [5] Chloe Albanesius. PC Magazine - Internet Explorer 10 Released for Windows 7. *PCMAG*, Nov 2012.
- [6] AORI. Targeted ads: Convenient or creepy? <https://aori.com/blog/targeted-ads-creepy>, Mar 2019.
- [7] Mika D Ayenson, Dietrich James Wambach, Ashkan Soltani, Nathan Good, and Chris Jay Hoofnagle. Flash cookies and privacy ii: Now with html5 and etag respawning. *Available at SSRN 1898390*, 2011.
- [8] Arif Bacchus. Digital Trends : Millions of People Use 'Do Not Track' Tool Which Does Nothing. <https://www.digitaltrends.com/computing/do-not-tracking-tools-do-nothing/>, Oct 2018.
- [9] Muhammad Ahmad Bashir, Umar Farooq, Maryam Shahid, Muhammad Fareed Zaffar, and Christo Wilson. Quantity vs. quality: Evaluating user interest profiles using ad preference managers. In *NDSS*, 2019.
- [10] Doug Beaver, Sanjeev Kumar, Harry C. Li, Jason Sobel, and Peter Vajgel. Finding a needle in haystack: Facebook's photo storage. In *Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation, OSDI'10*, pages 47–60, Berkeley, CA, USA, 2010. USENIX Association.
- [11] Hal Berghel and David Hoelzer. Disk wiping by any other name. *Commun. ACM*, 49(8):17–21, August 2006.
- [12] Nataliia Bielova. Web tracking technologies and protection mechanisms. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS '17*, pages 2607–2609, New York, NY, USA, 2017. ACM.

- [13] Nick Bilton. Girls around me: An app takes creepy to a new level. *The New York Times*, Mar 2012.
- [14] Dominik Birk and Christoph Wegener. Technical issues of forensic investigations in cloud computing environments. *2011 Sixth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering*, 2011.
- [15] Michael Björn. Why targeted advertising is becoming creepy. <https://www.ericsson.com/en/blog/2019/1/why-targeted-advertising-is-becoming-creepy>, Nov 2019.
- [16] Sophie C. Boerman, Sanne Kruike-meier, and Frederik J. Zuiderveen Borgesius. Online behavioral advertising: A literature review and research agenda. *Journal of Advertising*, 46(3):363–376, 2017.
- [17] Dan Bogdanov, Liina Kamm, Swen Laur, and Ville Sökk. Implementation and evaluation of an algorithm for cryptographically private principal component analysis on genomic data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, page 1–1, 2018.
- [18] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS '17*, pages 1175–1191, New York, NY, USA, 2017. ACM.
- [19] Dan Boneh and Richard J Lipton. A revocable backup system. In *USENIX Security Symposium*, pages 91–96, 1996.
- [20] Raphael Bost, Raluca Ada Popa, Stephen Tu, and Shafi Goldwasser. Machine learning classification over encrypted data. *Proceedings 2015 Network and Distributed System Security Symposium*, 2015.
- [21] Tega Brain. The New Organs- Collection of stories about 'Does the Internet know more about you than you think it should?'. <https://neworgans.net/>. Retrieved January 16, 2020.
- [22] Juan Miguel Carrascosa, Jakub Mikians, Ruben Cuevas, Vijay Erramilli, and Nikolaos Laoutaris. I always feel like somebody's watching me: Measuring online behavioural advertising. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies, CoNEXT '15*, pages 13:1–13:13, New York, NY, USA, 2015. ACM.
- [23] Farah Chanchary and Sonia Chiasson. User perceptions of sharing, advertising, and tracking. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 53–67, Ottawa, July 2015. USENIX Association.
- [24] Winnie Chung and John Paynter. Privacy issues on the internet. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences, 2002. HICSS.*, page 193, 01 2002.

- [25] E. F. Codd. A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377–387, June 1970.
- [26] E F E F. Codd. Derivability, redundancy and consistency of relations stored in large data banks. *ACM SIGMOD Record*, 38(1):17, 2009.
- [27] Cranor, Lorrie Faith, and McDonald. Beliefs and behaviors: Internet users’ understanding of behavioral advertising. *TPRC 2010. Available at SSRN: <https://ssrn.com/abstract=1989092>*, Jan 2012.
- [28] Lorrie Faith Cranor. Can users control online behavioral advertising effectively? *IEEE Security & Privacy Magazine*, 10(2):93–96, 2012.
- [29] Lorrie Faith Cranor, Joseph Reagle, and Mark S Ackerman. Beyond concern: Understanding net users’ attitudes about online privacy. *The Internet upheaval: raising questions, seeking answers in communications policy*, pages 47–70, 2000.
- [30] Claire Dolin, Ben Weinshel, Shawn Shan, Chang Min Hahn, Euirim Choi, Michelle L. Mazurek, and Blase Ur. Unpacking perceptions of data-driven inferences underlying online targeting and personalization. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI 18*, 2018.
- [31] Charles Duhigg. How companies learn your secrets. *The New York Times*, 16:2012, 2012.
- [32] Peter Eckersley. How unique is your web browser? In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 1–18. Springer, 2010.
- [33] Marnie Eisenstadt. Private companies know where you’ve been, thanks to license plate cameras. https://www.syracuse.com/news/2015/01/private_companies_know_where_youve_been_thanks_to_license_plate_cameras.html, Jan 2015.
- [34] eMarketer Editors. Data suggests surprising shift: Duopoly not all-powerful. *eMarketer*, 2018.
- [35] Mike English. Restoring deleted files in linux from the ext3 journal. <https://spin.atomicobject.com/2012/06/29/restoring-deleted-files-from-the-ext3-journal/>, Apr 2018.
- [36] Facebook. Facebook. What are my ad preferences and how can i adjust them on facebook? <https://www.facebook.com/help>, July 2019.
- [37] Thomas J. Fitzgerald. Deleted but not gone. *The New York Times*, Nov 2005.
- [38] James Frew. Makeusof - How Advertisers Use Web Beacons to Track You on the Web and in Emails. <https://www.makeuseof.com/tag/how-web-beacons-track-web/>, Dec 2016.

- [39] Shaksham Garg and Rishabh Mahrsee. How cookies work? <https://www.geeksforgeeks.org/javax-servlet-http-cookie-class-java/>. Retrieved January 16, 2020.
- [40] Joanna Geary. Tracking the trackers: Introduction to cookies and web tracking. *The Guardian*, Apr 2012.
- [41] Antonio Ginart, Melody Guan, Gregory Valiant, and James Y Zou. Making ai forget you: Data deletion in machine learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 3513–3526. Curran Associates, Inc., 2019.
- [42] Joshua Gomez, Travis Pinnick, and Ashkan Soltani. Knowprivacy. https://www.techsoupcanada.ca/en/learning_center/articles/analytics_tools,2009.
- [43] Google. Google Support. Choose your privacy settings. <https://support.google.com/chrome/answer/114836?co=GENIE.Platform=Desktop&hl=en>. Retrieved January 16, 2020.
- [44] Google. Google Support. Control the ads you see. <https://support.google.com/ads/answer/2662856?co=GENIE.Platform=Android&hl=en>, May 2019. Retrieved January 16, 2020.
- [45] Google. Google Support. Data deletion on Google Cloud Platform. <https://cloud.google.com/security/deletion/#>, November 2019. Retrieved January 16, 2020.
- [46] Google. Google Support Document.Tracking Code Overview Google Analytics and Google Developers. <https://developers.google.com/analytics/resources/concepts/gaConceptsTrackingOverview>, October 2019. Retrieved January 16, 2020.
- [47] Google. Goolge ad personalization. <https://adssettings.google.com/>, August 2019.
- [48] Preston Gralla. How to stop ad trackers and beacons dead in their tracks. <https://www.itworld.com/article/2974461/how-to-stop-ad-trackers-and-beacons-dead-in-their-tracks.html>, Aug 2015.
- [49] Bjoern Grief. Cookies, fingerprinting Tracking methods clearly explained. <https://www.ghostery.com/blog/ghostery-news/cookies-fingerprinting-co-tracking-methods-clearly-explained/>, March 2018.
- [50] Elliot Harmon. Site Statistics and User Privacy for Nonprofit Websites. https://www.techsoupcanada.ca/en/learning_center/articles/analytics_tools, Oct 2009.

- [51] Todd Haselton. How to find out what google knows about you and limit the data it collects. *CNBC*. December, 6, 2017.
- [52] Nathan Heller. The New Yorker - The Age of Creepiness. *The New Yorker*, Jun 2017.
- [53] Raymond Hill. A fast and efficient ad blocker. easy on cpu and memory. <https://ublock.org/guide/>, Jul 2019.
- [54] Rebecca Jennings. Vox - The joy and horror of targeted Facebook ads. *Vox*, Sep 2018.
- [55] Yucheng Jin, Karsten Seipp, Erik Duval, and Katrien Verbert. Go with the flow: Effects of transparency and user control on targeted advertising using flow charts. In *Proceedings of the International Working Conference on Advanced Visual Interfaces, AVI '16*, pages 68–75, New York, NY, USA, 2016. ACM.
- [56] Robert P St John and William A Lloyd. Web tracking system, December 4 1984. US Patent 4,485,982.
- [57] Amruta Joshi, Abraham Bagherjeiran, and Adwait Ratnaparkhi. User demographic and behavioral targeting for content match advertising. In *Proceedings of the Fifth International Workshop on Data Mining and Audience Intelligence for Advertising (ADKDD 2011)*, pages 53–60. Citeseer, 2011.
- [58] Nikolai Joukov, Harry Papaxenopoulos, and Erez Zadok. Secure deletion myths, issues, and solutions. *Proceedings of the second ACM workshop on Storage security and survivability(StorageSS 06)*, 2006.
- [59] Samy Kamkar. evercookie, September 20, 2010. <https://samy.pl/evercookie/>.
- [60] Avita Katal, Mohammad Wazid, and Rayan H Goudar. Big data: issues, challenges, tools and good practices. In *2013 Sixth international conference on contemporary computing (IC3)*, pages 404–409. IEEE, 2013.
- [61] Kate Kaye. Study: Consumers don't know what adchoices privacy icon is. *AdAge*, January 29 2014. <https://adage.com/article/privacy-and-regulation/study-consumers-adchoices-privacy-icon/291374>.
- [62] Pedro Giovanni Leon, Justin Cranshaw, Lorrie Faith Cranor, Jim Graves, Manoj Hastak, Blase Ur, and Guzi Xu. What do online behavioral advertising privacy disclosures communicate to users? *Proceedings of the 2012 ACM workshop on Privacy in the electronic society - WPES 12*, 2012.
- [63] Bin Liu, Anmol Sheth, Udi Weinsberg, Jaideep Chandrashekar, and Ramesh Govindan. Adreveal. *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks - HotNets-XII*, 2013.
- [64] Bishop C M. Pattern recognition & machine learning. In *Information science and statistics*. New York, NY : Springer, 2006. - 738 p., 1992.

- [65] Jones M. IBM Developer - Anatomy of the Linux virtual file system switch. <https://developer.ibm.com/tutorials/l-virtual-filesystem-switch/>, August 31 2009. IBM Developer.
- [66] Moira Maguire and Brid Delahunt. Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education*, 9(3), 2017.
- [67] Miguel Malheiros, Charlene Jennett, Sneha Patel, Sacha Brostoff, and Martina Angela Sasse. Too close for comfort. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI 12*, 2012.
- [68] Kirsten Martin. Facebook (A): Beacon and privacy. Case BRI-1 006 (A), 2010. Institute for Corporate Ethics.
- [69] Francis T McAndrew and Sara S Koehnke. On the nature of creepiness. *New ideas in psychology*, 43:10–15, 2016.
- [70] Aleecia M. McDonald and Lorrie Faith Cranor. Americans’ attitudes about internet behavioral advertising practices. In *Proceedings of the 9th Annual ACM Workshop on Privacy in the Electronic Society*, WPES ’10, pages 63–72, New York, NY, USA, 2010. ACM.
- [71] William Melicher, Mahmood Sharif, Joshua Tan, Lujio Bauer, Mihai Christodorescu, and Pedro Giovanni Leon. (do not) track me sometimes: Users’ contextual preferences for web tracking. *Proceedings on Privacy Enhancing Technologies*, 2016(2):135–154, Jan 2016.
- [72] Microsoft. Microsoft - FileEncryption. <https://docs.microsoft.com/en-gb/windows/win32/fileio/file-encryption?redirectedfrom=MSDN>, May 2019.
- [73] Subramanian Muralidhar, Wyatt Lloyd, Sabyasachi Roy, Cory Hill, Ernest Lin, Weiwen Liu, Satadru Pan, Shiva Shankar, Viswanath Sivakumar, Linpeng Tang, and Sanjeev Kumar. F4: Facebook’s warm blob storage system. In *Proceedings of the 11th USENIX Conference on Operating Systems Design and Implementation*, OSDI’14, pages 383–398, Berkeley, CA, USA, 2014. USENIX Association.
- [74] Hamre. M Myhrvold. S. Too creepy for comfort? a study of personalized online advertising effects on attitude towards the ad and the advertised brand across high/low involvement and socially sensitive products, and the mediating role of the creepiness factor. Master’s thesis, BI Norwegian Business School - campus Oslo, 2018.
- [75] Leo Notenboom. How do i detect web beacons in email? <https://askleo.com/how-do-i-detect-web-beacons-in-email/>, Feb 2019.
- [76] Karina Oertel, Oliver Hein, and Antje Elsner. The realeyex-project: Usability evaluation with eye tracking data. In *INTERACT*, pages 733–734, 2001.

- [77] Mark Oliver. 10 creepy ways companies collect data for targeted ads. <https://listverse.com/2018/05/27/10-creepy-ways-companies-collect-data-for-targeted-ads/>, Apr 2019.
- [78] Stefanie Olsen. Nearly undetectable tracking device raises concern. *CNET News*, January 2 2002. <https://www.cnet.com/news/nearly-undetectable-tracking-device-raises-concern/>.
- [79] Wladimir Palant. Adblock plus: The world's # 1 free ad blocker. <https://adblockplus.org/en/about>. Retrieved January 16, 2020.
- [80] Javier Parra-Arnau, Jagdish Prasad Acharya, and Claude Castelluccia. Myadchoices: Bringing transparency and control to online advertising. *ACM Trans. Web*, 11(1):7:1–7:47, March 2017.
- [81] Radia Perlman. Secure deletion of data. In *Proc. of the third international IEEE Security In Storage Workshop*, 2005.
- [82] Zachary NJ Peterson, Randal C Burns, Joseph Herring, Adam Stubblefield, and Aviel D Rubin. Secure deletion for a versioning file system. In *FAST*, volume 5, 2005.
- [83] Chanda Phelan, Cliff Lampe, and Paul Resnick. Its creepy, but it doesnt bother me. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI 16*, 2016.
- [84] Angelisa C. Plane, Elissa M. Redmiles, Michelle L. Mazurek, and Michael Carl Tschantz. Exploring user perceptions of discrimination in online targeted advertising. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 935–951, Vancouver, BC, August 2017. USENIX Association.
- [85] Ivens Portugal, Paulo Alencar, and Donald Cowan. The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, 97:205–227, 2018.
- [86] Pixel Privacy. Browser fingerprinting: What is it and what should you do about it? <https://pixelprivacy.com/resources/browser-fingerprinting/>. Retrieved January 16, 2020.
- [87] Paul Reber. What is the memory capacity of the human brain? <https://www.scientificamerican.com/article/what-is-the-memory-capacity/>, May 2010.
- [88] Nithya Sambasivan, Garen Checkley, Amna Batool, Nova Ahmed, David Nemer, Laura Sanely Gaytán-Lugo, Tara Matthews, Sunny Consolvo, and Elizabeth Churchill. “Privacy is not for me, it’s for those rich women”: Performative privacy practices on mobile phones by women in south asia. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*, pages 127–142, 2018.

- [89] Paul Sarbanes. Sarbanes-oxley act of 2002. In *The Public Company Accounting Reform and Investor Protection Act. Washington DC: US Congress*, 2002.
- [90] J. Ben Schafer, Joseph Konstan, and John Riedi. Recommender systems in e-commerce. *Proceedings of the 1st ACM conference on Electronic commerce - EC 99*, 1999.
- [91] Ryan Schoen. Expanding user protections on the web. <https://blog.chromium.org/2017/11/expanding-user-protections-on-web.html>, Nov 2017.
- [92] Barry Silverstein. *Internet Marketing for Information Technology Companies: Proven Online Techniques to Increase Sales and Profits for Hardware, Software and Networking Companies*. Independent Publishers Group, 2001.
- [93] Tom Simonite. Facebook’s like buttons will soon track your web browsing to target ads. *MIT Technology Review*, September 16 2015. <https://www.technologyreview.com/s/541351/>.
- [94] Lindsay Simpkins, Xiaohong Yuan, Jwalit Modi, Justin Zhan, and Li Yang. A course module on web tracking and privacy. *Proceedings of the 2015 Information Security Curriculum Development Conference on - InfoSec 15*, 2015.
- [95] Lindsay Simpkins, Xiaohong Yuan, Jwalit Modi, Justin Zhan, and Li Yang. A course module on web tracking and privacy. In *Proceedings of the 2015 Information Security Curriculum Development Conference, InfoSec ’15*, pages 10:1–10:7, New York, NY, USA, 2015. ACM.
- [96] Janice C. Sipior, Burke T. Ward, and Ruben A. Mendoza. Online privacy concerns associated with cookies, flash cookies, and web beacons. *Journal of Internet Commerce*, 10(1):1–16, 2011.
- [97] Brad Smith. 31 advertising statistics to know in 2018. <https://www.wordstream.com/blog/ws/2018/07/19/advertising-statistics>. Retrieved January 16, 2020.
- [98] Richard M. Smith. The Web Bug FAQ. https://web.archive.org/web/20010729060646/www.eff.org/Privacy/Marketing/web_bug.html, 2019. Electronic Frontier Foundation.
- [99] Christopher Soghoian. The History of the Do Not Track Header. *Slight Paranoia*, 2011. <http://paranoia.dubfire.net/2011/01/history-of-donot-track-header.html>.
- [100] Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris Jay Hoofnagle. Flash cookies and privacy. In *2010 AAAI Spring Symposium Series*, 2010.
- [101] Louise Story. Apologetic, facebook changes ad program. *New York Times*, 2007.
- [102] Andrew S Tanenbaum and Herbert Bos. *Modern operating systems*. Pearson, 2015.

- [103] Helma Torkamaan, Catalin-Mihai Barbu, and Jurgen Ziegler. How can they know that? *Proceedings of the 13th ACM Conference on Recommender Systems - RecSys 19*, 2019.
- [104] Christina Tsuei. How advertisers use internet cookies to track you. *The Wall Street Journal*, July 30 2010. <https://www.wsj.com/video/how-advertisers-use-internet-cookies-to-track-you/92E525EB-9E4A-4399-817D-8C4E6EF68F93.html>.
- [105] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart, useful, scary, creepy. *Proceedings of the Eighth Symposium on Usable Privacy and Security - SOUPS 12*, 2012.
- [106] Natalija Vlajic, Marmara El Masri, Gianluigi M. Riva, Marguerite Barry, and Derek Doran. Online tracking of kids and teens by means of invisible images: Coppa vs. gdpr. In *Proceedings of the 2Nd International Workshop on Multimedia Privacy and Security, MPS '18*, pages 96–103, New York, NY, USA, 2018. ACM.
- [107] W3Schools. HTML5 Web Storage. https://www.w3schools.com/html/html5_webstorage.asp, 2019.
- [108] Kurt Wagner. This is how facebook collects data on you even if you don't have an account. <https://www.vox.com/2018/4/20/17254312/facebook-shadow-profiles-data-collection-non-users-mark-zuckerberg>, 2018.
- [109] Tom Warren. Facebook has been collecting call history and sms data from android devices. <https://www.theverge.com/2018/3/25/17160944/facebook-call-history-sms-data-collection-android>, Mar 2018.
- [110] Elizabeth Weise. No, facebook doesn't secretly listen via your microphone to target ads at you. *USA Today*, April 10 2018. <https://www.usatoday.com/story/tech/2018/04/10/no-facebook-doesnt-secretly-listen-via-your-microphone-target-ads-you/505257002/>.
- [111] Craig E. Wills and Can Tatar. Understanding what they do with what they know. *Proceedings of the 2012 ACM workshop on Privacy in the electronic society - WPES 12*, 2012.
- [112] Queenie Wong. Facebook's ad targeting has created a creepy image problem it can't shake. *CNET News.com*, May 2019.
- [113] C. P. Wright, J. Dave, and E. Zadok. Cryptographic file systems performance: What you don't know can hurt you. In *Second IEEE International Security in Storage Workshop*, pages 47–47, Oct 2003.
- [114] Yaxing Yao, Davide Lo Re, and Yang Wang. Folk models of online behavioral advertising. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW 17*, 2017.
- [115] YouTube. Youtube Ad Feedback. <https://www.youtube.com/>, 2019.

Appendix A

Study 1

This appendix contains all the recruitment documents used for study 1, questionnaires, mock-ups and study script. The flow of this Appendix is as follows:

- Consent Form
- Poster
- Email Recruitment
- Online Recruitment
- Questionnaire
- Mock ups
- Study Script

Research Consent Form

Name and Contact Information of Researchers:

Vidhi Kirit Shah, Carleton University, School of Computer Science

Tel.: 613-400-9799

Email: vidhikiritshah@cmail.carleton.ca

Supervisor and Contact Information: *Prof. Anil Somayaji, Carleton University, School of Computer Science.*

Project Title

Users Perceptions of Targeted Advertisements and Online Tracking

Carleton University Project Clearance

Clearance #: CUREB-B Clearance #111390

Date of Clearance: September 06, 2019

Invitation

The information in this form is intended to help you understand what we are asking of you so that you can decide whether you agree to participate in this study. Your participation in this study is voluntary, and a decision not to participate will not be used against you in any way. The researcher for this study is a master's student, Vidhi Kirit Shah. She is working under the supervision of Prof. Anil Somayaji in the Computer Science Department. As you read this form, and decide whether to participate, please ask all the questions you might have, take whatever time you need, and consult with others as you wish.

What is the purpose of the study?

The aim of this study is to explore participants' perceptions of targeted advertisements and online tracking. Users of online applications are constantly bombarded with targeted ads based on their online behavior which includes their browsing history, location, their time spent on websites, etc. In some cases, users find some ads very creepy as user feel that the ad network knows more about them than they would like. This information may be embarrassing or otherwise sensitive. This research revolves around whether changes in existing tracking systems could improve user acceptance of online tracking for advertising purposes.

What will I be asked to do?

If you agree to take part in the study, we will ask you to answer some multiple choice and open-ended questions regarding your knowledge about targeted advertisements and web tracking. As part of this, we will show you an informational video which will help you understand what targeted advertisement are and how they work. We will show you some screenshots and mock-ups related to YouTube ads and Google Ad settings to understand your awareness about them. Please note that you are not being tested; we are only interested in your perception of cybersecurity warning messages. This research study will be audio-recorded for the purposes of transcription and analysis.

The study will take around 45 minutes and will involve four rounds of questions, watching information video, observing mock-ups followed by post-mockup interviews. Each round of questionnaires will include some set of scale questions followed by open-ended questions. There are no predictable risks in participating in this study. You will not be asked to disclose any personally identifiable information.

Risks and Inconveniences

We do not anticipate any risks to participating in this study.

Possible Benefits

You may not receive any direct benefit from your participation in this study. However, you may learn more about how online ad tracking works. The results from this study will be used to make recommendations for better opt-out options from creepy targeted ads.

Compensation/Incentives

As a token of appreciation, you will receive a \$10 Tim Horton's gift card.

No waiver of your rights

By signing this form, you are not waiving any rights or releasing the researchers from any liability.

Withdrawing from the study

If you withdraw your consent during the study, all information collected from you before your withdrawal will be discarded. You may withdraw at any time before September 30, 2019.

Confidentiality

We will treat your personal information as confidential, although absolute privacy cannot be guaranteed. No information that discloses your identity will be released or published without your specific consent. Research records may be accessed by the Carleton University Research Ethics Board in order to ensure continuing ethics compliance.

The results of this study may be published or presented at an academic conference or meeting, but the data will be presented so that it will not be possible to identify any participants unless you give your express consent. You will be assigned a code so that your identity will not be directly associated with the data you have provided. Recordings and questionnaires will be destroyed after they are transcribed. Transcribed data and analysis will be kept in a password-protected file on a secure computer. The master list associating your name with your code will be kept on paper on a master list stored in a secure location. This list will be destroyed in six months.

What if I do not want to be audio-recorded?

If you choose to not be audio-recorded, you may still participate in the study. We will take notes of what you said during the study.

Data Retention

The audio recording and the questionnaires shall be destroyed once they are transcribed. The transcriptions will be kept for two years for the purpose of publication. Participant's contact information will not be stored for future recruitment and shall be deleted after the project completion (at most two years). The anonymized analysis shall be archived (in the CCSL and on the researcher's personal machine) after thesis is defended and approved by the committee.

New information during the study

In the event that any changes could affect your decision to continue participating in this study, you will be promptly informed.

Ethics review

This project was reviewed and cleared by the Carleton University Research Ethics Board B. Should you have any ethical concerns with the study, please contact the REB Chair, Carleton University Research Ethics Board-B (by phone: 613-520-2600 ext. 4085 or by email: ethics@carleton.ca). For all other questions about the study, please contact the researcher.

Statement of consent – print and sign name

I voluntarily agree to participate in this study.

Yes

No

I agree to be audio recorded.

Yes

No

Signature of participant

Date

Research team member who interacted with the subject

I have explained the study to the participant and answered any and all their questions. The participant appeared to understand and agree. I provided a copy of the consent form to the participant for their reference.

Signature of researcher

Date



Participate in a Study about Targeted advertisement and Web Tracking.

To participate in this study, you must be:

- Familiar with online ads
- At least 18 years old
- Comfortable in the English language

This is a 45-minute study. You will be asked to answer questions about targeted advertisement and web tracking done by ad networks.

Participants will be compensated with a \$10 Tim Horton's gift card.

The ethics protocol for this project has been reviewed and cleared by the Carleton University Research Ethics Board. If you have any ethical concerns with the study, please contact Dr. Bernadette Campbell, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at ethics@carleton.ca). [insert information for CUREB-, if appropriate – see instructions.]

If you are interested in participating, please email Vidhi Kirit Shah at: vidhikiritshah@mail.carleton.ca

This research has been cleared by the Carleton University Research Ethics Board (CUREB-B), REB clearance #.

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Email Invitation

Subject: A research study to understand user perception on Creepy targeted ads and their acceptance of online tracking.

Dear Sir/Madam,

My name is Vidhi and I am a master's student in the Computer Science department at Carleton University. I am working on a research project under the supervision of Prof. Anil Somayaji.

I am writing to you today to invite you to participate in a study entitled “Users Perceptions of Targeted Advertisements and Online Tracking”. This study aims to understand users’ perception on creepiness of targeted advertisement and the acceptability of online tracking.

The study will take approximately 45 mins. As a participant in the study, you will observe screenshots and mock ups, watch a video, answer multiple choice questions, and discuss open-ended questions.

We are looking for adult participants over 18 years old who have a minimal understanding of targeted ads and web tracking and are comfortable speaking and reading in English.

You will have the right to end your participation in the study at any time, for any reason, up until September 30, 2019. If you choose to withdraw, all the information you have provided will be destroyed.

Participants will be compensated with a \$10 Tim Horton’s gift card.

If you are interested in participating, please email Vidhi Kirit Shah at: vidhikiritshah@email.carleton.ca

This research has been cleared by the Carleton University Research Ethics Board (CUREB-B), REB clearance #.

If you have any ethical concerns with the study, please contact Dr. Bernadette Campbell, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca).

Sincerely,

Vidhi Kirit Shah

Online Invitation

Post on Carleton Research Participants Facebook group.

Volunteers needed for research study on targeted advertisement and web tracking.

My name is Vidhi and I am a master's student in the Computer Science department at Carleton University. I am working on a research project under the supervision of Prof. Anil Somayaji.

I am writing to you today to invite you to participate in a study entitled “Users perception of Targeted advertisement and acceptance on online tracking”. This study aims to understand users’ perception on creepiness of targeted advertisement and the acceptability of online tracking.

The study will take approximately 45 mins. As a participant in the study, you will observe screenshots and mock ups, watch a video, answer multiple choice and discuss open-ended questions.

We are looking for adult participants over 18 years old who have minimal understanding of targeted ads and web tracking and are comfortable speaking and reading in English.

You will have the right to end your participation in the study at any time, for any reason, up until September 30, 2019. If you choose to withdraw, all the information you have provided will be destroyed.

Participants will be compensated with a \$10 Tim Horton’s gift card.

If you are interested in participating, please email Vidhi Kirit Shah at: vidhikiritshah@gmail.com

This research has been cleared by the Carleton University Research Ethics Board (CUREB-B), REB clearance #.

If you have any ethical concerns with the study, please contact Dr. Bernadette Campbell, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca).

Sincerely,

Vidhi Kirit Shah

Questionnaires - 1

1. How often do you see online ads?
 - Always
 - Very Often
 - Sometimes
 - Rarely
 - Never

2. How useful do you find online ads?
 - Not at all useful
 - Slightly useful
 - Moderately useful
 - Very useful
 - Extremely useful

3. How often do you see ads related to your interest?
 - Always
 - Very Often
 - Sometimes
 - Rarely
 - Never

4. Do you think interest related ads are better than random ads?
 - Definitely
 - Very Probably
 - Probably
 - Probably Not
 - Definitely Not

5. How do you find online tracking nowadays?
 - More than I would like
 - About right
 - Less than I would like

6. Do you use any technological means to avoid online ads or tracking?

- Yes
- No
- I didn't know if it's possible

7. How often do you use ad-blocking tools to avoid online ads and tracking?

- Always
- Very Often
- Sometimes
- Rarely
- Never
- I don't know

Questionnaires - 2

Definition of Creepiness: - Creepiness in general is a feeling you get when you are non-consensually observed while engaging in private behavior. For example, a stranger watching looking at you through your window while you are doing your own thing at home is normally creepy.

1. How much time do you spend online every day?
2. What is the first thing that comes in your mind when you hear 'online ads'?
3. Where do you most commonly see online ads?
4. What kind of online ads do you normally see?
5. Do you ever see ads related to your interests? Can you describe one or two ads related your interests?
6. Are you aware of online tracking? If yes, please describe it in your own words and, if possible, give an example.
7. What do you know about targeted advertisements?
8. What information you think the ad companies collect about you for targeting ads?
9. What is the definition of creepiness for you in terms of tracking and online ads?
10. Do you feel that online tracking is creepy? Why or Why not?
11. Do you find targeted ads creepy? If yes which ads do you find creepy?
12. What do you do (if anything) when you see a creepy ad?

Questionnaires - 3

1. Were you surprised by the content of this video?
 - To a Great Extent
 - To Somewhat
 - Very Little
 - Not at All

2. Did the video contain significant information that you did not know previously about online tracking and ads?
 - Strongly Agree
 - Agree
 - Undecided
 - Disagree
 - Strongly Disagree

3. After watching this video, are you more or less interested in using tools to block ads and limit tracking?
 - To a Great Extent
 - Somewhat
 - Very Little
 - Not at All

4. How comfortable are you with the amount of tracking done by ad companies?
 - Extremely
 - Very Moderately
 - Slightly
 - Not at all

Questionnaires - 4

1. How have your views on targeted ads changed after watching the video? Describe in your words.
2. Would you like ads based on your interests? Why or why not?
3. What online activity would you want ad networks to not track?
4. What do you do when you want to avoid online tracking?
5. Have you used any ad-blocking tools? Which ones? Did you have any issues when using them?
6. After watching the video do online ads seem creepier than before?

Questionnaires - 5

1. How often do you try to get more information about why an ad is displayed?

- Always
- Very Often
- Rarely
- Never
- I didn't know this was possible

2. How often do you provide ad feedback?

- Always
- Very Often
- Sometimes
- Rarely
- Never

3. Do the feedback options influence your decision in providing the ad-feedback?

- To a Great Extent
- Somewhat
- Very Little
- Not at All

4. Do you feel that the ad feedback changes the ads you see in the future?

- Definitely
- Very Probably
- Probably
- Probably Not
- Definitely Not

5. Does the ability to provide ad feedback makes the online ads more acceptable?

- Strongly Agree
- Agree
- Not sure
- Disagree
- Strongly Disagree

Questionnaires - 6

1. Have you seen ad information online as shown in the screenshots?
2. Do you remember seeing this type of information in any ads? Did you get further information on any ads?
3. For what kind of ads will you provide ad feedback and select “stop seeing this ad”?
4. Are you aware of any ad-feedback system apart from the one you just saw?
5. Do you think providing ad feedback will make online ads less creepy?

Questionnaires - 7

1. Do you think it is feasible for ad companies to forget (delete) information that they had previously gathered about individuals?

- Yes
- No
- I'm not sure

2. If ad companies delete information that you don't want them to save, will that change your acceptance of online tracking and targeted ads?

- Definitely
- Very Probably
- Probably
- Probably Not
- Definitely Not

3. If you had better control on tracking will you be more open towards accepting online tracking and online ads?

- Strongly Agree
- Agree
- Undecided
- Disagree
- Strongly Disagree

Questionnaires - 8

1. What comes in your mind when you hear forgetting tracked information? Which type of information you are thinking of?
2. Would having control over what ad networks stored about you (if you could omit information from your profile) change how comfortable you were with online tracking? What about targeted ads?
3. Do you think having a “forget” option for ads would help make targeted ads less creepy?
4. Do you think being able to delete information from your online history would make online tracking less creepy?
5. Do you have any suggestions or comments?

Study Script

Hello! Greetings. How are you today? So, as you know you are here to participate in a study on Targeted ads and web tracking. Let's start with some introduction.

- *What is your name?*
- *What is your age?*
- *What is your highest education?*
- *How much time do you spend online?*

Thank you for your response.

Let's begin some multiple-choice questions. These are also called scale questions. You will have four rounds of scale questions. You just need to choose an option to answer the questions as per your best knowledge and feelings.

- *So, are you ready?*

So here you go. Please fill these questions by selecting any one option and let me know when you are done.

- *Hand participant Questionnaire 1*

Thank you.

- *How did you find the questions? Did you face any difficulty or any problem understanding the questions?*

Let's proceed further with a small interview. You don't need to write but just explain your answer/ response in your words. It is just like an interview. Before proceeding with next set of questions I need to explain you the definition of creepiness in terms of this study.

Creepiness in general is a feeling when you are been watched by a stranger while you are doing something on you own. Ex. Some stranger watching you from your window is when you feel creepy.

- *Do you have any questions or concern before starting the next round?*
- *Ask questions from Questionnaire 2*

I hope you are finding everything ok. Now I am going to show you a video which will provide you more information about targeted ads and web tracking. This video will also help you understand questions in upcoming rounds.

- *Show them seven-minute information video,
“How advertisers use internet cookies to track you”
<https://www.wsj.com/video/how-advertisers-use-internet-cookies-to-track-you/92E525EB-9E4A-4399-817D-8C4E6EF68F93.html>*

(Note this video was used in a study reported in “Smart , Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising.” Blase Ur, Pedro G. Leon, Lorrie Faith Cranor, Richard Shay and Yang Wang)

- *How did you find the video? Was it informative?*

Here is second round of scale questions for you. Please fill these questions by selecting an option. If you have any query please feel free to ask.

- *Handing participant Questionnaire 3*

Thank you again. Are you ready for the second interview? Now, you must be aware of the interview format? Am I correct? Let's begin.

- *Asking questions from Questionnaire 4*

We are doing pretty good. So, now we have something interesting for you. I am going to show you few mock-ups. Do you know what mock up is? If not, mock-up is a pictorial representation of a model or a structure. In our case these are screenshots/pictures of web application YouTube.

- *Showing them mock up 1.*

These are the screenshots taken from YouTube on an iPhone. These screenshots were taken when you see a YouTube ad. Few screenshots show the ad feedback system of YouTube and ad information. Also, it displays Google Settings page where it shows you why they have shown you this ad and what information they have of you. Have a look. If you don't understand anything just let me know.

So, first page is the YouTube home page that is what you see when you open YouTube.

Next page is showing the YouTube ad. I suppose you must have seen YouTube ads very often. If not, this is how it looks like.

- *Have you ever clicked on any YouTube ad before? Have you seen the ad information? If not, here is how it looks.*

After clicking on the three-dot drop-down menu here what we see.

Further if you click on 'Why this ad?' it will display you the below information.

What happens when you further click on Google's Ad Settings link on the pop-up. Here is what you see.

So, you saw the Google ad personalization page. Now next is what you see when you click on 'Stop seeing this ad'

When you select not to see this ad YouTube will ask you to provide optional feedback. And this is how YouTube Ad feedback system will appear.

If you choose to provide your feedback you get a confirmation from YouTube which looks like these...

Ok. So, now after this mock-up we have another set of scale question related to the mock-up.

- *Hand participants Questionnaire 5*

Great! We are almost at the mid of our study. Next we have another interview related to previous mock-up we just saw.

Asking Questions from Questionnaire 6

Great! Now it is time for next round of mock ups. So here we go. You already saw how YouTube ad feedback system looks like. You also saw the options there. Now we have just made one improvement in the existing ad feedback system. Let's see. You need to guess the difference.

- *Showing them Mock-up 2*

So, this is how the new ad feedback will look like.

If you can make ad companies 'forget' information about you, which is used to display creepy ads.

Were you able to guess what was the difference? If not, we have the mock-up of existing YouTube ad feedback system for your revision.

I need to explain you the purpose of forget in this context. Let me give you an example. If you are often searching for baby products and pregnancy then google assumes that you are either a parent or going to be parent soon so they create a profile that way assuming that you are either pregnant or a parent and will show you related ads on YouTube and other social platforms. In order to get rid of this information stored within Google you can select forget. Forget will help you get rid of such creepy ads.

Finally, we have managed to reach to the last segment of our study. So, we are just left with one scale question round and one interview. So, here we go with the last scale question round.

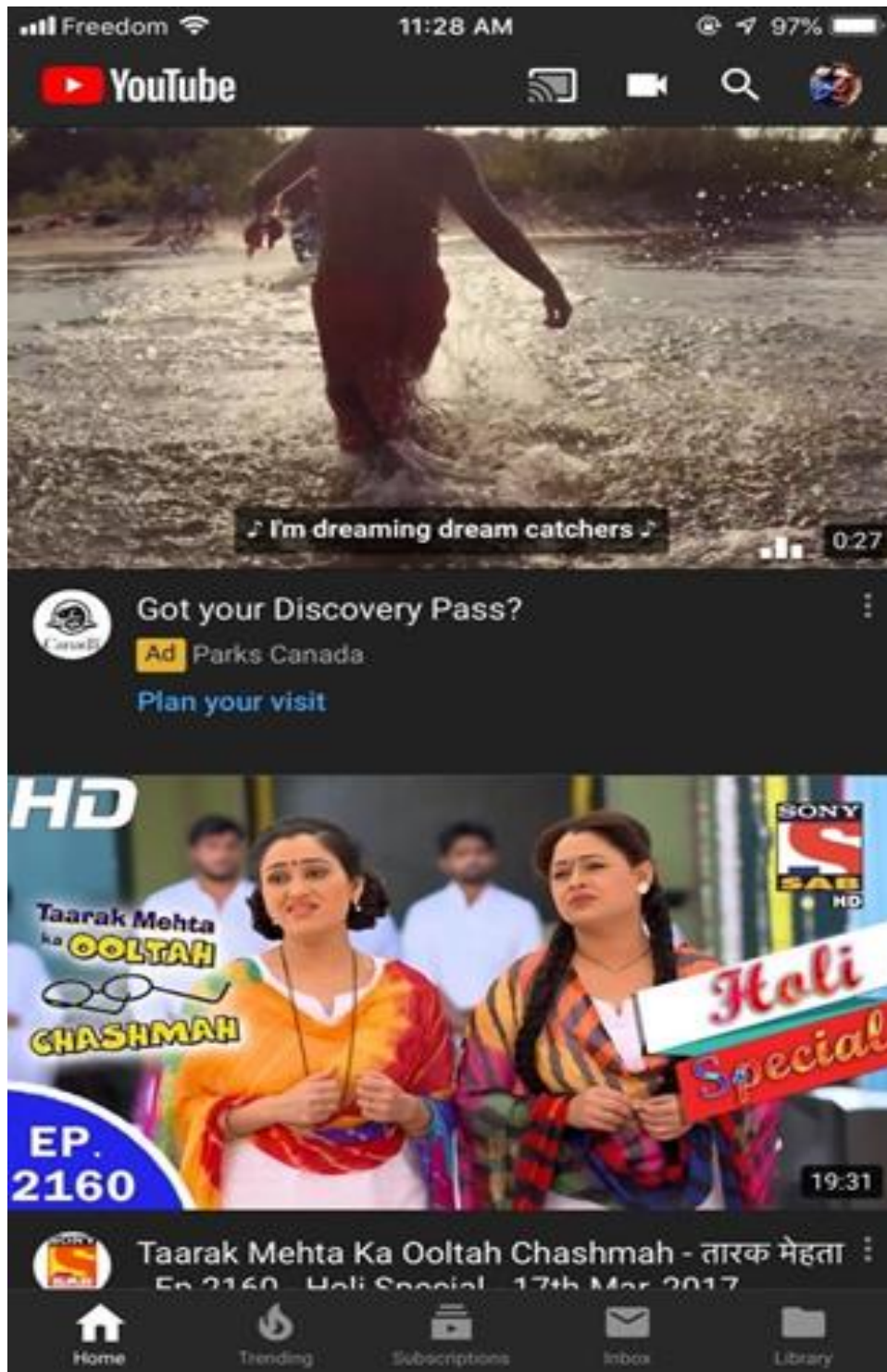
- *Hand participant Questionnaire 7*

- *So, are you ready for the last interview? Here we go.*

Asking Questions from Questionnaire 8

Mock up - 1

1. YouTube Home page



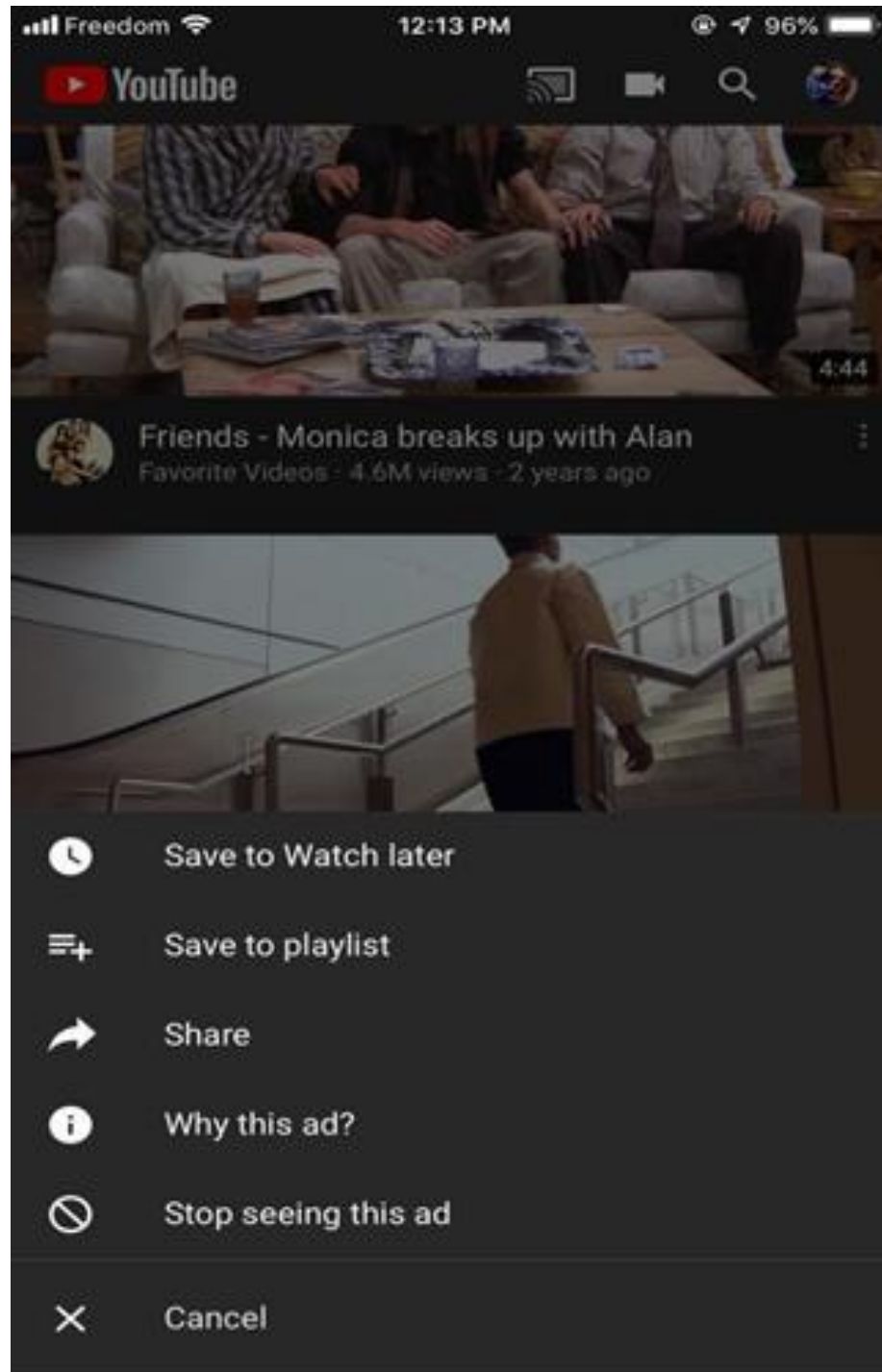
2. YouTube Ad: -

The image is a screenshot of a YouTube mobile application interface. At the top, the status bar shows 'Freedom' carrier, Wi-Fi signal, '11:53 AM' time, and '97%' battery. The main content area features a video player. The first video is from the channel 'Friends', titled 'Monica breaks up with Alan', with 4.6M views and posted 2 years ago. The video thumbnail shows three people sitting on a couch. The second video is an advertisement for O-Train Confederation Line, titled 'Get Ready for Rail', with a yellow 'Ad' label. The thumbnail shows a red and blue double-decker bus with 'OC TRANSP' and '8157' on its destination sign. A text overlay on the bus reads 'when you transfer at a major station on the Oh train Confederation line you will'. Below the ads, there is a banner for 'EPISODE # 06' in HD 1080p. At the bottom, a navigation bar contains icons for Home, Trending, Subscriptions, Inbox, and Library.

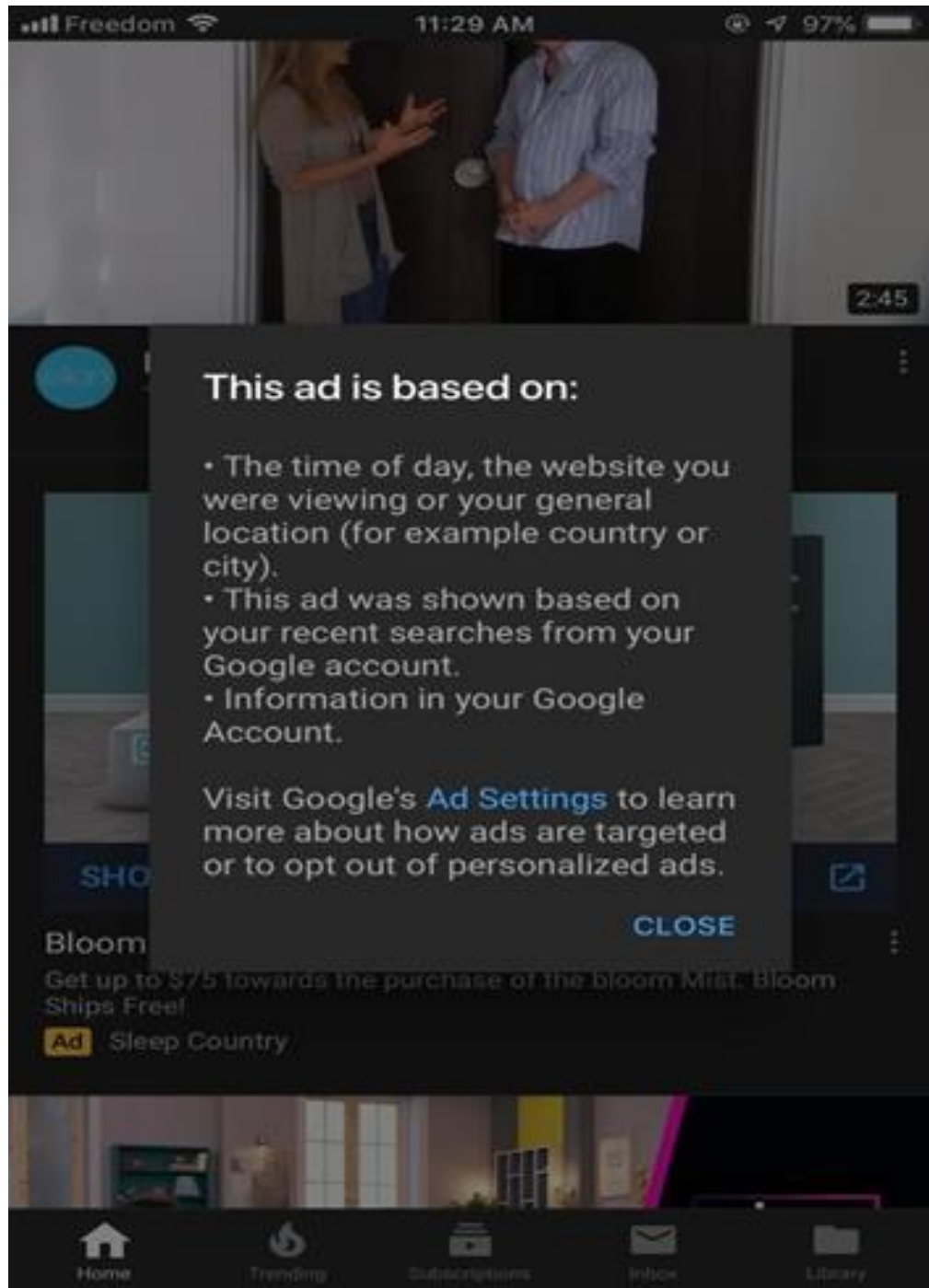
3. Click on Ad information



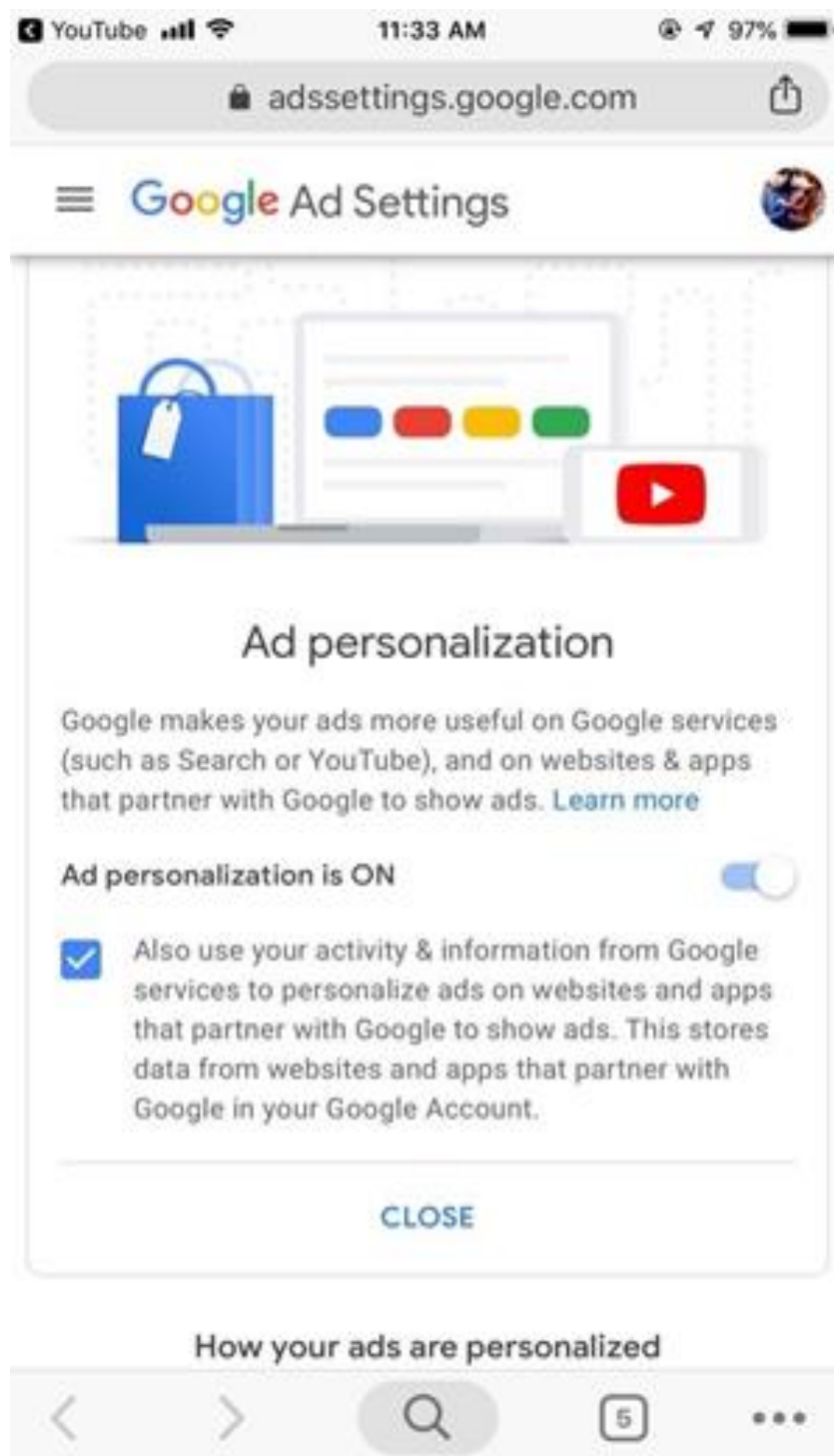
4. Ad information menu



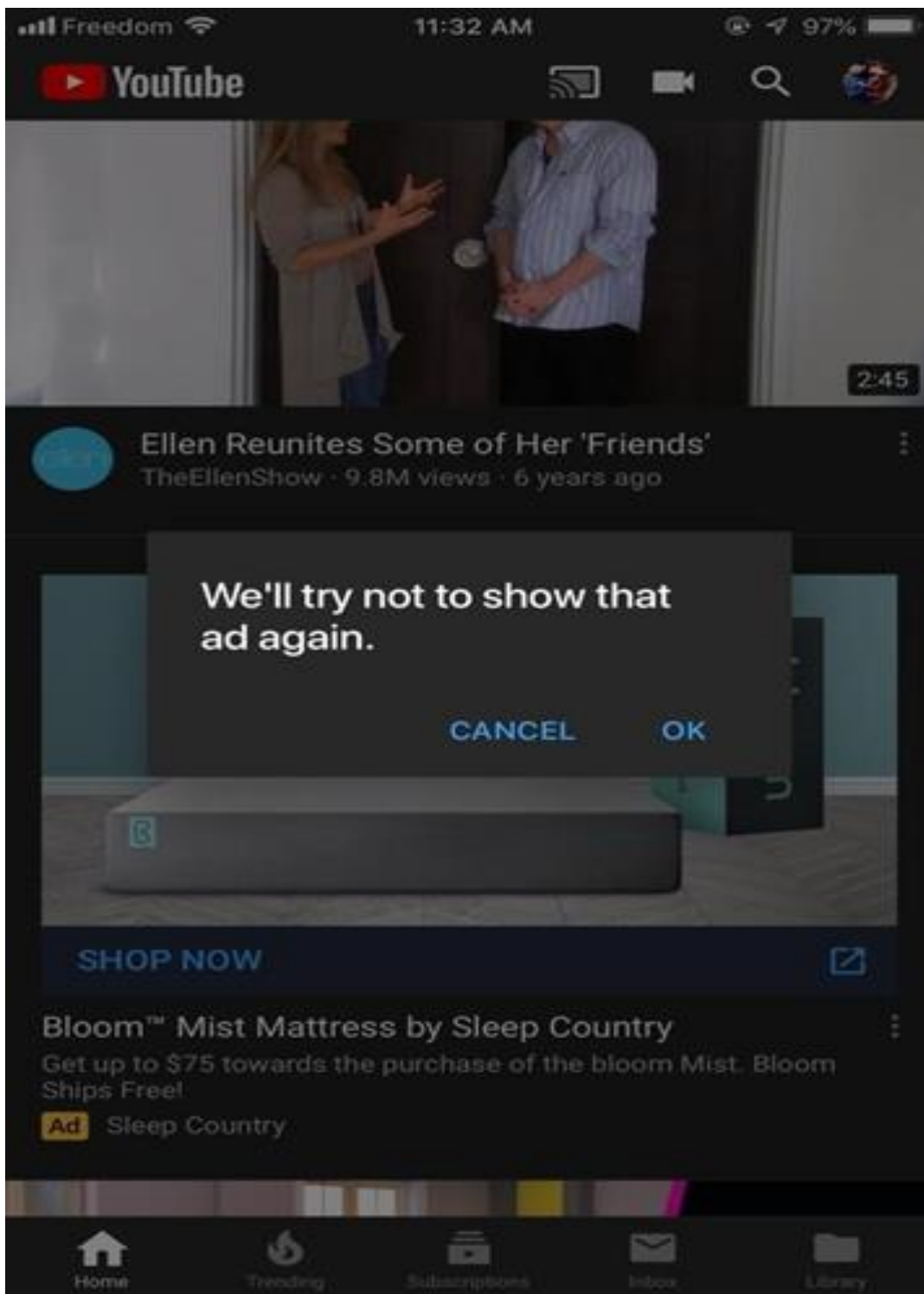
5. When you click on 'Why this Ad?'



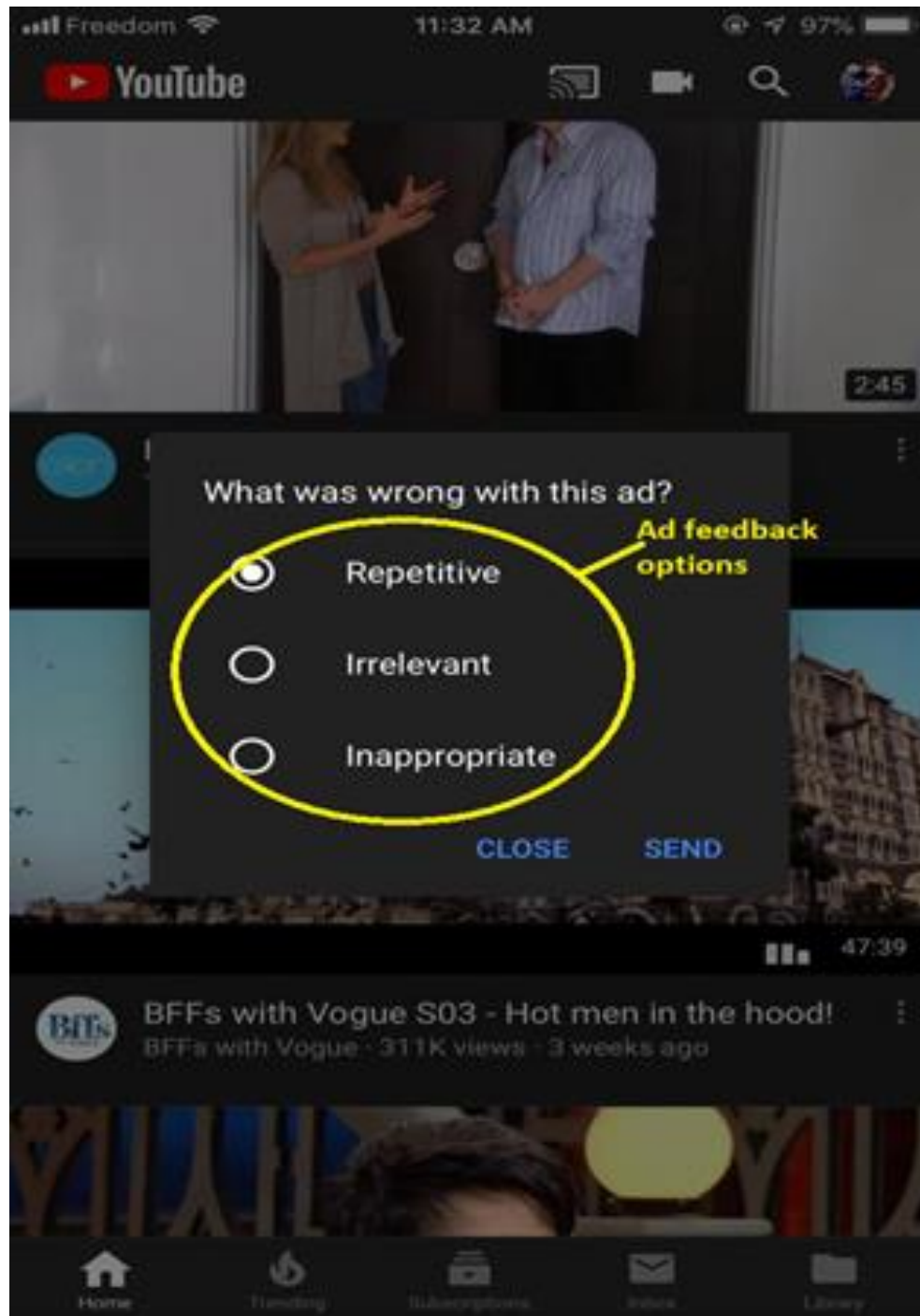
6. When you click on 'Google's Ad Settings'



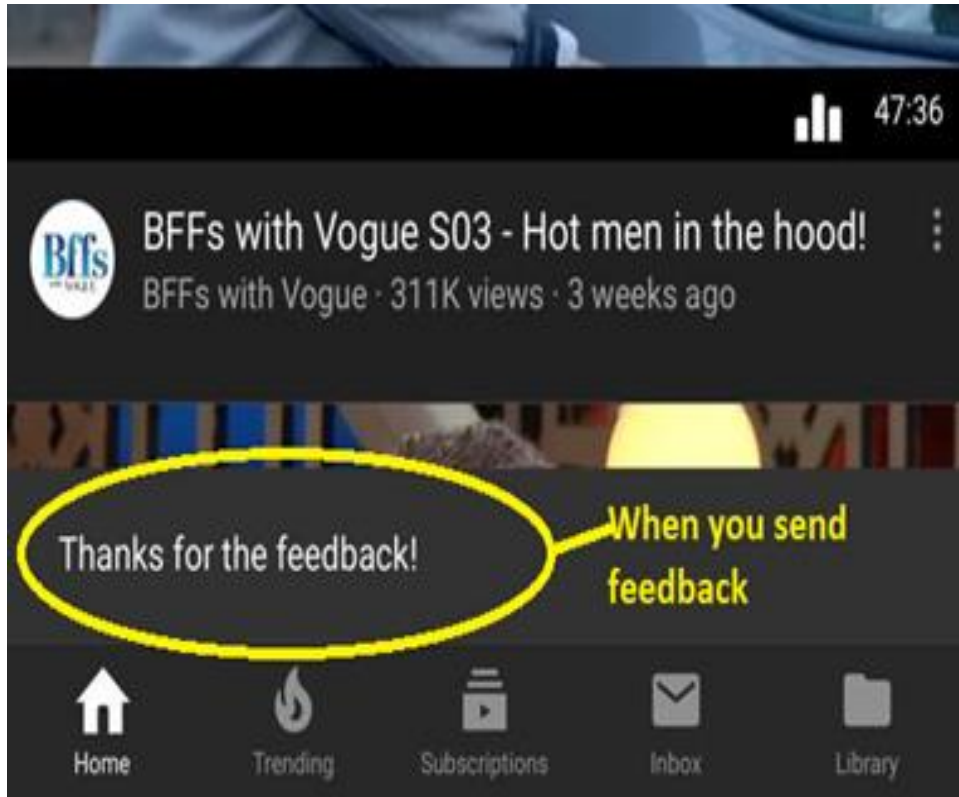
7. When you click on 'Stop seeing this ad'



8. YouTube asking for Ad feedback

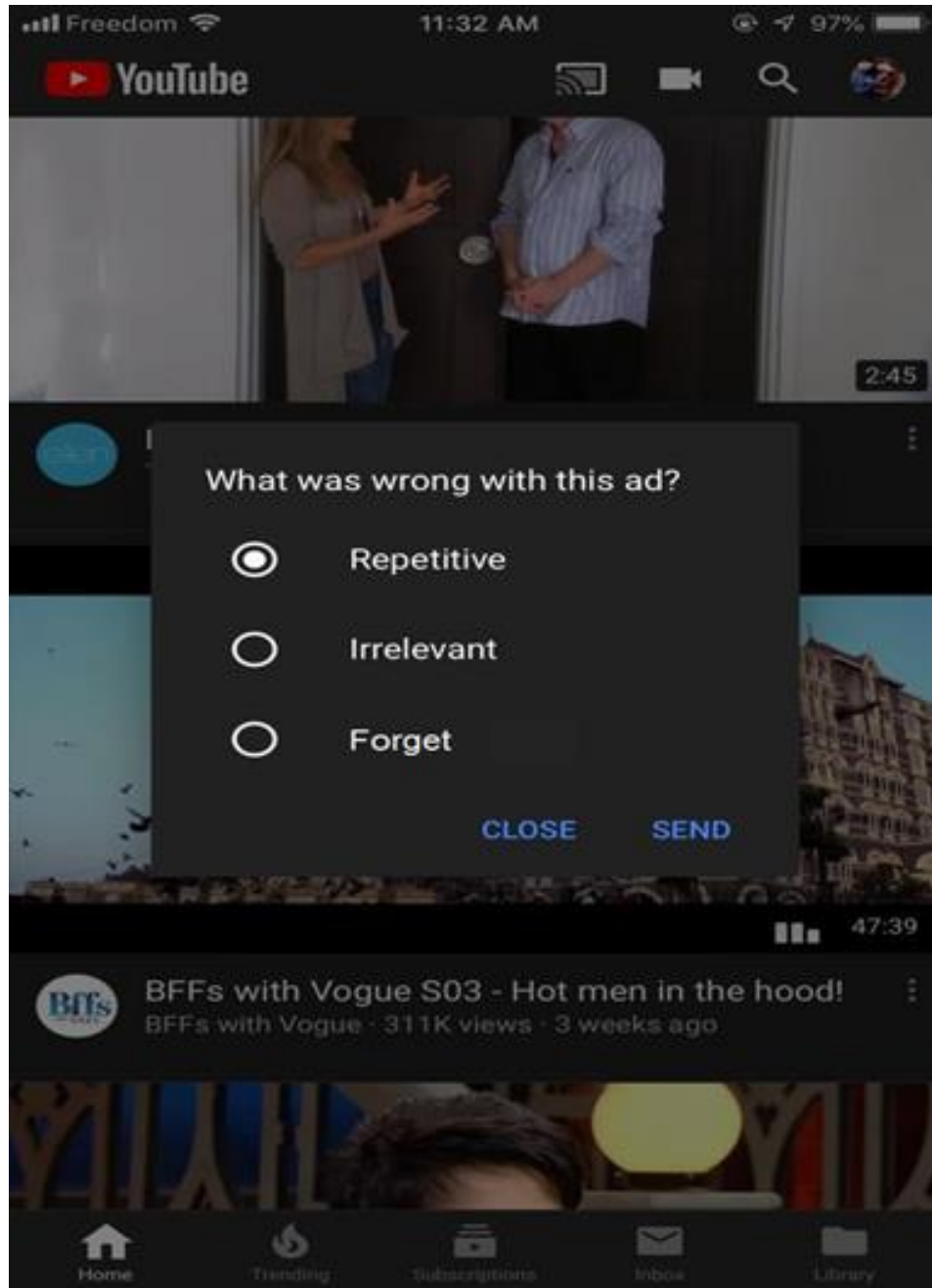


9. When you send the feedback

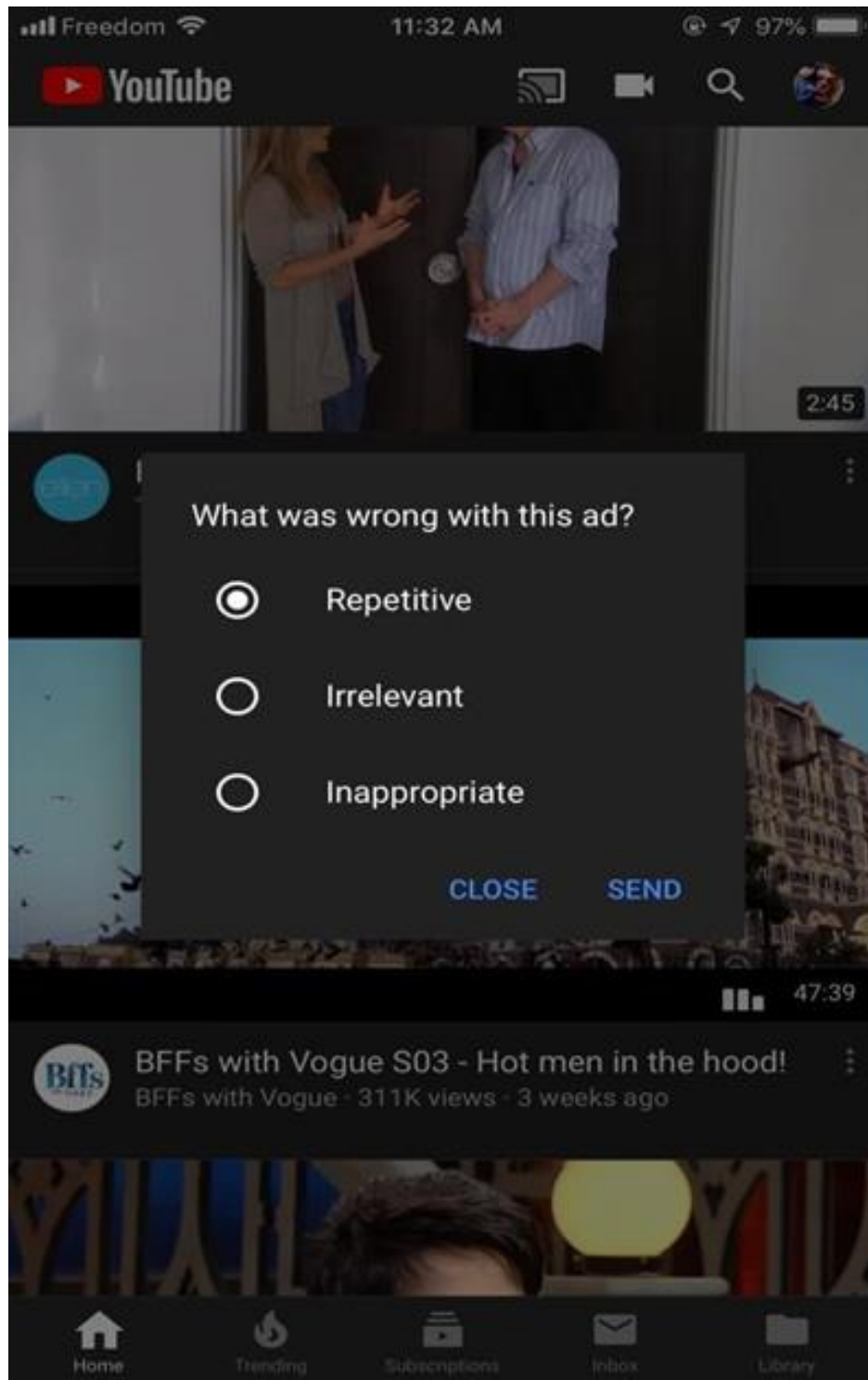


Mock up - 2

1. If you can make ad companies 'forget' information about you which they use to display creepy ads: -



2. Existing YouTube Ad Feedback options: -



Appendix B

Study 2

This appendix contains questionnaires and mock-ups used for study 2. The flow of this Appendix is as follows:

- Consent Form
- Poster
- Email Recruitment
- Online Recruitment
- Questionnaire
- Mock up 1
- Mock up 2
- Study Script

Research Consent Form

Name and Contact Information of Researchers:

Vidhi Kirit Shah, Carleton University, School of Computer Science

Tel.: 613-400-9799

Email: vidhikiritshah@cmail.carleton.ca

Supervisor and Contact Information: *Prof. Anil Somayaji, Carleton University, School of Computer Science.*

Project Title

Users Perceptions of Targeted Advertisements and Online Tracking

Carleton University Project Clearance

Clearance #: CUREB-B Clearance #111390

Date of Clearance: October 16, 2019

Invitation

The information in this form is intended to help you understand what we are asking of you so that you can decide whether you agree to participate in this study. Your participation in this study is voluntary, and a decision not to participate will not be used against you in any way. The researcher for this study is a master's student, Vidhi Kirit Shah. She is working under the supervision of Prof. Anil Somayaji in the Computer Science Department. As you read this form, and decide whether to participate, please ask all the questions you might have, take whatever time you need, and consult with others as you wish.

What is the purpose of the study?

The aim of this study is to explore participants' perceptions of targeted advertisements and online tracking. Users of online applications are constantly bombarded with targeted ads based on their online behavior which includes their browsing history, location, their time spent on websites, etc. In some cases, users find some ads very creepy as user feel that the ad network knows more about them than they would like. This information may be embarrassing or otherwise sensitive. This research revolves around whether changes in existing tracking systems could improve user acceptance of online tracking for advertising purposes.

What will I be asked to do?

If you agree to take part in the study, we will ask you to answer some open-ended questions regarding your knowledge about targeted advertisements and web tracking. As part of this, we will show you some screenshots and mock-ups related to YouTube ads and Google Ad settings to understand your awareness about them. Please note that you are not being tested; we are only interested in your perception of cybersecurity warning messages. This research study will be audio-recorded for the purposes of transcription and analysis.

The study will take around 30 minutes and will involve three rounds of questions and observing mock-ups followed by post-mockup interviews. Each round of questionnaires will include open-ended questions. There are no predictable risks in participating in this study. You will not be asked to disclose any personally identifiable information.

Risks and Inconveniences

We do not anticipate any risks to participating in this study.

Possible Benefits

You may not receive any direct benefit from your participation in this study. However, you may learn more about how online ad tracking works. The results from this study will be used to make recommendations for better opt-out options from creepy targeted ads.

Compensation/Incentives

As a token of appreciation, you will receive a \$10 Tim Horton's gift card.

No waiver of your rights

By signing this form, you are not waiving any rights or releasing the researchers from any liability.

Withdrawing from the study

If you withdraw your consent during the study, all information collected from you before your withdrawal will be discarded. You may withdraw at any time before October 30, 2019.

Confidentiality

We will treat your personal information as confidential, although absolute privacy cannot be guaranteed. No information that discloses your identity will be released or published without your specific consent. Research records may be accessed by the Carleton University Research Ethics Board in order to ensure continuing ethics compliance.

The results of this study may be published or presented at an academic conference or meeting, but the data will be presented so that it will not be possible to identify any participants unless you give your express consent. You will be assigned a code so that your identity will not be directly associated with the data you have provided. Recordings and questionnaires will be destroyed after they are transcribed. Transcribed data and analysis will be kept in a password-protected file on a secure computer. The master list associating your name with your code will be kept on paper on a master list stored in a secure location. This list will be destroyed in six months.

What if I do not want to be audio-recorded?

If you choose to not be audio-recorded, you may still participate in the study. We will take notes of what you said during the study.

Data Retention

The audio recording and the questionnaires shall be destroyed once they are transcribed. The transcriptions will be kept for two years for the purpose of publication. Participant's contact information will not be stored for future recruitment and shall be deleted after the project completion (at most two years). The anonymized analysis shall be archived (in the CCSL and on the researcher's personal machine) after thesis is defended and approved by the committee.

New information during the study

In the event that any changes could affect your decision to continue participating in this study, you will be promptly informed.

Ethics review

This project was reviewed and cleared by the Carleton University Research Ethics Board B. Should you have any ethical concerns with the study, please contact the REB Chair, Carleton University Research Ethics Board-B (by phone: 613-520-2600 ext. 4085 or by email: ethics@carleton.ca). For all other questions about the study, please contact the researcher.

Statement of consent – print and sign name

I voluntarily agree to participate in this study. Yes No

I agree to be audio recorded. Yes No

Signature of participant

Date

Research team member who interacted with the subject

I have explained the study to the participant and answered any and all their questions. The participant appeared to understand and agree. I provided a copy of the consent form to the participant for their reference.

Signature of researcher

Date



Participate in a Study about Targeted advertisement and Web Tracking.

To participate in this study, you must be:

- Familiar with online ads
- At least 18 years old
- Comfortable in the English language

This is a **30-minute** study. You will be asked to answer questions about targeted advertisement and web tracking done by ad networks.

Participants will be compensated with a \$10 Tim Horton's gift card.

The ethics protocol for this project has been reviewed and cleared by the Carleton University Research Ethics Board. Should you have any ethical concerns with the study, please contact the REB Chair, Carleton University Research Ethics Board-B (by phone: 613-520-2600 ext. 4085 or by email: ethics@carleton.ca). For all other questions about the study, please contact the researcher.

If you are interested in participating, please email Vidhi Kirit Shah at: vidhikiritshah@mail.carleton.ca

This research has been cleared by the Carleton University Research Ethics Board (CUREB-B), REB clearance #(CUREB-B Clearance #111390).

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Targeted Advertisement
vidhikiritshah@mail.carleton.ca

Email Invitation

Subject: A research study to understand user perception on Creepy targeted ads and their acceptance of online tracking.

Dear Sir/Madam,

My name is Vidhi and I am a Master's student in the Computer Science department at Carleton University. I am working on a research project under the supervision of Prof. Anil Somayaji.

I am writing to you today to invite you to participate in a study entitled "Users Perceptions of Targeted Advertisements and Online Tracking". This study aims to understand users' perception on creepiness of targeted advertisement and their acceptability of online tracking.

The study will take approximately 30 mins. As a participant in the study, you will observe screenshots and mock-ups and discuss open-ended questions.

We are looking for adult participants over 18 years old, who have minimal understanding of targeted ads and web tracking and is comfortable speaking and reading in English.

You will have the right to end your participation in the study at any time, for any reason, up until October 30, 2019. If you choose to withdraw, all the information you have provided will be destroyed.

Participants will be compensated with a \$10 Tim Horton's gift card.

If you are interested in participating, please email Vidhi Kirit Shah at: vidhikiritshah@gmail.com

This research has been cleared by the Carleton University Research Ethics Board (CUREB-B Clearance #111390).

Should you have any ethical concerns with the study, please contact the REB Chair, Carleton University Research Ethics Board-B (by phone: 613-520-2600 ext. 4085 or by email: ethics@carleton.ca). For all other questions about the study, please contact the researcher.

Sincerely,

Vidhi Kirit Shah

Online Invitation

Post on Carleton Research Participants Facebook group.

Volunteers needed for research study on targeted advertisement and web tracking.

My name is Vidhi and I am a master's student in the Computer Science department at Carleton University. I am working on a research project under the supervision of Prof. Anil Somayaji.

I am writing to you today to invite you to participate in a study entitled “Users Perceptions of Targeted Advertisements and Online Tracking”. This study aims to understand users’ perception on creepiness of targeted advertisement and the acceptability of online tracking.

The study will take approximately 30 mins. As a participant in the study, you will observe screenshots and mock-ups and discuss open-ended questions.

We are looking for adult participants over 18 years old who have minimal understanding of targeted ads and web tracking and are comfortable speaking and reading in English.

You will have the right to end your participation in the study at any time, for any reason, up until October 30, 2019. If you choose to withdraw, all the information you have provided will be destroyed.

Participants will be compensated with a \$10 Tim Horton’s gift card.

If you are interested in participating, please email Vidhi Kirit Shah at: vidhikiritshah@gmail.com

This research has been cleared by the Carleton University Research Ethics Board (CUREB-B Clearance #111390).

Should you have any ethical concerns with the study, please contact the REB Chair, Carleton University Research Ethics Board-B (by phone: 613-520-2600 ext. 4085 or by email: ethics@carleton.ca). For all other questions about the study, please contact the researcher.

Sincerely,

Vidhi Kirit Shah

Questionnaires - 1

Definition of Creepiness: - Creepiness in general is a feeling you get when you are non-consensually observed while engaging in private behavior. For example, a stranger watching looking at you through your window while you are doing your own thing at home is normally creepy.

1. How much time do you spend online every day?
2. Do you ever see ads related to your interests? Can you describe one or two ads related your interests?
3. What do you know about online tracking and targeted advertisement? Are you aware of it? If yes, can you describe it in your word?
4. What information you think the ad companies collect about you for targeting ads?
5. What is the definition of creepiness for you in terms of tracking and online ads? Can you give an example?
6. What online activity would you want ad networks to not track?
7. Do you find any ad creepy? If yes, what kind of ad (which ad) do you usually find creepy?
8. What do you do (if anything) when you see a creepy ad?
9. Are you aware of any opt-out options for online tracking? If yes, can you describe them.

Questionnaires - 2

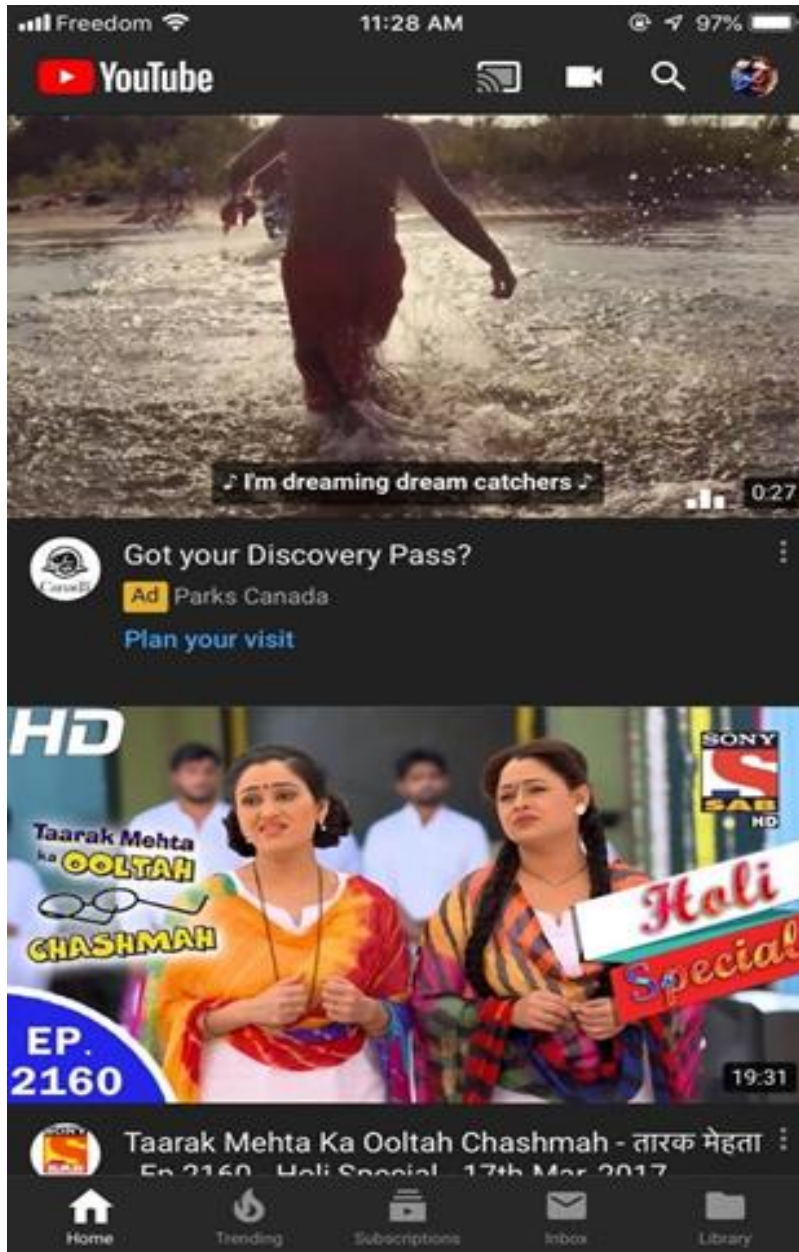
1. How much useful do you find the 'This ad is based on:' information shown in the screenshots? Can you explain.
2. How useful do you find the ad options? (When you click on the three dots on the ad you get an ad menu with several options. How useful do you think these options are?)
3. Have you ever visited Google Ad Settings page to check how your ads are personalised?
If not, were you aware that this type of information exists?
4. How useful do you feel is the Google Ad Settings page?

Questionnaires - 3

1. Would having control over what ad networks stored about you (if you could omit information from your profile) change how comfortable you were with online tracking and targeted ads?
2. Do you think having a “click to forget searches related to this ad” option will provide you more control on online tracking and targeted ads?
3. For what type of ads would you choose to click on the forget searches related to this ad? Is there any ad in particular?
4. On the scale of 1 to 5 how useful do you think the forget option is in terms of getting control over tracked information and creepy ads? 1 is least and 5 is most.
5. Do you have any suggestions or comments?

Mock Up 1 for Study 2

1. YouTube Home page



2. When ad related to Credit Card is shown: -

The screenshot shows a mobile device interface with a YouTube video player. At the top, the status bar displays 'Freedom' carrier, signal strength, Wi-Fi, time '1:16 PM', location, and 74% battery. The video title is 'Real Life Couples' with a duration of 3:47. Below the video, the channel name is 'Celebrity News' and the video title is 'Real Life Partners of Yeh Rishta Kya Kehlata Hai Actors - 15 August 2018 - Today Episode', with 4M views and posted 1 year ago. An advertisement for Capital One credit cards is displayed. The ad features an image of a Capital One Guaranteed Mastercard with the number 5457 5656 7890 1234 and the name R THOMAS. The text of the ad says 'Get a credit limit between \$300 and \$7,000!' and includes an 'APPLY NOW' button with a share icon. Below the ad, the text reads 'Get a \$300-\$7,000 Credit Limit' and 'It's easy to apply for a Guaranteed Mastercard® from Capital One®.' An 'Ad' label and 'Capital One Canada' are also present. At the bottom of the video player, there are three small images of women. The bottom navigation bar includes icons for Home, Trending, Subscriptions, Inbox, and Library.

3. When you click on Ad information: -

Freedom 1:16 PM 74%

Real Life Couples

3:47

Real Life Partners of Yeh Rishta Kya Kehlata Hai Actors - 15 August 2018 - Today Episode
Celebrity News · 4M views · 1 year ago

Get a credit limit between \$300 and \$7,000!

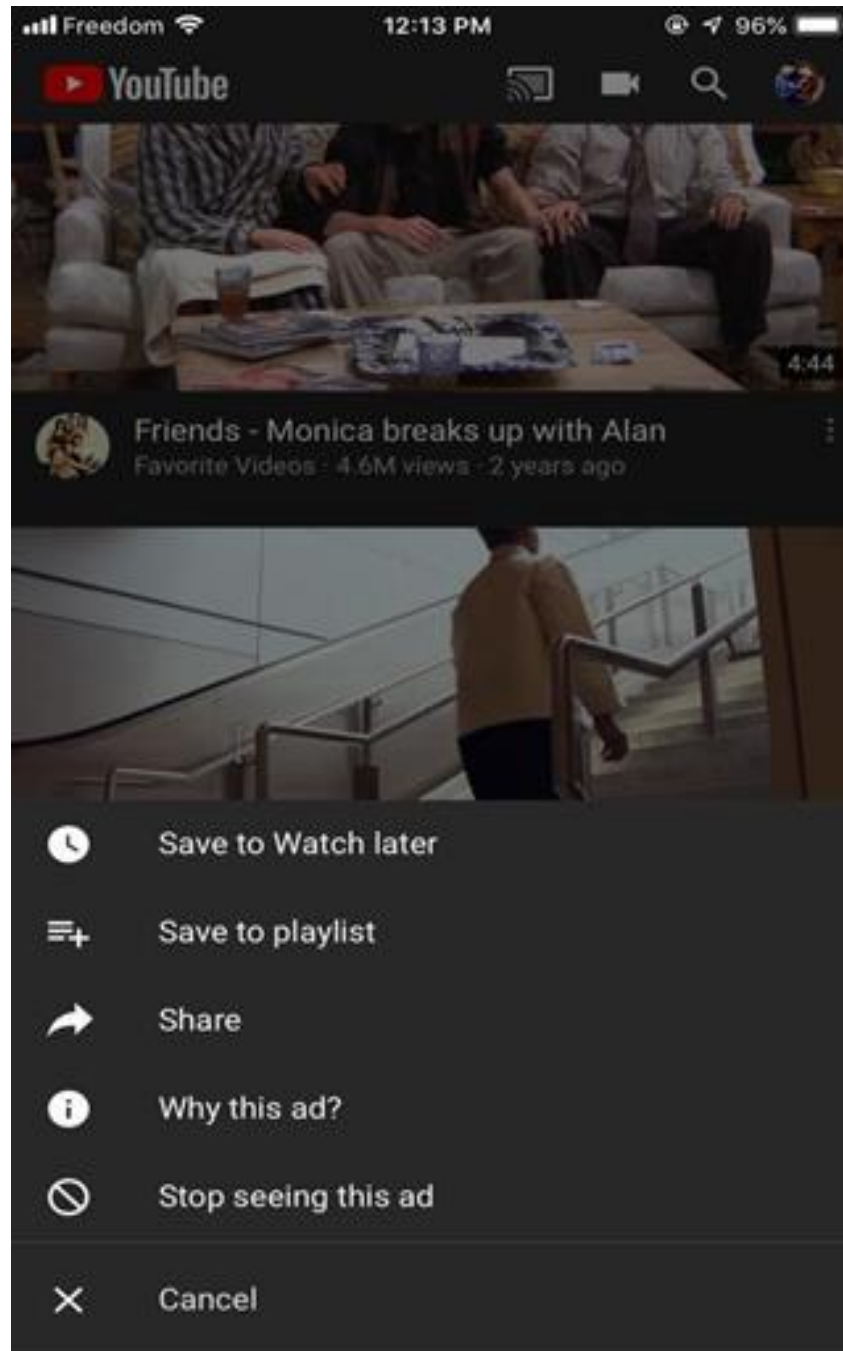
APPLY NOW

Get a \$300-\$7,000 Credit Limit
It's easy to apply for a Guaranteed Mastercard® from Capital One®.
Three-dot for more information on ad.

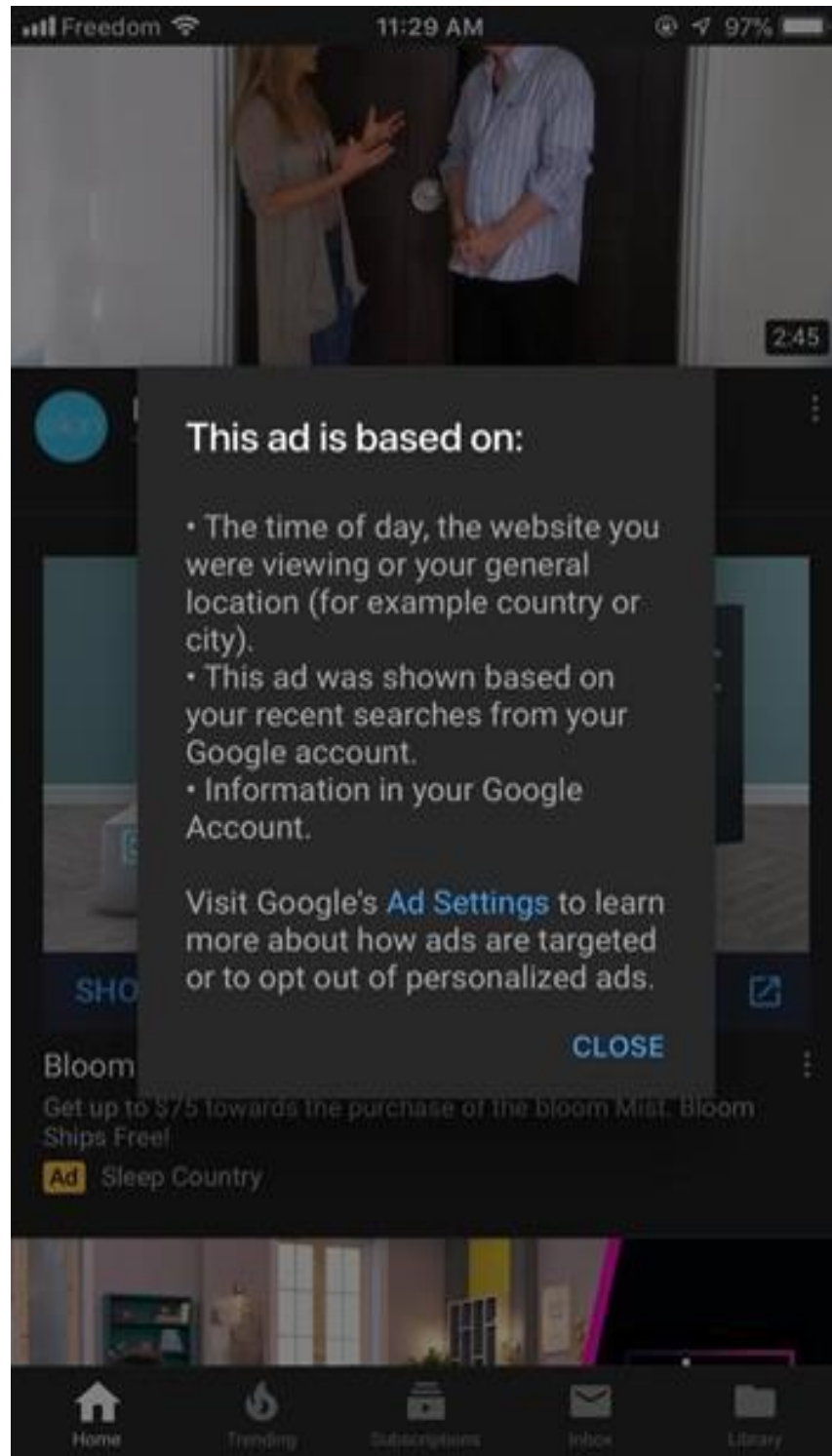
Ad Capital One Canada
Shows that this is an ad.

Home Trending Subscriptions Inbox Library

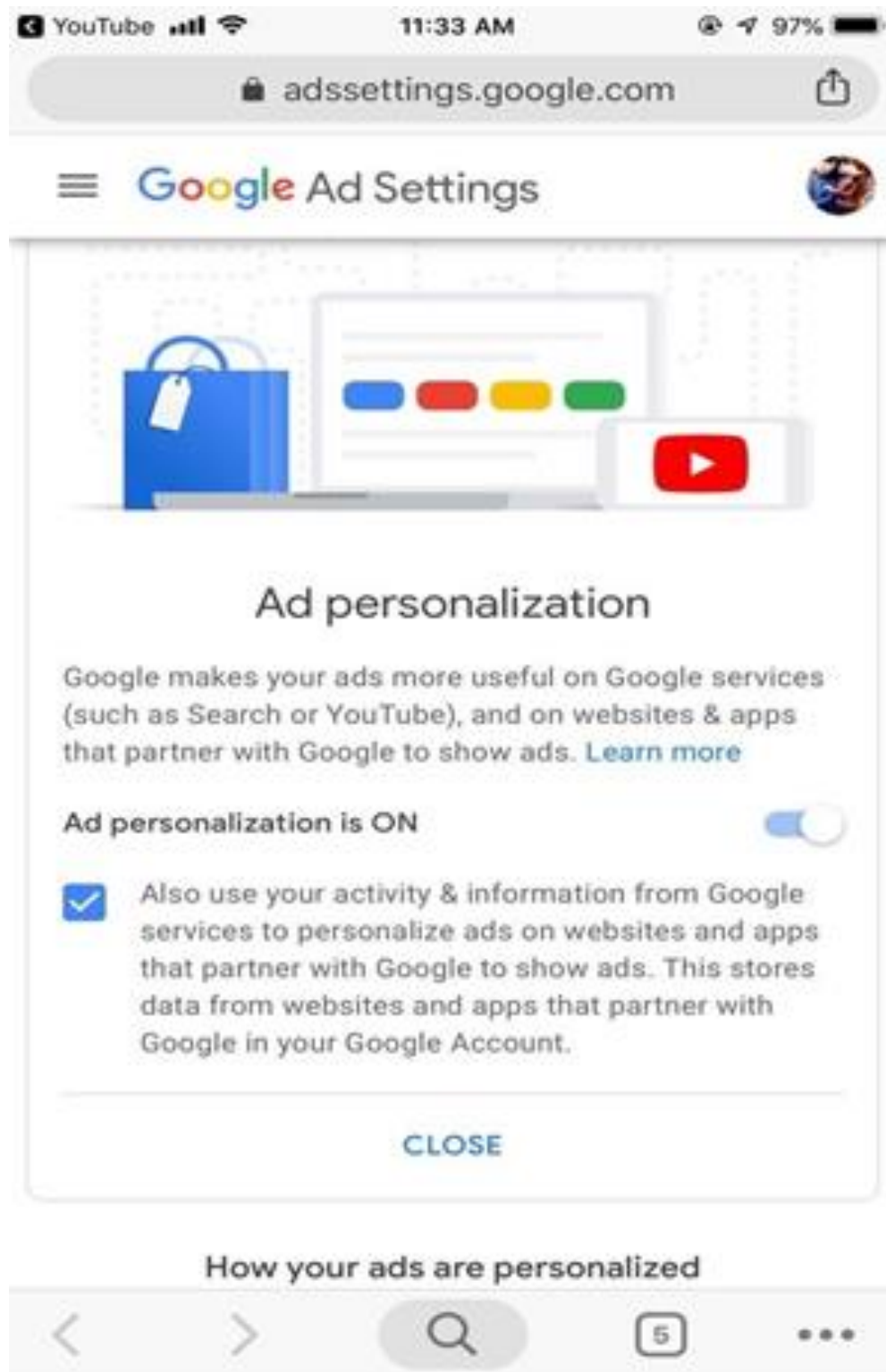
4. Ad information menu: -



5. When you click on 'Why this Ad?'



6. When you click on 'Google's Ad Settings'



7. Google Ad Settings page show all your interest under your ad profile: -

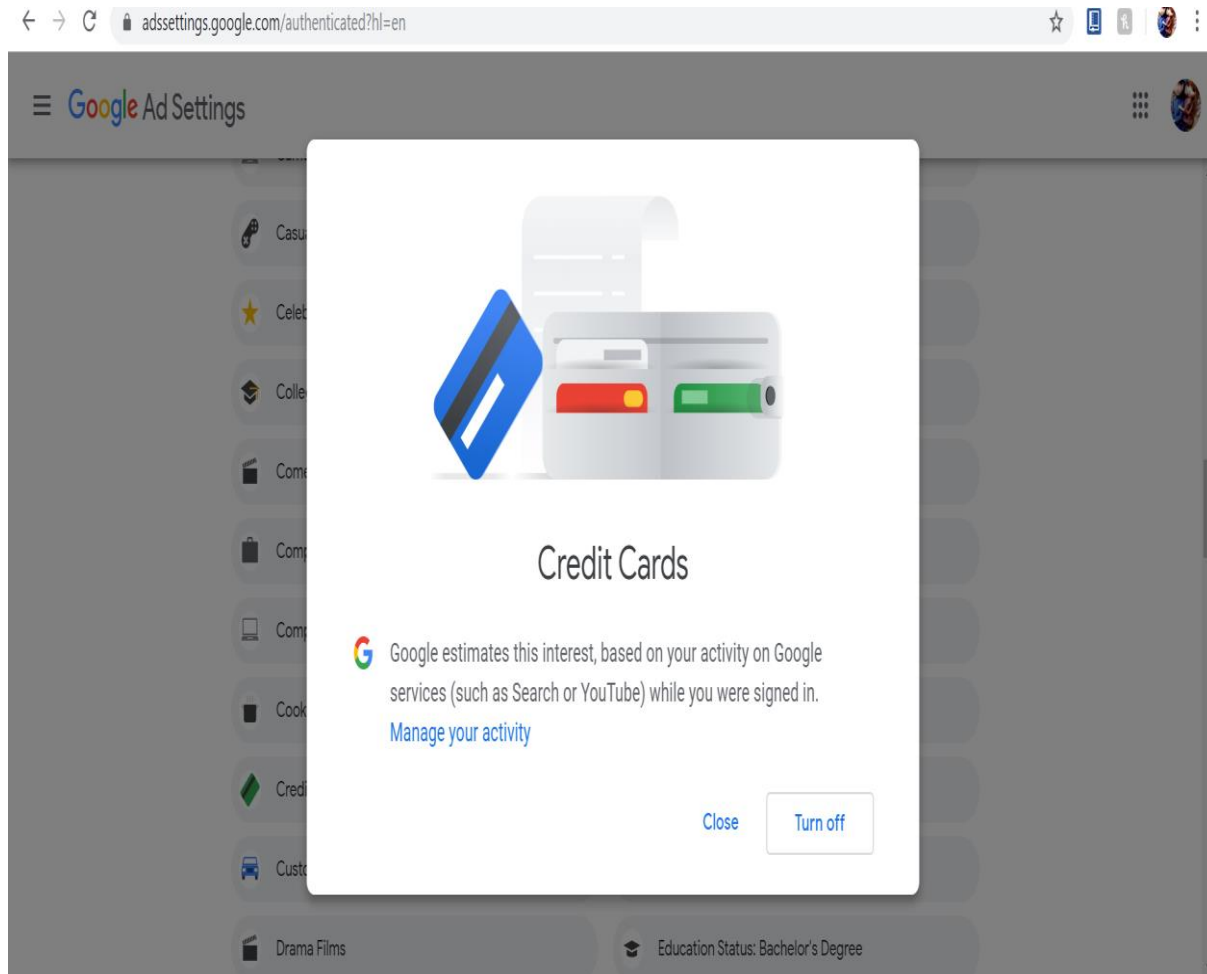
The screenshot shows the Google Ad Settings page. At the top, it indicates that "Ad personalization is ON" with a toggle switch. Below this, there is a "MORE OPTIONS" link. The main section is titled "How your ads are personalized" and explains that ads are based on personal info, advertiser data, and Google's interest estimation. A "Learn more" link is provided. Below the text, there is a grid of interest categories, each with an icon and a label:

- 25-34 years old (Candle icon)
- Female (Person icon)
- Hyundai Auto Canada (H icon)
- Walmart (W icon)
- tripcentral.ca™ (T icon)
- Advertising & Marketing (Megaphone icon)
- Android OS (Android robot icon)
- Antivirus & Malware (Checkmark icon)

This screenshot shows a more detailed view of the Google Ad Settings page. The browser address bar and the "Google Ad Settings" header are visible. Below the header, there is a grid of interest categories, each with an icon and a label. The "Books & Literature" category is highlighted with a grey background:

- Android OS
- Antivirus & Malware
- Apparel
- Apple iOS
- Audio Equipment
- Autos & Vehicles
- Beaches & Islands
- Beauty & Fitness
- Blues
- BMW
- Board Games
- Bollywood & South Asian Film
- Books & Literature
- Business & Productivity Software
- Business News
- Business Services
- Camera & Photo Equipment
- Canada
- Casual Games
- Cats
- Celebrities & Entertainment News
- Classical Music

8. Google Ad Settings page showing you interest in Credit Cards: -



The screenshot shows the Google Ad Settings page with a central notification modal for "Credit Cards". The modal contains an illustration of a blue credit card and a grey card reader. Below the illustration, the text reads: "Credit Cards" followed by "Google estimates this interest, based on your activity on Google services (such as Search or YouTube) while you were signed in." and a link "Manage your activity". At the bottom of the modal are two buttons: "Close" and "Turn off".

adssettings.google.com/authenticated?hl=en

Google Ad Settings

Credit Cards

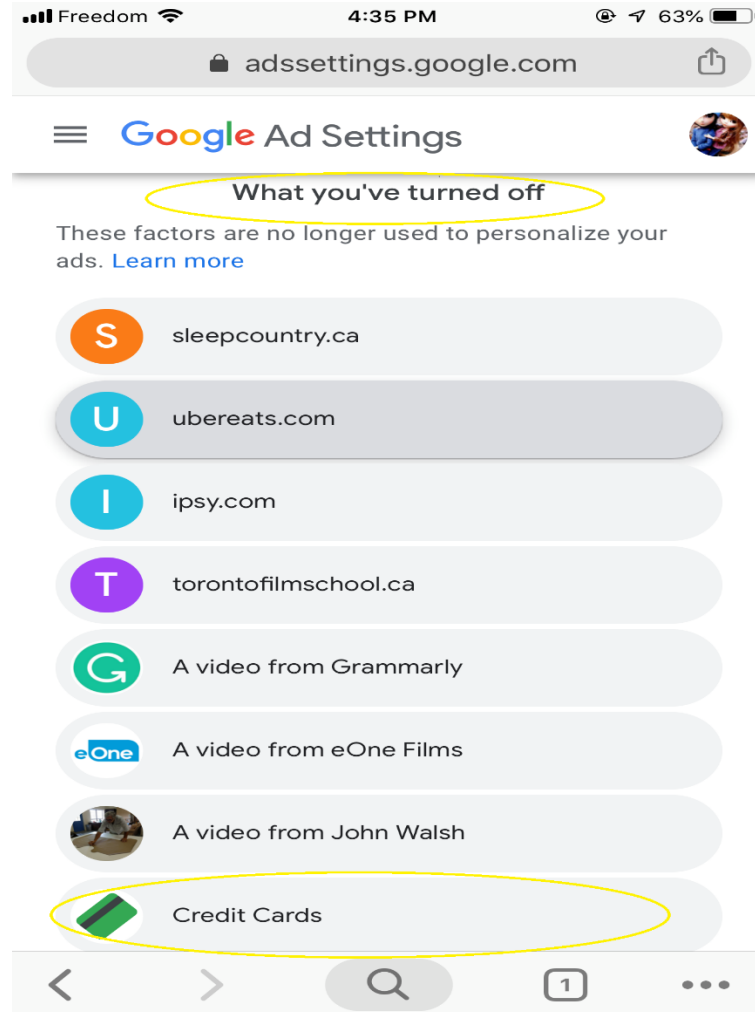
Google estimates this interest, based on your activity on Google services (such as Search or YouTube) while you were signed in.

[Manage your activity](#)

Close Turn off

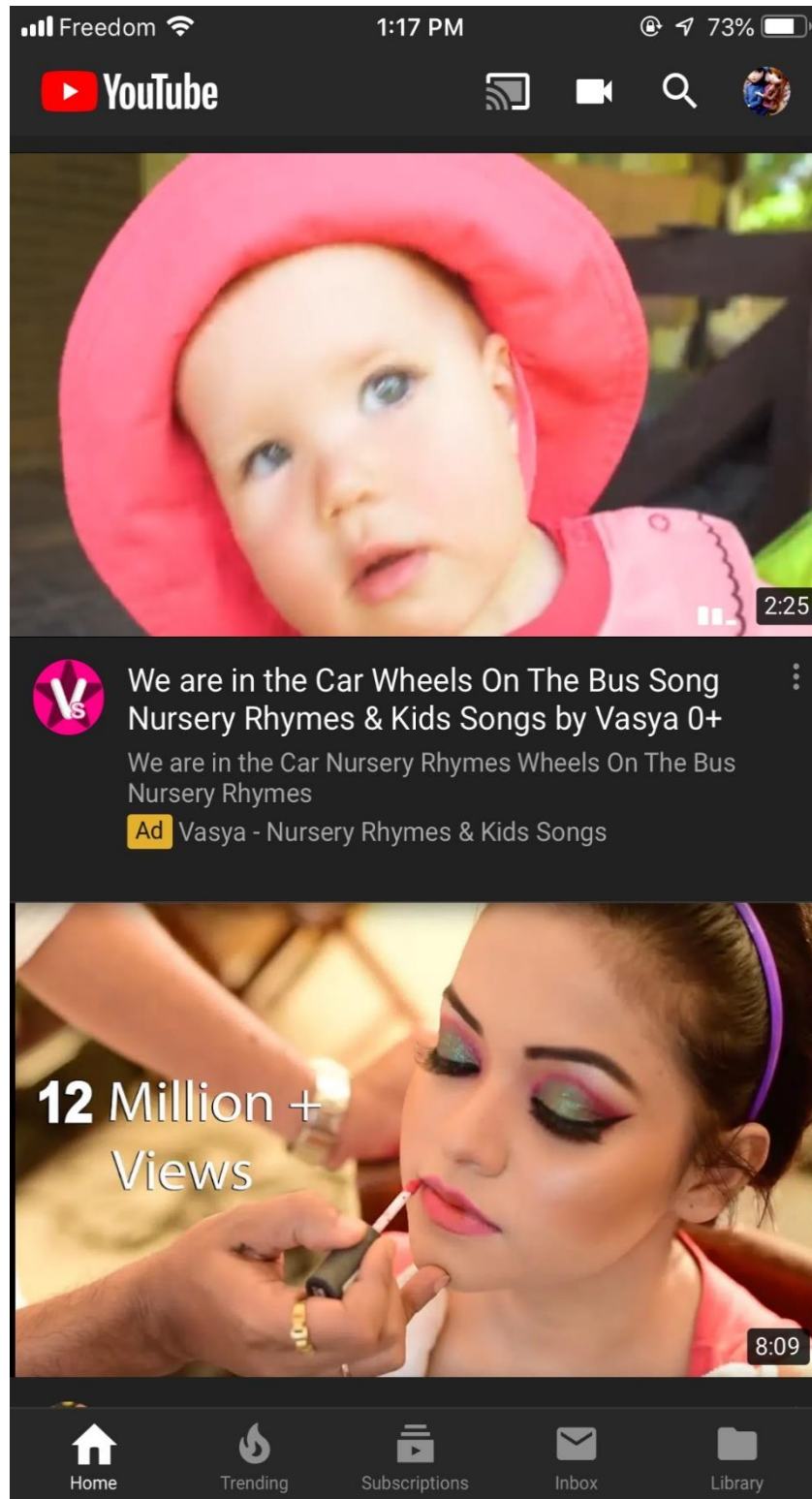
Drama Films Education Status: Bachelor's Degree

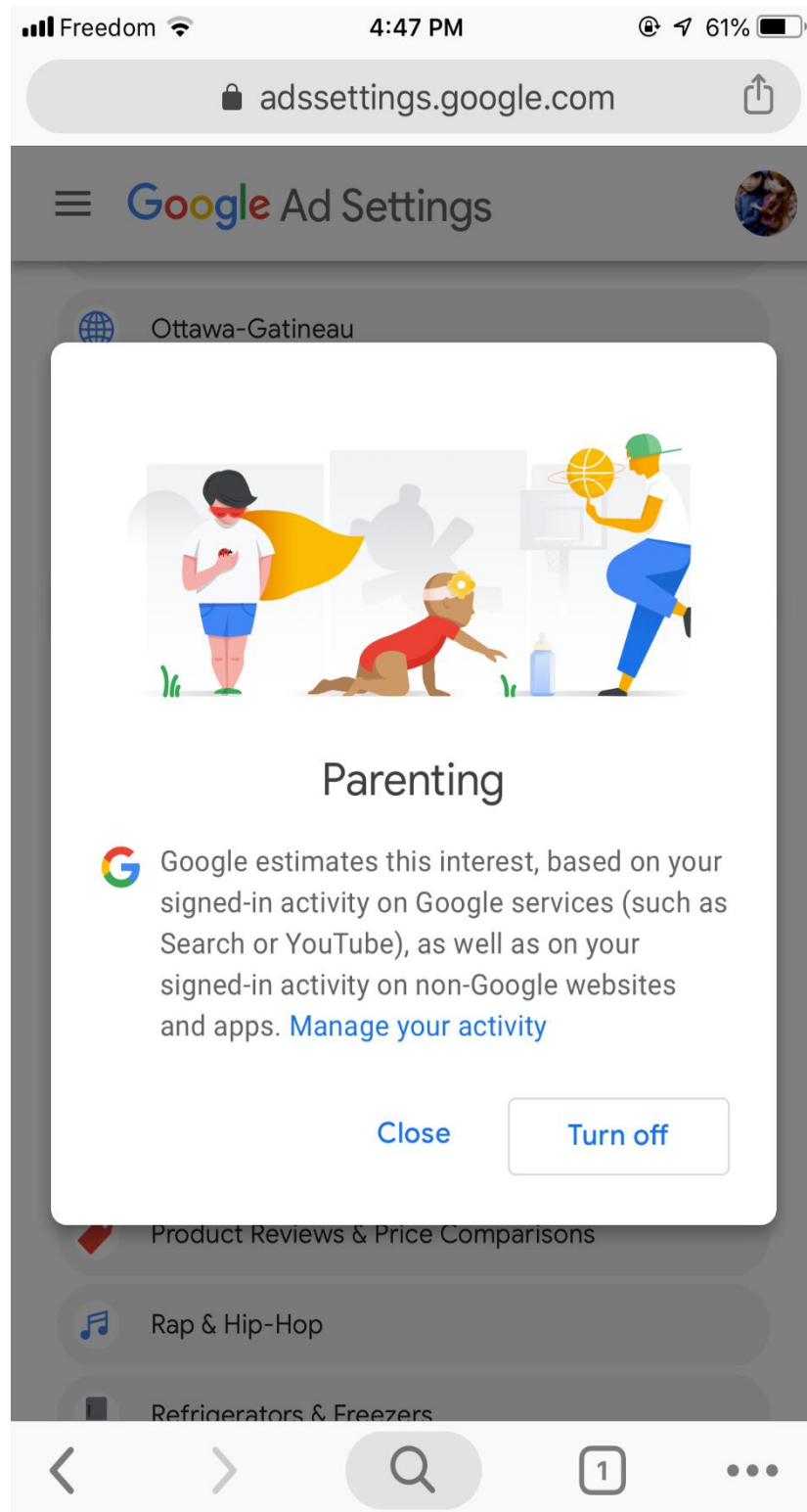
9. If you click on Turn off option seen on the last screenshot, Google Ad setting will remember that you have asked them to turn off ads for credit card. They won't show you ads related to it, but they still have the record that you are not interested in credit cards.



Example 2: -

10. Likewise, if you are shown ads related to child products or parenting: -



11. Google Ad Settings page showing Parenting under your profile: -

The screenshot shows a mobile browser interface with the URL `adssettings.google.com`. The page title is "Google Ad Settings" and the location is "Ottawa-Gatineau". A white modal window is centered on the screen, titled "Parenting". It features an illustration of a child in a superhero costume, a crawling baby, and a man playing basketball. Below the illustration, the text reads: "Google estimates this interest, based on your signed-in activity on Google services (such as Search or YouTube), as well as on your signed-in activity on non-Google websites and apps. [Manage your activity](#)". At the bottom of the modal are two buttons: "Close" and "Turn off".

Freedom 4:47 PM 61%

adssettings.google.com

Google Ad Settings

Ottawa-Gatineau

Parenting

Google estimates this interest, based on your signed-in activity on Google services (such as Search or YouTube), as well as on your signed-in activity on non-Google websites and apps. [Manage your activity](#)

Close Turn off

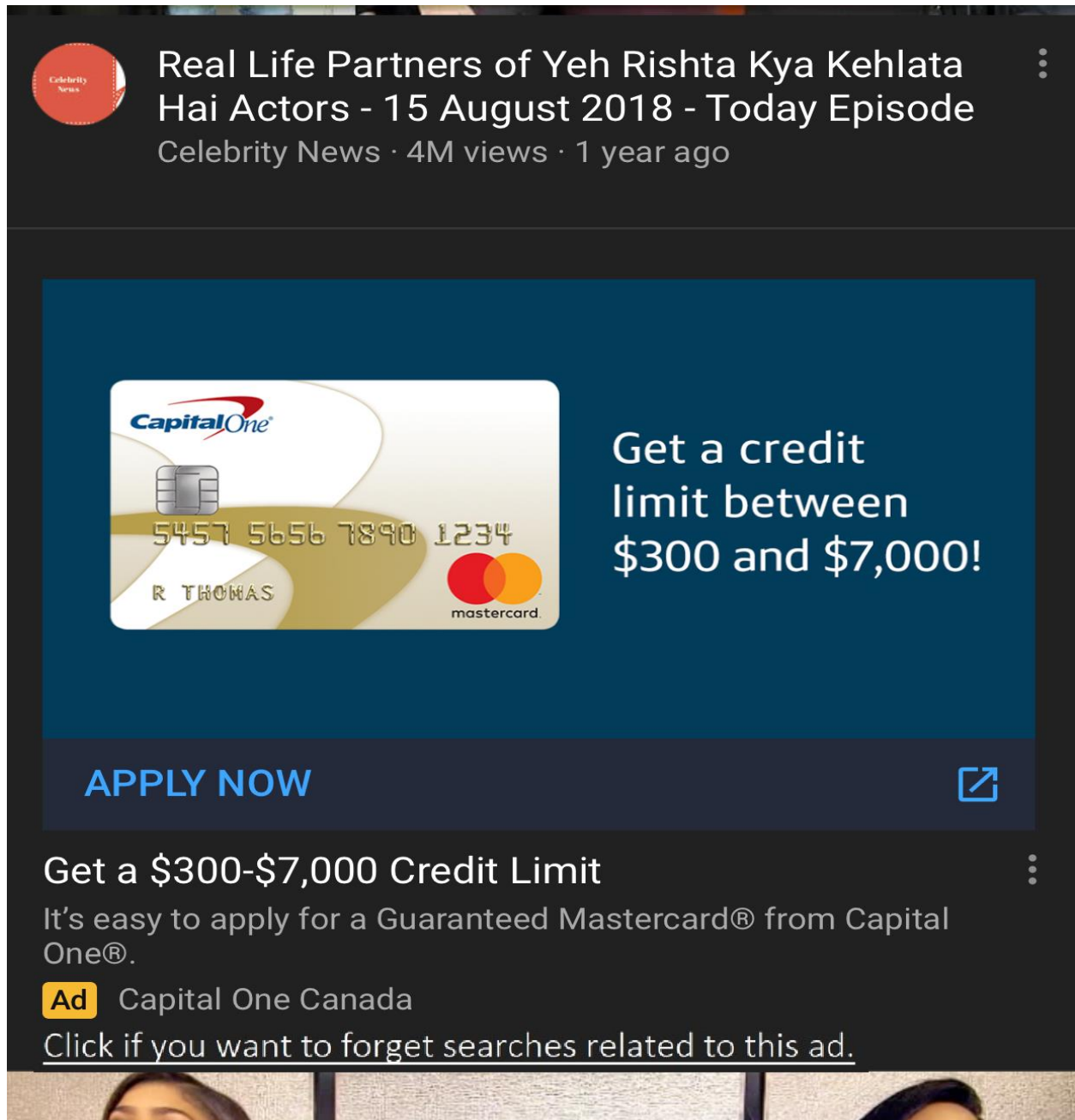
Product Reviews & Price Comparisons

Rap & Hip-Hop

Refrigerators & Freezers

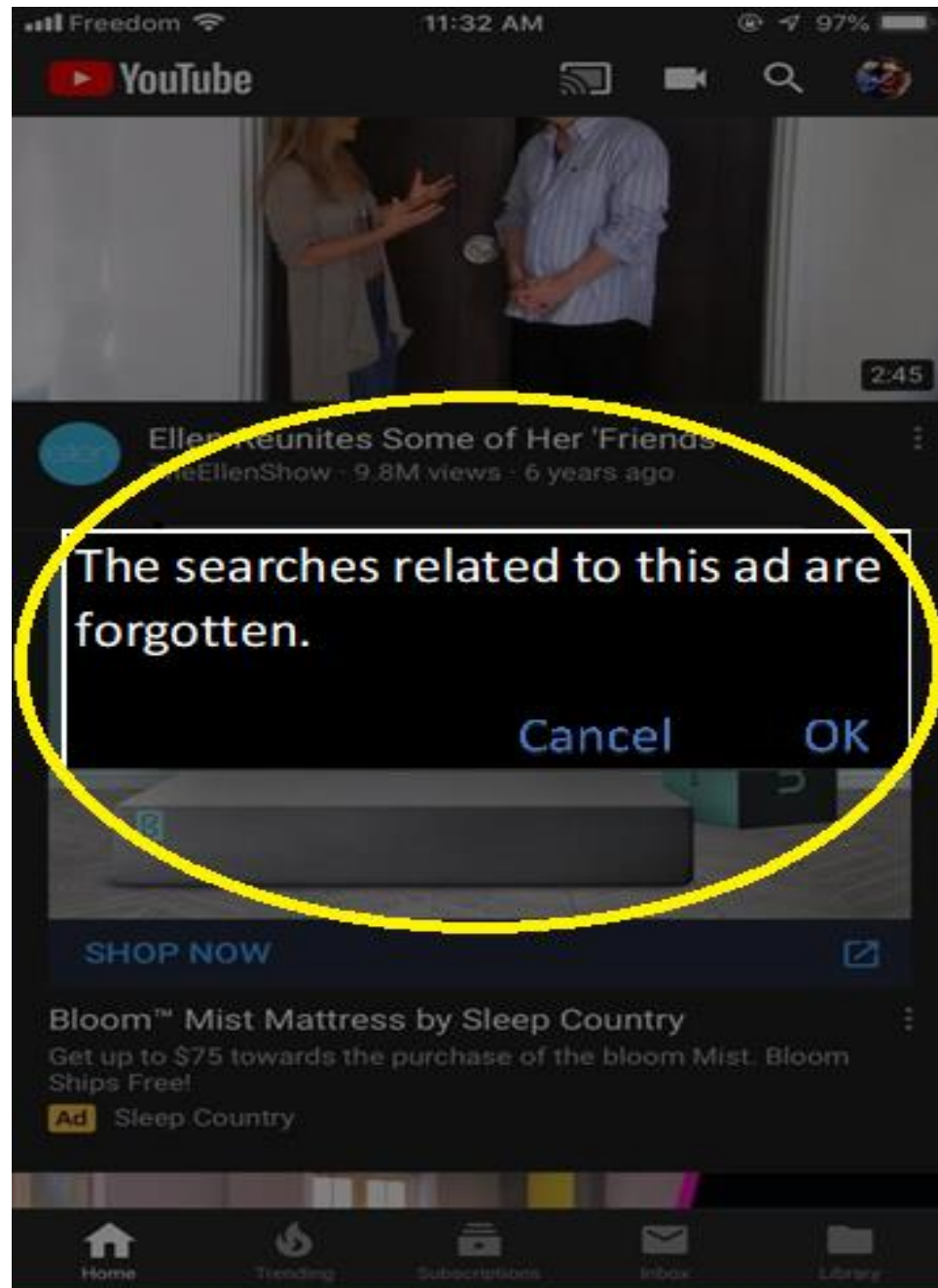
Mock up 2 for Study 2

1. Imagine if you have an option to make ad companies forget all searches related to an ad. In the example below they believe you are potentially interested in new credit cards (say, because you were searching for credit card offers and loans). By clicking on this link : Click if you want to forget searches related to this ad , you are making ad networks forget all the searches stored in your profile related to this credit card ad.

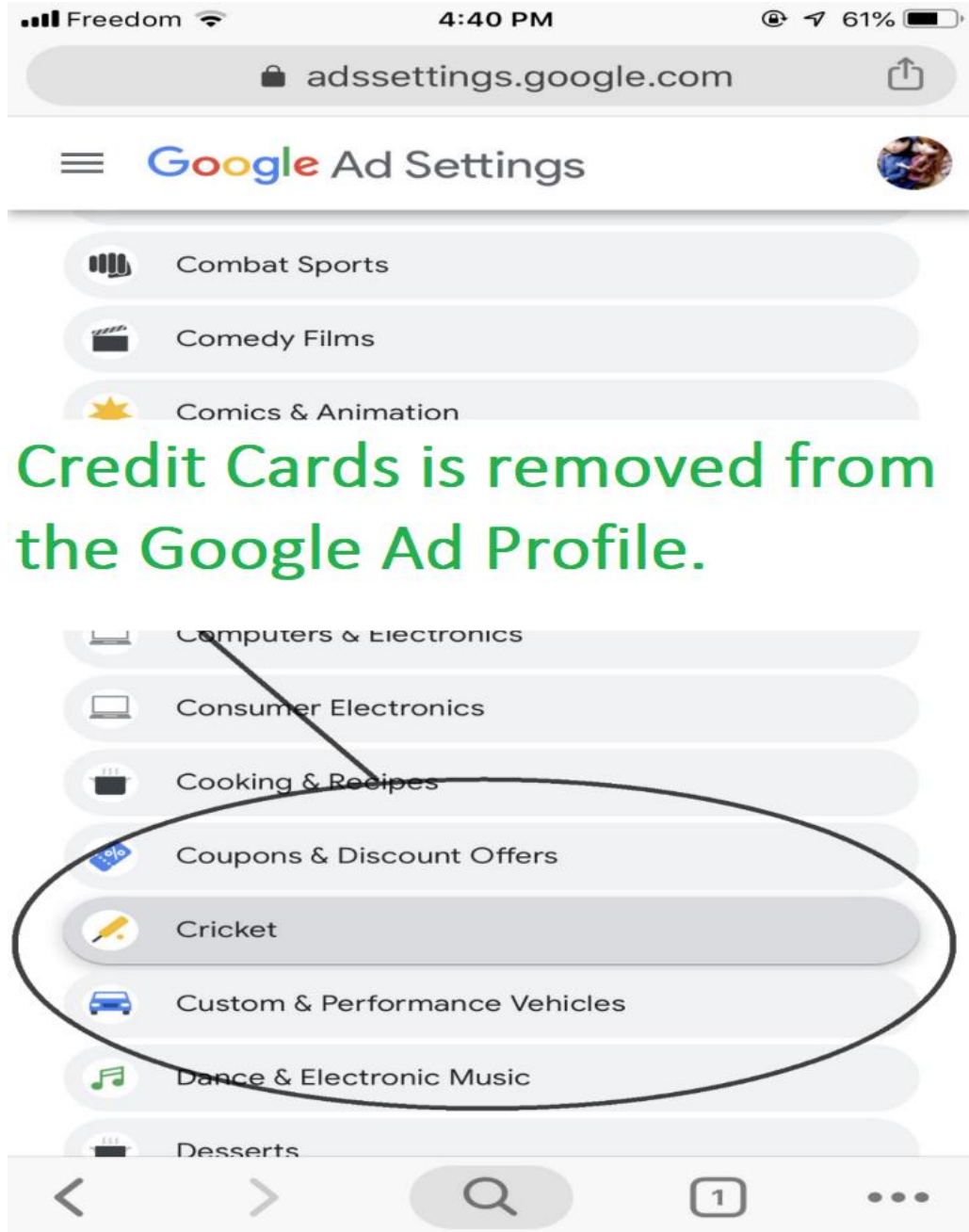


The screenshot shows a YouTube video player interface. At the top, there is a red circular logo with the text "Celebrity News". The video title is "Real Life Partners of Yeh Rishta Kya Kehlata Hai Actors - 15 August 2018 - Today Episode" and it has "4M views · 1 year ago". The main content is a credit card advertisement for Capital One. The ad features a Capital One credit card with the number "5457 5656 7890 1234" and the name "R THOMAS". To the right of the card, the text reads "Get a credit limit between \$300 and \$7,000!". Below the card, there is a blue button that says "APPLY NOW" and a share icon. At the bottom of the ad, there is a text overlay: "Get a \$300-\$7,000 Credit Limit" followed by "It's easy to apply for a Guaranteed Mastercard® from Capital One®." and "Ad Capital One Canada". At the very bottom of the ad, there is a link: Click if you want to forget searches related to this ad.

2. When you click on the link the ad disappears from your YouTube feed: -

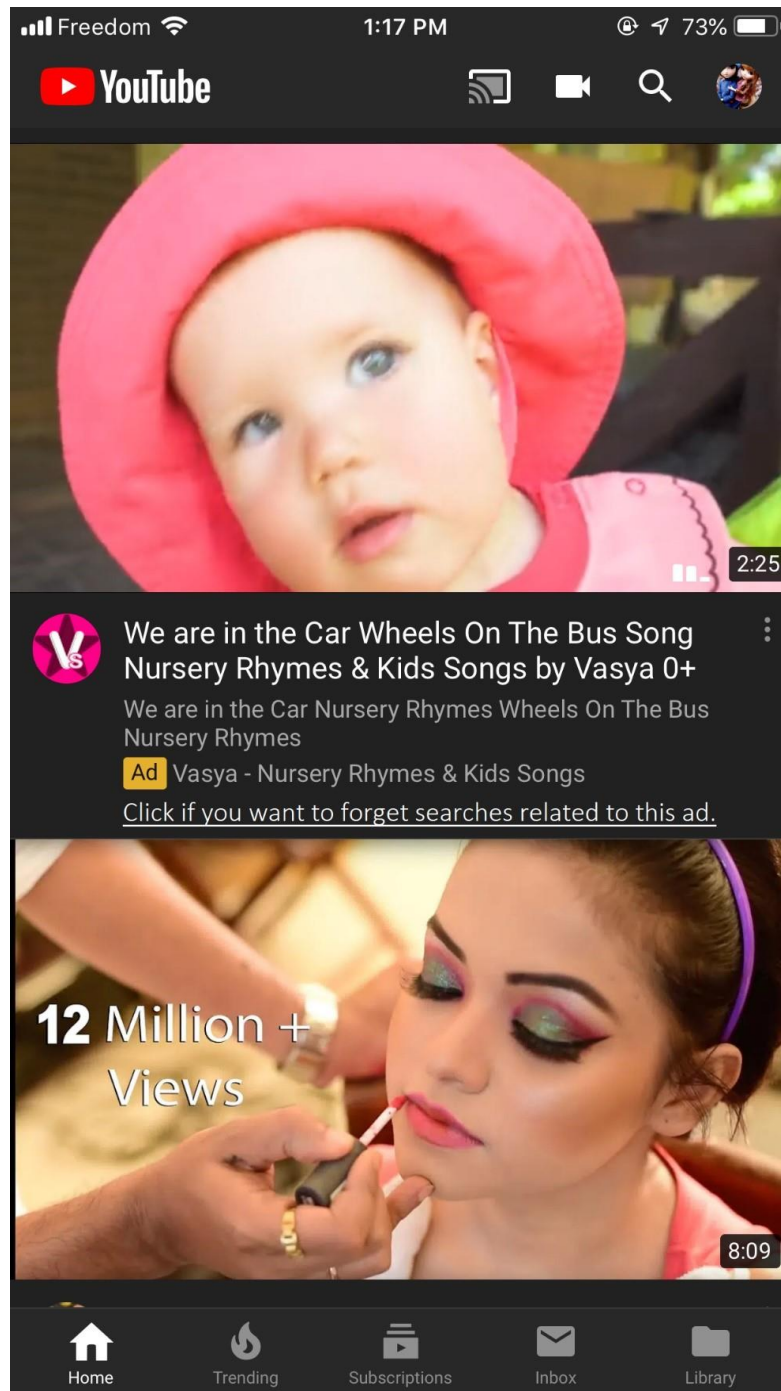


3. Google Ad settings page would show that your interest in Credit Cards is removed from your profile. Unlike the existing system Google won't store the forgotten interest of credit cards in what you've turned off section. That's the difference with forget. It has been removed completely from your ad profile.

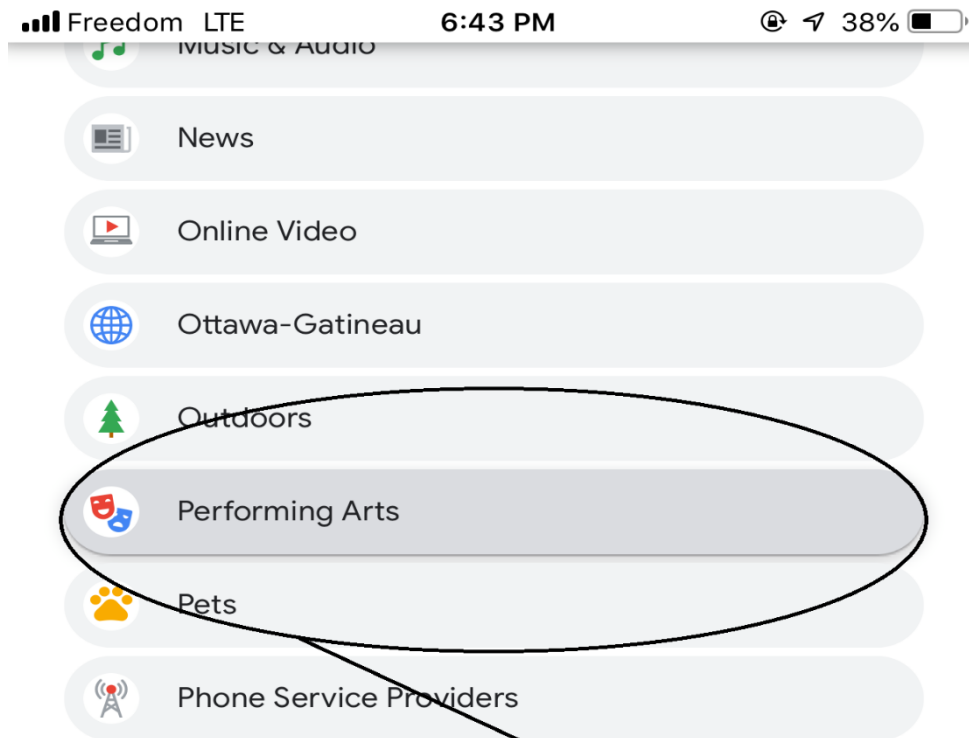


Example 2: -

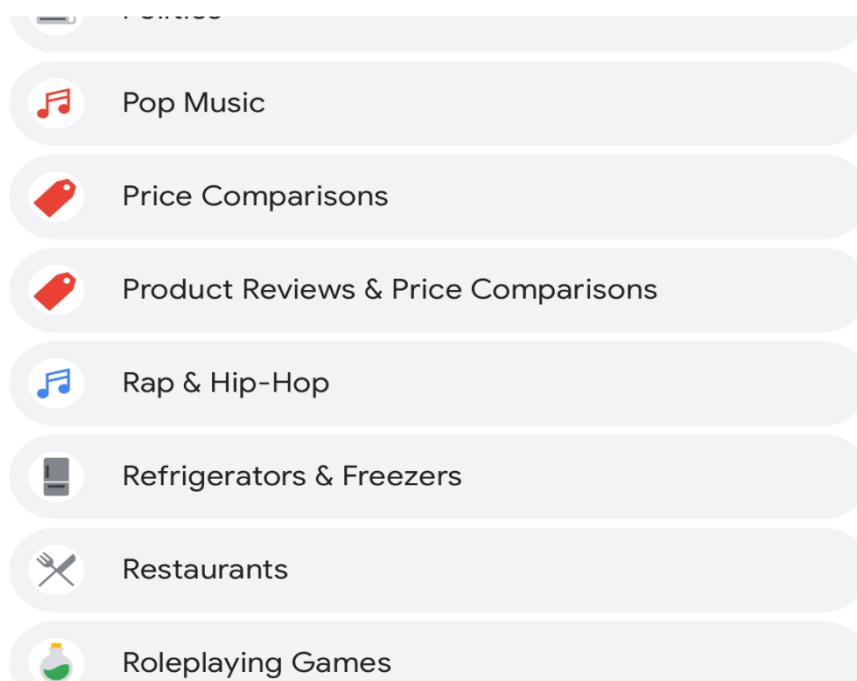
4. When you click on 'Click to forget searches related to this ad' on Nursery ads: -



5. Clicking on Forget link on the ad for Nursery Rhymes and Kids song will remove Parenting from my ad profile.



Parenting is removed from Google Ad interest profile.



Script for the second user study “Users Perceptions of Targeted Advertisements and Online Tracking”

Hello! Greetings. How are you today? So, as you know you are here to participate in a study on Targeted ads and web tracking.

I am going to turn on the audio recorder so everything you say from now will be recorded.

Start audio recording.

- *What is your name?*
- *What is your age?*
- *What is your highest education?*
- *How much time do you spend online?*

Thank you for your responses.

{Removed the information about first round of scale question.}

Let's begin with a small interview. You don't need to write but just explain your answer/ response in your words. It is just like an interview. Before proceeding with the questions, I need to explain you the definition of creepiness in terms of this study.

Creepiness in general is a feeling when you are been watched by a stranger while you are doing something on you own. Ex. Some stranger watching you from your window is when you feel creepy.

- *Do you have any questions or concern before starting the first round?*
- *Hand participant Questionnaire 1 and discuss each question with them.*

I hope you are finding everything ok.

[Removed the information about the video]

We are doing pretty good. So, now we have something interesting for you. I am going to show you a mock-up. Do you know what mock up is? If not, mock-up is a pictorial representation of a model or a structure. In our case these are screenshots/pictures of web application YouTube.

- *Show participant mock-up 1.*

These are the screenshots taken from YouTube on an iPhone. These screenshots were taken when you see a YouTube ad. Few screenshots show the ad information and some of them displays the Google Ad Personalization page where it shows you based on what information they show you specific ads. The Google Ad Settings page displays all you interest stored within Google Ad network. Have a look. If you don't understand anything just let me know.

So, first page is the YouTube home page that is what you see when you open YouTube.

Next page is showing the YouTube ad of credit card. I suppose you must have seen YouTube ads very often. If not, this is how it looks like.

- *Have you ever clicked on any YouTube ad before? Have you seen the ad information? If not, here is how it looks.*

After clicking on the three-dot drop-down menu here what we see.

Further if you click on 'Why this ad?' it will display you the below information.

What happens when you further click on Google's Ad Settings link on the pop-up. Here is what you see.

So, you saw the Google ad personalization page. Now next is all your interest stored within Google Ad network displayed on Google Ad settings page. Have you seen such information before?

So, it has information about your gender, your age range and everything that you have looked so far, or you are interested in.

Moving further if you notice Google Ad settings page have stored credit card as my interest so they showed me a YouTube ad for credit card which we saw earlier in this mock-up.

If you click on the turn-off option on the credit card interest it will remove your interest but as shown in the screenshot it will have this information stored in the 'What you've turned off' section.

We have another similar example of Nursery Rhymes and Kids songs YouTube ad. If you follow the same steps and check your Google Ad Settings page you will find that they have stored parenting as my one of my interest which lead to the nursery rhymes and kids' song ads.

Ok, so that's all we have for mock-up one. Do you have any questions or concern before proceeding to the next round?

[Removed information about scale questions]

Ok. So, now we have another set of interview question to discuss which are related to the screenshots we saw.

- *Hand participant Questionnaire 2 and discuss each question with them.*

Great! Now it is time for a next round of mock-up. These are simulated screenshots that show how YouTube would look with a key change to the ad interface. Can you tell what is different?

- *Show participant Mock-up 2*

Were you able to see what was the difference? If not, we have the mock-up of existing YouTube ad for you to review.

So, this is how YouTube ad would look if you could make ad companies 'forget' information about you.

I need to explain to you the purpose of forget in this context. Let me give you an example. If you are often searching for baldness treatment, then Google is going to show ads related to baldness treatments. If you are watching YouTube with your friends, it is possible for such an ad to be displayed, potentially telling your friends about your hair loss, potentially embarrassing you. To help avoid such situations, we propose forget as a possible solution. When you see ads which you are not comfortable with, you can select forget. Forget will instruct the ad network to forget the information that led to the selection of that specific ad. As a result, you won't be shown any ads related to it in future.

So, in mock-up 1 we saw ads of credit card and nursery rhymes and when we checked Google Ad Setting, we found that they have stored credit card and parenting as our interest which caused them to display these ads to us. But what if I don't want these data to be stored within the Google Ad network?

In the previous screenshots you saw that even after selecting the turn off option for credit card, Google ad network store this information in 'what

you've turned off so that the existing Google ad network can remember that you have asked it to forget about credit cards or not show ads related to credit cards.

Unlike that, forget mechanism proposed in the mock-up 2 will completely remove the data from your ad profile. It won't store that information in 'what you've turned off'. The 'Click to forget searches related to this ad' will allow user to remove the data behind creepy targeted ads. It will provide us more control over what data should be stored within the ad network.

Finally, we have managed to reach to the last segment of our study. So, we are just left with one interview. So, here we go.

- Hand participant Questionnaire 3 and discuss each question with them.*

Thank you for participating in this study! Here is your gift card.