

Internet Privacy Implications

A Major Qualifying Project

Worcester Polytechnic Institute

Submitted to the Faculty of the Worcester Polytechnic
Institute in partial fulfillment of the requirements for
the Degree of Bachelor of Science.



WPI

Submitted By:

Goncharova, Masha

Harnois, Jeffrey

Advisors:

Wills, Craig PhD

Doyle, James PhD

May 6, 2021

Abstract

Our research focused on understanding the effectiveness of browsers, extensions, mobile applications, and search engines and their protection of user privacy. We ran test cases on the top 100 Alexa sites, using a Fiddler proxy to capture traffic, with certain configurations of tools mentioned to see which ones were efficient in blocking user tracking technologies. We found that Brave and Firefox in Strict mode are the best browsers in terms of tradeoff between percent of websites with degradation versus percent trackers remaining. uBlock Origin, Ghostery and Privacy Badger are the best browser extensions in terms of the same tradeoff.

Based on our results, we created a recommendation system using a survey approach. We suggest a combination of tools that are personalized to users based on their reported privacy preferences and desire to switch their current browsing setup. In order to better understand users' views on privacy, we additionally showed participants their own data Google has synthesized about them to evaluate if that would change their responses. A ceiling effect appeared in our responses, indicating that no matter the condition, all our participants indicated a willingness to switch to the tools that we were recommending.

Table of Contents

Abstract	1
Table of Contents	2
List of Tables	6
List of Figures	7
Double Major Note	9
1. Introduction	10
2. Background	12
2.1 Related Work	12
2.2 User Concerns Associated with Internet Tracking	13
2.2.1 Privacy in the Media	14
2.2.2 Human Profiling	15
2.2.3 Dynamic Pricing	15
2.2.4 Filter Bubbles	16
2.3 Tracking Techniques	17
2.3.1 Location and Location-Based Advertising	17
2.3.2 Cookies	17
2.3.2.1 First-Party Cookies vs. Third-party Cookies	18
2.3.2.2 Single Session Cookies	18
2.3.2.3 Persistent Cookies and Super Cookies	18
2.3.2.4 Cookie Sharing and Cookie Syncing	19
2.3.3 Canvas Fingerprinting	20
2.4 Third-party Domain Classification	20
2.5 Most Popular User Privacy Technologies	21
2.6 Performance and Browser Extensions	21
2.7 Summary	21
3. Research Questions	22
3.1 Privacy Protection of Browsers	22
3.2 Privacy Protection of Browser Extensions	22
3.3 Privacy Tools vs. Website Degradation	23
3.4 Privacy Protections of Mobile Applications	23
3.5 Privacy Protections of Search Engines	24
3.6 User Privacy Concerns and Survey	24
3.7 Summary	25
4. Methodology	26
4.1 Browsers	26

4.1.1 List of Browsers Tested	26
4.1.2 Evaluation	29
4.1.2.1 Logging Traffic	29
4.1.2.2 Driving Browsers	30
4.1.2.3 Categorizing Domains	31
4.2 Extensions	35
4.2.1 List of Extensions Tested	35
4.2.2 Evaluation	42
4.3 Website Degradation	42
4.3.1 Evaluation	42
4.4 Summary	43
5. Results for Browsers and Extensions Privacy Protections	44
5.1 Counting Third Parties Per First-party Site	44
5.2 Browser Results	46
5.3 Extensions Results	50
5.4 Website Degradation Results	55
5.5 Website Degradation vs. Privacy Improvement	57
5.6 Summary	58
6. Mobile Applications Privacy Protections	59
6.1 Popular Mobile Privacy Applications	59
6.2 Methodology	60
6.2.1 Fiddler Proxy on Mobile	61
6.2.2 Certificate Pinning	61
6.2.2.1 Editing APK Code	62
6.2.2.2 Using an Android Emulator on Desktop	63
6.2.2.3 Frida Hook and Objection	63
6.2.2.4 Root Android Device to Bypass SSL Pinning	63
6.2.2.5 Edit Browser Configuration File	65
6.2.2.6 End Result	65
6.3 Mobile Results	66
6.4 Summary	68
7. Search Engines	69
7.1 List of Search Engines Tested	69
7.2 Methodology	72
7.3 Search Engine Results	76
7.4 Summary	79
8. Recommendation Survey	80

8.1 Research Questions	80
8.1.1 Switching to Privacy Tools	80
8.1.2 Lack of Control vs. Google Synthesization	80
8.1.3 Lack of Control vs. Seeing Invasive Ads	81
8.1.4 Societal Issue vs. Seeing More Invasive	81
8.2 Methodology	81
8.2.1 Participants	81
8.2.2 Conducting Survey Research	82
8.2.3 Creation of Questions	82
8.2.4 Experimental and Control Groups	83
8.2.5 Survey Design	85
8.2.6 IRB Approval	86
8.2.7 Revised Survey	87
8.2.8 Survey Logic	88
8.2.9 Survey Testing	90
8.3 Summary	90
9. Survey Results	91
9.1 Experimental Survey Results	91
9.1.1 Switching to Privacy Tools Per Condition	92
9.1.2 Level of Control vs. Google Accuracy	92
9.1.3 Level of Control vs. Presence of Invasive Ads	93
9.1.4 Invasive Ads vs. Societal Concerns	93
9.2 Revised Survey Results	93
9.2.1 Level of Control vs. Presence of Invasive Ads	94
9.2.2 Invasive Ads vs. Societal Concern	95
9.2.3 Possible Response Bias	95
9.3 Discussion and Limitations	95
9.4 Summary	96
10. Conclusion	97
11. Future Work	99
References	102
Appendix A: Examples of Minor and Major Website Degradation	107
Appendix B: List of Survey Questions	111
Appendix C: IRB Approval & Training Certificate	153

List of Tables

Table 4.1. List of Browsers tested and their features.	p. 25
Table 4.2. List of browser extensions tested during the project and their features.	p.36
Table 6.1. List of the popular mobile applications tested in the project	p.59
Table 7.1. List of Search Engines Tested.	p.69
Table 7.2. Induced Behavioral Interests	p.73
Table 7.3: Induced Sensitive Interests.	p.74
Table 7.4: A setup of interests generated in each search engine.	p.75
Table 8.1: List of Survey Results.	p.88

List of Figures

Figure 4.1 Python code developed to split the hostname and extract just the domain name.	p.30
Figure 4.2. Screenshot of the Ghostery extension on Chrome, when visiting cnn.com, showing Bounce Exchange and detected tracker URL.	p.31
Figure 4.3. Screenshot showing Ghostery Global Blocking List and the categories inside.	p.32
Figure 4.4. Screenshot showing Ghostery's Blocking list, specifically Advertising category.	p.32
Figure 4.5. Screenshot showing Chrome code inspector with a snippet of code from the Ghostery extensions.	p.33
Figure 4.6. Line of code responsible for capturing screenshots during web testing.	p.42
Figure 4.7. Code used to calculate the average number of third parties per first party.	p.44
Figure 4.8. Code for the helper function used.	p.45
Figure 5.1. Average Number of Third Parties per Visited Site for Each Browser	p.46
Figure 5.2. Average Number of Third Parties per Visited Site for Each Browser by Category.	p.47
Figure 5.3. Percent Trackers Remaining for Each Browser	p.48
Figure 5.4. Percent Trackers Remaining vs. Browser with Privacy Features Configured.	p.49
Figure 5.5. Average Number of Third Parties per Visited Site for Each Extension	p.50
Figure 5.6. The average number of third parties per visited site for Each Chrome Extension by Category.	p.51
Figure 5.7. The average number of third parties per visited site for Each Firefox Extension by Category.	p.52
Figure 5.8. Percent Trackers Remaining for Each Extensions	p.53

Figure 5.9. Percent Trackers Remaining Average on Chrome and Firefox.	p.54
Figure 5.10. Percent of Websites with Degradation for Each Browser by Severity.	p.55
Figure 5.11. Percent of Websites with Degradation for Each Extension by Severity.	p.56
Figure 5.12 Percent Trackers Remaining vs. Percent Websites with Degradation	p.57
Figure 6.1. Code change in the header of the AndroidManifest.xml file to lower the APK and API version.	p.61
Figure 6.2 An example of code that needs can be added to define user added certificates.	p.62
Figure 6.3. Average Number of Third-Parties Per first-party site for Each Mobile App.	p.65
Figure 6.4. Average Number of Third Parties per first-party site for Each Mobile App by Category.	p.66
Figure 6.5. Percent Trackers Remaining for Each Mobile App.	p.67
Figure 7.1. Total Number of Advertisements related to search terms for Each Search engine.	p.76
Figure 7.2. Total Number of Advertisements related to search terms for Each Search engine by Induced Interest.	p.77
Figure 7.3. Induced Interest Term vs. the Total Number of Advertisements Shown Related to the Induced Interest.	p.78
Figure 8.1. Snippet of flowchart containing answers for our survey recommendation questions	p.82
Figure 8.2. A manipulated list of the interests of “the common WPI student” for the control group of our experimental survey	p.84

Double Major Note

This MQP project was also submitted as partial fulfillment of the requirements for the Degree of Bachelor of Science in Psychological Science. Chapters 8 & 9, and the supplementing work that was carried out described in those chapters, were completed by Jeffrey Harnois to fulfill these additional requirements. The rest of the entirety of this paper was a joint effort by both authors for the finalization of the Computer Science degrees.

1. Introduction

There have been many studies that show that users are growing increasingly more concerned about their data on the Internet and how it is being used. Many lack the fundamental understanding of the technologies being used and actionable steps they can take to protect themselves. Transcend - a company that focuses on user privacy and giving users control over their personal data - released a study in 2020 titled the “The Data Privacy Feedback Loop” that has solidified the idea and proved with tangible data the fact that users are more concerned about their privacy than before (*The Data Privacy Feedback Loop*, 2020).

The main method used by the majority of websites to track users are through the use of cookies, although there are other more complex methods that have gained popularity in recent years. Cookies by themselves are not inherently bad and were designed to improve user experience. Functional benefits of cookies include saving location data to suggest relevant weather and map information, saving their login information to reduce the number of times they need to login in, provide cross social platform sharing and offer more relevant information in general. With the increase in the overall tracking technology, rise and growth of analytics companies and lack of user awareness regarding their privacy, there are a lot of concerns about the future use of these technologies.

This research in this paper is not intended to portray all cookies and advertisements to be bad. At the time of conducting this research, there is unprecedented growth of companies and individuals that rely on the Internet to make their living. With a mass availability of content on the Internet, users have grown to expect high quality content at their fingertips for free and many are not willing to pay subscription fees to access content behind a paywall. News sites, blogs, Internet services (including Google Maps, Gmail, Youtube), independent Youtube and video creators, and small business owners turn to advertisements on the Internet to either make profit directly from advertisers or use smart analytics technologies to attract more users to their service or shop.

There are different business models and different ethics standards employed by companies when it comes to showing ads or gathering information about users. Ethical advertisements that do not play on user’s emotions and do not gather sensitive data are a great way to continue making free, high-quality content on the Internet while allowing creators to benefit financially. Advertisements that do not track users across platforms and have no insight into the user's emotional state are thought to be less effective driving product and service companies to more deceptive and tricky ways of capturing the interest of a consumer (Reczek et al., 2016). Recent research has demonstrated that since users are growing increasingly more concerned about their privacy, they favor privacy-oriented companies that are transparent about their policies and are willing to switch from the current companies that they purchase from in order to prioritize data security (*Data Privacy Feedback Loop*, 2020).

The following research talks about tools such as browsers and extensions that users can employ to protect their privacy and companies that follow alternative advertising practices. For example, DuckDuckGo search engine which, unlike Google, shows users ads only based on what they have typed into a search bar and anything previously searched for will not follow the user around (Evangelho, 2018).

The goal of this work is to build on the previous research and findings by Wills, and Uzunoglu from 2016 and Mihajlo Zeljkovic from 2010, to see how the landscape of user privacy on the Internet has changed, as well as what new tools were developed to protect the user's privacy and how effective they are. We aim to provide a recommendation to the users that consists of a tool or a combination of several tools that are intuitive to use and set up, are free, do not break the main website features and prevent sites from gathering information about people in ways depending on their privacy concerns determined by a survey we develop.

The following paragraph presents a roadmap for the rest of the work. Chapter 2 serves as the background for our work, examining in detail user concerns associated with privacy violations, and explores various tracking mechanisms used on the web. Chapter 3 contains the research questions for the work and reasons behind the importance of those questions. Chapter 4 presents a methodology to answer the first three research questions regarding browsers, extensions and website degradation and Chapter 5 presents the results for those questions. Chapter 6 and Chapter 7 present the methodology and answers the fourth and fifth research question regarding mobile applications and search engines respectively. Chapter 8 details the design process for the survey used to answer the last research question and Chapter 9 analyzes the survey results. Finally, Chapter 10 makes suggestions regarding future work and Chapter 11 serves as the conclusion for the project.

2. Background

Before we could study Internet privacy and its implications on the users, we had to perform some background research to establish what similar work has been done before and what conclusions have the researchers reached, what the top user concerns associated with privacy invasions on the web and which specific mechanisms can be used in order to track users.

2.1 Related Work

The following section highlights six papers that were particularly important to the research we were conducting and influenced our research questions and methodology.

1. “Understanding ADBlockers” (Uzunoglu, 2016). The project evaluated ad blockers to give more information about existing ad blocking tools, how they operate and how good they are. The project focused on analyzing the effectiveness of existing ad blockers against different types of third-party content on popular websites.
2. ”What ADBlockers Are (and Are Not) Doing” (Wills & Uzunoglu, 2016). Same project as above, but a more concise six page version.
3. “Privacy Awareness of Web Users” (Zeljko, 2010). The work examined web browsing habits of Internet users and their level of awareness regarding their privacy on the Web. The researchers created a website that provided information about common tracking mechanisms and showed them based on their individual browsing history what trackers were embedded on the sites they visited.
4. “Block Me If You Can” (Merzdovnik, 2017). The work studied the effectiveness of third-party trackers on a large scale by analyzing 100,000 popular websites. Their work established that rule-based (or list based) browser extensions outperform learning-based ones.
5. “Comparison of Web Privacy Protection Techniques” (Mazel, 2019). The work proposed a methodology to compare privacy protection techniques when crawling many websites, assessed webpage quality degradation and examined different types of extensions based on their blocking technique (i.e. heuristics, list, other). They also performed manual analysis because they found that automated analysis sometimes fails to run or identify the metrics they were looking for. The group found that Ghostery and uBlock Origin provide the best trade off between protection and webpage quality, but Ghostery had to be configured, since out-of-the-box setup did not provide optimal protection.
6. “Understanding What They Do with What They Know” (Wills & Tatar, 2012). The work examined what web advertisers know about the users and what they can do with the information available to them. The researchers analyzed the ads shown to them during their controlled browsing tests and developed a set of induced interests for sensitive and non-sensitive interests. They visited a selection of sites that were equivalent to the

induced interest and analyzed the type of advertisements that they saw in a controlled browsing environment.

2.2 User Concerns Associated with Internet Tracking

Internet privacy is characterized as the level at which one is able to maintain their autonomy while interacting with other end users or hosts (H. Wang, Lee, & C. Wang, 1998). The main goal of cybersecurity is to provide the element of the CIA triad, confidentiality, integrity, and availability as protections to users. Privacy on the Internet plays a crucial part in being able to protect said users. If one's privacy is compromised, it might jeopardize the triad and the aim of cybersecurity (Harris & Maymi, 2018).

Privacy concerns on the Internet come in many different forms. Stakeholders in every aspect of the Internet must be aware of their own privacy to maintain their own protections. Stakeholders include basic end users, companies and large businesses, and governmental agencies. Over the years, concerns over privacy have been increasing as more news stories continue to be published, exposing the public to possible situations that violate user privacy.

The act of invading one's privacy has been broken down into two different categories, static and dynamic private information. The difference between the two, as the names suggest, refer to levels at which the information changes. A social security number, for example, would remain static while a personal identification number to someone's debit card is dynamic and can change over time. The disclosure of any sort of personal or organizational information such as this is considered a breach in privacy (Wang, Lee, & Wang, 1998).

A study conducted in 2007 found that people tended to have a high level of fear in disclosing their information to the sites that they visit. 84% reported that they would like laws to be passed that would regulate what sites were able to capture in the context of personal information on their users (Turow & Hennessy, 2007). Another survey conducted in 2020 yielded similar results. 93% of all Americans polled claimed that they would switch to a company that advertised better privacy protections. 70% claimed that if they were able to see what companies knew about them, that they would actually trust the company more (*The Data Privacy Feedback Loop*, 2020).

It seems that there is much concern with what sites actually know about us and how much they are willing to protect our personal information. The attitudes towards Internet privacy have not changed over the past decade. Privacy concerns are just as prevalent now as they used to be in 2007. Events continue to surface that expose companies for being too with their ad tracking.

2.2.1 Privacy in the Media

Recent developments that have caught the public's attention include advertisements aimed at specific users on the Internet. These include targeting their emotional states, attempting to influence voting decisions, and creating filter bubbles. Targeted advertising has long been a topic of debate. It is unclear whether or not this type of online advertisement is inherently good or bad. What counts as manipulation versus suggestion needs to get decided first. The following note of the common areas of concern as well provide suggestions of tools for those users that share the same concerns.

One big area of concern has been targeted advertisements to those that are in a vulnerable emotional state, especially the teenage audiences. In 2017, a report stated that Facebook could monitor teenage activity on the platform (Susser, et al., 2018) to target them with ads at the moments when they are the most vulnerable (feel stressed, insecure, anxious). These advertisements can be dangerous for those struggling with weight loss, diabetes, and cancer which could have lasting negative, psychological effects on these teen-aged users.

The influence of advertisements on the results of elections in countries such as the United States, the United Kingdom, Germany, and France is another concern as well. In March of 2018, Cambridge Analytica's impact on the presidential election resulted in a large outcry from the public. By 2020, Cambridge Analytica became synonymous with the worst that technology has to offer. It became the most popular example that people use when they speak about the negative consequences of user tracking technologies. Alexander Nix, CEO of the Cambridge Analytica, stated that his company is able to personalize messages to users whether via regular mail or media ads by using data points across social platforms to target audiences (Susser, et al., 2018).

Laws and regulations have been passed as explained by Miyazaki and Fernandez which focus on certain aspects of privacy. These include the Children's Online Privacy Protection Act of 1998 which aims to protect the privacy of children under thirteen years of age and the E-Privacy Act and Secure Public Networks Act, which aid in protecting E-transactions (Miyazaki, et al., 2000). The issue with creating laws and regulation, which many Americans seem to be in support of, is that there is a major lack of resources in law enforcement to uphold them (Wang, Lee, & Wang, 1998). Given that violations happen to come from foreign adversaries and as Wang and others explained, it would be hard to enforce punishment without the creation of international agreements.

The lack of such agreements leaves the burden on the user to be aware of their actions online and take steps that they need to in order to keep themselves safe. Data Privacy Feedback Loops synthesized, 59% of Americans are not educated in the realm of privacy to their own level of satisfaction. In fact, 88% of them reported that they are frustrated that they are not able to get access to what kinds of information that companies have on them. The levels of knowledge, education, and awareness of the typical person is not as high as they should be for these users to effectively achieve strong privacy protections while online (*Data Privacy Feedback Loop*, 2020).

2.2.2 Human Profiling

Advertisement companies build detailed profiles on their users based on their site traffic. A user might trigger a cookie to record specific information just by searching for specific items like in 2012 when Target detected that someone's searched items indicated them being pregnant before other family members knew (Hill, 2016). These profiles have algorithms run on them in an attempt to increase the effectiveness of their targeted ads. Multiple studies have recreated this phenomenon.

For many years now, human profiling has been studied to evaluate people's perceptions of its outcomes. A study in 2020 was conducted that showed how users would respond to 'hyper-personalized' ads. Researchers gathered information from users using online tracking and data brokers to form personalized advertisements for participants. When asked, about half of all users report to have negative responses from observing such ads (Hanson et al., 2020). In 2019, the study had participants download a browser extension that would collect private information on them as they went on browsing the internet as they would in their everyday life. The extension gathered information based on the users interactions with other sites. With the information gathered, they were able to make clear conclusions about interests and behaviors of their participants. Sites such as Facebook and Google were able to gather inferences on their hobbies and interests as well as observe what privacy-protective measures participants might be taking. Interaction with their sites, other sites, and other tools allow these companies to gather statistics on their users (Bashir, et al., 2019). These studies show how easy it is for sites and tools to gather much information about us.

Our privacy is being invaded in many ways. Sites have even begun to make monetary incentives to others for trading information they have synthesized on their users in order to increase their ad targeting effectiveness. It is the general consensus that after being presented with ads tailored to their characteristics, users wish there to be stronger anonymity while browsing (Bashir, et al., 2019). Many tools and companies make bold claims that their business models revolve around protecting the privacy of their users. This might not be the case behind the scenes.

2.2.3 Dynamic Pricing

One of the most deceptive tactics that the retailers use is altering online prices based on the location. There have been several companies over the years that received serious criticism for their practice of dynamic pricing based on the user's location, operating system, profile or device.

Prices change all the time based on the store's physical location which is understandable since there are serious considerations such as cost of rent that go into pricing. Most of the users assume that when they are shopping online, everyone is getting the same prices (Turow et. al, 2005). As it turns out that's far from the truth. In an attempt to increase profit, companies employ a practice of dynamic pricing to price a product online based on the amount they think a user will

be willing to pay. In 2005, University of Pennsylvania conducted a study on dynamic pricing and found that consumers are deeply uncomfortable with the practice of dynamic pricing and the idea that someone else might be paying less for the same product online. Notably, the study found that “64% of American adults who have used the Internet recently do not know it is legal for an online store to charge different people different prices at the same time of day” (Turow et. al, 2005).

In 2000, Amazon tested dynamic pricing by offering different prices to different customers and almost instantly received angry comments from their customers and dropped the practice. Wall Street Journal identified several retailers including Staples, Rosetta Stone, and Home Depot that were constantly adjusting their prices based on a range of characteristics they were able to discover about a consumer (Klosowki, 2013). In 2012, Orbitz received intense criticism after their consumers found that the site tends to charge as much as 30% more for users that book hotels from Mac computers (Mattioli, 2012). In 2019, an investigation drew customer’s attention to Target when it was uncovered that Target changes prices on certain items when the customer is inside or outside the store. The cost of the TV, vacuums and car seats changed for the customer depending on whether they were in the parking lot of the store versus their home (Hrapsky, 2019). In 2020, DuckDuckGo newsletter noted that Target changes prices based on the local market data for online shoppers as well. Target charges the customers based on their set preferred location, so consumers in New York will pay more than a consumer in Minnesota, even though they are both shopping online and the price increase has nothing to do with how much it costs to deliver something to the customer.

There are a few techniques available to help the customers such as changing their IP address, such as by using a proxy server or a virtual private network (VPN). Additionally, users can browse in incognito or private mode and disable or block third-party cookies. Another option for users is using one browser for shopping that does not store any data and a separate browser for everyday activities, however that method is not foolproof or complete.

2.2.4 Filter Bubbles

The Cambridge Analytica case is an example of more direct manipulation of citizens to influence their vote in the elections. There are also concerns about the rise of filter bubbles that have garnered more attention over the years. In March of 2011, Eli Pariser gave a TED talk regarding online “filter bubbles” that are created on the Internet, especially social media platforms, driven by the algorithms designed to improve the quality of information a user was seeing and the cookie tracking technologies that gathered the information about the user (Pariser, 2011). Algorithms build profiles on their users and show them information based on their profile. Initially, these algorithms were used to improve the relevancy of the information given to the users, however people such as Pariser argued that there is a danger in only seeing the information that the search engine determines the user will agree with. Pariser noted in 2011, that the

algorithms that were originally created by social media, were being heavily adopted by the online news sites such as Huffington Post, Washington Post, New York Times in an attempt to grab user's attention in an ever increasingly competitive digital news space where it is hard to get users to pay for content. As Parser defined it, a person's filter bubble is "your personal unique universe of information that you live in online. And what's in your filter bubble depends on who you are, and it depends on what you do. But the thing is that you do not decide what gets in. And more importantly, you do not actually see what gets edited out" (Parser, 2011).

2.3 Tracking Techniques

Increasingly there are more and more sites that rely on third-party cookies, cross-site tracking and fingerprinting to build a user profile in order to refine the information shown and in many cases make suggestions for the products and services to purchase. At the same time, the level of user awareness about their privacy on the Internet and what technologies exist to store their information remains low. The following section explores a few different techniques that can be used to track users in more detail.

2.3.1 Location and Location-Based Advertising

Location has been and remains one of the most basic and cornerstone techniques for tracking users. Location based services such as maps or weather apps can be helpful for showing relevant data, however social networks and other sites can also collect user's data for targeted advertising. Location-based advertising (LBA) is a term used for advertisements that are tailored to the location where a potential customer accesses or sees the information. LBA is not only used in digital content, roadside advertisements and billboards are also considered LBA. Even if billboards are static and bound to their location, advertisers can still choose specific countries or regions where they believe their ad will be most relevant. LBA on computers is much more dynamic and can be determined by detecting the IP address of a user either with their consent or without. Location-based advertising is not new, however, since the massive rise in popularity of mobile devices, advertising can be more dynamic and accurate. GPS can be used on mobile devices to give a location accurate to roughly 3.5 meters. Using this information, advertisers can target the user depending on their proximity to different stores and restaurants given that people tend to carry their devices with them at all times (Bauer & Strauss, 2016).

2.3.2 Cookies

Cookies are the most common way to remember user preferences and track people on the Internet. Cookies are text files that websites can place on an individual's computer that can be used for identifying purposes. They are stored on the client-side and the client can transmit the information back to the server during interactions, such as when they load a page. Cookies were invented to remember the individual's information to reduce the number of logins, remember

preferences and show most relevant information. Advertisement and analytics companies have been making use of cookie technology for years to track user's browsing history and search habits to display targeted advertisements (Anlim, 2016, p.104).

2.3.2.1 First-Party Cookies vs. Third-party Cookies

First-party cookies are the cookies where its domain matches the URL visited. first-party cookies are not used to track user activity in order to collect and aggregate data about them or pass information from one site to another. first-party cookies are useful to save the login information and location data to display more relevant information to the user. That does not mean that the owner of the website cannot collect data through first-party cookies. The owner could collect data and use it to change how the website appears or the type of information displayed. If the owner collects data and sells it to outside organizations, that should be explained in the website's privacy policy.

Third-party cookies are the cookies where its domain is different from the website visited. Not all third-party cookies are aimed at tracking the users and invading their privacy. Cookies can provide performance enhancing services, widgets or cloud storage solutions. Online advertising and user analytics seems to be the most common type of third-party cookies (Wills & Uzunoglu, 2016). Third-party domains that can appear as first-party. For example upon first look "metrics.cnn.com" appears as if it was part of cnn.com. Upon further inspection and lookup it can be determined that the alias is cnn.122.2o7.net which is a third-party domain (Wills & Uzunoglu, 2016, p.2).

Although first-party and third-party cookies are most often the focus of researchers, there is also a potential category of second-party cookies. The existence of second-party cookies "has been a subject of contention" (Wlosik & Sweeney, 2020) for a while. The best definition of a second-party cookie are "cookies that are transferred from one company (the one that created first-party cookies) to another company via some sort of data partnership. For example, an airline could sell its first-party cookies (and other first-party data such as names, email addresses, etc.) to a trusted hotel chain to use for ad targeting, which would mean the cookies become classed as second-party" (Wlosik & Sweeney, 2020).

2.3.2.2 Single Session Cookies

Single Session cookies are usually first-party cookies and perform the same functions as such. The only notable difference is that these types of cookies are erased after the user quits their browser session (Anlim et al., 2016, p.104).

2.3.2.3 Persistent Cookies and Super Cookies (Zombie Cookies & Flash Cookies)

Persistent cookies are these types of cookies that are saved on a user's computer so when a browser application is closed, they can be retrieved next time. These types of cookies have an expiration date and should be removed from the system upon expiration, although this can be

years in the future and is up to the discretion of a programmer deciding on the cookie parameters (Anlim et al., 2016, p.104).

Super cookies are much harder to remove and cannot simply be removed by clearing the browser cookies. There are several different names that refer to these types of cookies depending on the method they use to stay on the user's computer and are referred to by the umbrella term of super cookies. Super cookies do have legitimate uses, one example being online video games to prevent users from cheating. They have been known to track users without their consent and install malicious software on a user's device. Verizon was one of the companies that came under a lot of fire from consumers for their use of supercookies (Peterson, 2015).

Flash cookies are a type of super cookies that can respawn themselves or other HTTP cookies. The idea of flash cookies became prominent in part because of the study "Flash Cookies and Privacy" from 2009. In the paper, the researchers determined that many websites during that era were using Flash cookies in order to make them more resilient to removal than regular cookies. Flash cookies can be set when a user visits a website containing Adobe Flash Technology content on the page, even if the user does not click on the banner or the ad itself (Soltani et al., 2009).

Zombie cookies earned their name because they can return to life after the user has deleted them. They are stored in multiple client side repositories and if the user neglects to delete one of them, they will repopulate. The practice behind the use of zombie cookies is widely frowned upon by consumers and in many countries and cases, the use of zombie cookies are a violation of user privacy laws (Sorensen, 2013).

2.3.2.4 Cookie Sharing and Cookie Syncing

Cookie sharing or cookie syncing as defined in the 2014 paper involves different websites that are not related to each other exchanging cookie information through embedded third-party trackers (Acar, 2014). Even if a user might never visit certain sites, tracking companies already have pre-built profiles and information on users which can be used across different sites. For more detailed explanation and diagrams, see the work by Englehardt (Englehardt, 2014).

Google Analytics is present on a huge portion of the Web's pages and makes use of cookie sharing. Google's cookie sharing script is able to receive identifiers from first-party cookies and link the request back to the user profile (Cyphers, 2019). In addition, since 2014 Google's algorithms have gotten much smarter and can use other identifying information such as IP address and even TLS state to link different cookie values to the same user (Cyphers, 2019). Google Analytics, Chartbeat, Nexac, Amazon Alexa Metrics and BounceX all implemented cookie sharing techniques in recent years.

2.3.3 Canvas Fingerprinting

Canvas Fingerprinting was first documented in 2012 by Mowery and Shacham (Mowery, 2012) and further researched in a paper from 2014 by researchers from Princeton University (Acar, 2014). Their paper was the first large scale study of the three most advanced web tracking mechanisms: canvas fingerprinting, evercookies, and cookie syncing.

Mowery and Sacham first found that by using Canvas API of modern browsers, interested parties can use the differences in the rendering of text or images to make identifications about operating systems, graphics cards, graphics drivers and browsers (Mowery, 2012). Furthermore, with the increase in the number of browsers adds-on, companies started using the number, the type and the version of the extensions that users have installed on their machine to build a unique fingerprint. A research project Panopticlick, by the Electronic Frontier Foundation (EFF), uses many metrics that companies might use such as screen resolution, color depth, timezone of the system fonts installation on the computer, system platform, language and others to build a uniqueness score (*Panopticlick 3.0*, n.d.).

Acar and others stated that the only effective technique that works as a countermeasure to fingerprinting is adding noise to the pixel data (Acar et al, 2014). The paper concluded that asking user permissions for each canvas read attempt is the best solution and pointed out that it is exactly the technique adopted by the Tor browser, which many articles recommend as a tool against canvas fingerprinting. Tor Browser “returns an empty image from all the canvas functions that can be used to read image data. The user is then shown a dialog where they may permit trusted sites to access the canvas” (Acar et. al, 2014).

After thorough investigations Panopticlick project suggests that the only methods for reducing efficiency of canvas fingerprinting: using the Tor browser, disabling JavaScript (such as by using NoScript) or using less common browsers (*Panopticlick 3.0*, n.d.).

2.4 Third-party Domain Classification

Given the differences described above, we needed a classification regarding the category and the purpose of the cookies. Like stated previously, third-party cookies can be used for many different purposes, from advertisements and analytics, to displaying video and comment sections. One common example is a comment section on popular websites. If a website uses a free comment extension to add an ability to comment to their website, they might be unknowingly embedding third-party trackers into their site. Previous work done by Wills and Uzunoglu established six categories for types of third-party domain classifications: AdTrackers, Analytics, Beacons, Social, Widgets, Others (Wills & Uzunoglu, 2016, p.2). Their definition was heavily influenced by the categories established at the time by privacy tools such as Ghostery, as well as Abine and TrendMicro. Ghostery has since updated its categories and now has eight categories: Advertising, Site Analytics, Customer Interaction, Social Media, Essential, Audio/Video Player, Adult Advertising and Comments.

2.5 Most Popular User Privacy Technologies

There are several methods that can be used to improve the user's privacy. Most popular include ad blocking or privacy focused (tracker blocking) browser extensions, browsers with built in protections (cookie or fingerprinting blocking), privacy-oriented search engines that claim to not track user's across searches, apps for mobile devices, VPN to mask the location, and incognito mode that does not save web browsing history. There are also network based level protections, such as hosts file and local content filtering proxy servers methods, which are less traditional and are not browser based methods (Uzunoglu, 2016). Examples include blocking domain names from being resolved via a PiHole or using interception proxy like Privoxy (Borgolte & Feamster, 2020).

2.6 Performance and Browser Extensions

For a long time there has been a conflict between browser developers (such as Google Chrome) who are financially motivated to track users' online behavior and browser extensions that try to offer enhanced protections to the users, but have frequently had their fair share of security and privacy issues. In addition to the security argument, browser vendors argue that the extensions that are not coded by the browser developers, can be used to reduce performance, negatively impact user experience and negatively impact user engagement and profit (Borgolte & Feamster, 2020).

A thorough research paper by Borgolte and Feamster carefully studied and researched the question of performance when it comes to browser extensions. They evaluated eight privacy focused extensions: Adblock plus, Decentraleyed, Disconnect, Ghostery, HTTPS Everywhere, NoScript, Privacy Badger and uBlock Origin and “[found] no evidence that privacy focused extensions fundamentally degrade performance in any way, but [their] results show that they improve performance across various metrics” (Borgolte & Feamster, 2020).

2.7 Summary

Users are growing increasingly concerned about their privacy as the amount and invasiveness of tracking techniques on the Internet grows. Users are justifiably concerned about the implications of these various tracking techniques and seek solutions to gain back the control over their data that can be used for emotional manipulation and increased sales of the products. Given these concerns, our project focuses on investigating various tools that can help users take control of their data and reduce the privacy invasion.

3. Research Questions

Given the prevalence of user concerns we establish in Section 2.2 and pervasiveness of various tracking mechanisms that can be used to violate user privacy that we explored in Section 2.3, the goal of our work is to explore which privacy tools exist on the market that can help the users to protect their privacy and reduce information leakage.

Our project had two main goals. First, to conduct a research study to understand the effectiveness of browsers, extensions, mobile applications, and search engines at protecting user privacy. Second, to create a recommendation system based on a survey approach to understand user views on privacy and based on the research in the first part, make recommendations on what they could do to better align their privacy practices with their personal concerns. To accomplish our goals, we developed the following research questions.

3.1 Privacy Protection of Browsers

Research question: Are there browsers that in their “out-of-the-box” state provide better privacy protections than others? Additionally, do features like “Block third-party” cookies on Google and Firefox Strict and Custom modes help to protect user privacy and to what extent?

We first evaluated whether different browsers provide better privacy than others. All of the browsers tested were based either on Chromium or Firefox source code. We used the Selenium library and Python script we developed to drive browsers through Alexa Top 100 Sites in the US. We were able to capture the HTTP/HTTPS traffic with Fiddler 4 as a proxy server and analyze the amount of privacy protection each browser provided. We measured the amount of privacy protection offered, by assigning each domain a category based on the Ghostery filter list (such as Advertising, Analytics, etc) and we decided whether certain browsers block more cookies and scripts than others in their out-of-the-box state.

3.2 Privacy Protection of Browser Extensions

Research question: Which of the popular browser extensions provide the best privacy protection and do ad blocking tools, list based privacy tools or heuristic based privacy tools offer the best protections?

We seek to build on previous studies that focused on the privacy protections offered by browser extensions. There have been many prior studies that looked into browser extensions and the additional privacy they provide to users (Wills, 2016) (Merzdovnik, 2017) (Mazel, 2019). We wanted to repeat the research in 2020, given that the extension landscape and the filters that they use changes year-to-year. We used a large number of the most popular extensions to determine which are the best for the users. Our work was influenced by several papers. First, work by Wills and Uzunoglu in 2016 is the first paper we examined that analyzed the effectiveness of existing ad blockers. Our work builds on their questions, methods, and findings, but aims to examine more tools, including browsers, which were not examined in their work. We also looked at the

work by Merzdovnik from 2017 and Mazel in 2019, which performed similar research to the work of Wills and Uzunoglu and also studied different extensions and assessed their effectiveness. Their work also influenced our research in terms of the extensions we choose to analyze and serve as a good reference point since it was more recent than Wills and Uzunoglu work from 2016. The most similar and recent work related to our research question is by Mazel et al. from 2019. Their research tested a large number of extensions and received compelling results. We wanted to repeat the experiment to see how our findings compare and whether we come to similar results.

3.3 Privacy Tools vs. Website Degradation

Research question: What is the amount of website degradation caused by different privacy tools?

In addition to studying the effectiveness of different privacy tools, we also wanted to examine the website degradation that they cause. It has been established by other researchers (Uzunoglu, 2016) (Merzdovnik, 2017) (Mazel, 2019) that an extension like NoScript while providing best privacy protection also results in the biggest website degradation and prevents users from interacting with websites in the usual way. We set out to analyze the amount of website degradation caused, categorize it as minor or major degradation depending on the severity of the issue in order to in the second part of this work make a recommendation to the users based on the desired amount of privacy protection, as well as tolerance for website degradation.

3.4 Privacy Protections of Mobile Applications

Research question: What applications for mobile devices (Android) provide best privacy protections for a user?

Mobile devices and tablets are growing increasingly more sophisticated and popular. Most mobile and tablet devices have enough processing power to replace a desktop or laptop computer for daily operations. Users are becoming increasingly more mobile with their devices. In 2019, mobile devices accounted for 54.44% of the market share, while desktop/laptop computers accounted for 40.63% (NetMarketShare), providing an increasing incentive for advertising and analytics companies to invest into techniques that are able track their mobile users. To further exacerbate the issue, stores and companies can use location data to show advertisements to their users based on the locations that they visited or locations that are close. We aim to examine protections that exist for mobile platforms, especially since we were unable to find any similar research in the field.

3.5 Privacy Protections of Search Engines

Research question: Are there search engines that provide their users with better levels of privacy protections than others?

According to the NetMarketShare, Google search engine accounted for 82.23% of the searches across all devices in 2019. While search engines such as DuckDuckGo are growing in popularity, it only accounted for 0.33% in 2019. Search engines, just like browsers, make more money when more consumers use their platform. Popular search engines such as Google and Bing make a portion of their profits from tracking users around the web. Given the user concerns associated with all the ways they are tracked on the web, several companies such as DuckDuckGo, SwissCows, Privado and Qwant emerged as competitors to Google and Bing. The previously listed companies offer their own search engine and claim they respect user privacy. In the context of search engines, respecting user privacy means that these search engines claim they only show advertisements related to the search term typed into the engine at the time, and not any of the previously collected information. Given that we were unable to find any similar research in the field, we set out to investigate alternatives to Google and Bing and what privacy protection they offer.

3.6 User Privacy Concerns and Survey

Research question: to understand user concerns on privacy and make recommendations on what they could do to better align their privacy practices with their personal concerns.

In order to best understand user concerns on privacy and make a potential recommendation, we develop a survey. We decided to make more personalized recommendations based on the desires of an individual. We designed a survey to provide custom recommendations for each user based on how much they indicate that they wish to protect themselves while online. A recommendation tool, via a survey software - Qualtrics - was designed and also acts as the perfect mask for additional experimental work that we wanted to conduct. Since this project is interdisciplinary with psychology, we are using the survey to evaluate additional questions that focus on a variety of questions. We wanted to identify which types of advertisements users specifically care about, if the types of technologies owned and ran by users had any effect on their knowledge of privacy, and providing the user with a bit of their own personal data that was synthesized by Google to see if that would affect their willingness to switch to more privacy-protection tools. The work by Zeljkovic from 2010 which examined different web browsing habits of users by creating a website that showed personalized tracker information gave our group the idea to build a personalized survey. The researchers created a website that provided information about common tracking mechanisms and showed them based on their individual browsing history what trackers were embedded on the sites they visited. Given a decade of

improvements in user privacy and additional mechanisms that are built into the modern browsers, and our concern for the privacy of our users, our group did not pull data from user's browsers about the sites they visited, but rather conducted our own research on a test set of Alexa top 100 sites and analyzed which browsers, extensions and search engines offered the best protection, and provided that recommendation to the user based on some of their own personal preferences and tolerance for website breakage through the user of a Qualtrics survey.

3.7 Summary

Users concerned about their privacy are looking for the best tools that they can use to reduce the amount of information that companies collect on them. The questions that we propose will ultimately help to establish the best combination of tools to recommend to the users. The next Chapter will present a methodology that will be used to find answers to those questions.

4. Methodology

In order to answer research questions listed in Section 3, we developed the following methodology. Given that the main focus of our work was browsers, extensions and the degradation caused, the following section presents the methodology to answer the first three research questions from Section 3.1-Section 3.3. Methodology and the results for Sections 3.4-3.6 is presented later in the paper.

4.1 Browsers

Although Google Chrome is the most popular browser (NetMarketShare), there are a lot of alternative browsers that come equipped with privacy enhancing features. While users have access to a variety of privacy-oriented extensions, they might still prefer to select an alternative browser which, out-of-the-box, comes stripped of many tracking techniques and other privacy measures built in, which eliminates the need for users to take additional steps to download extensions.

Additionally, users also have access to browser extensions that can perform similar functions to privacy-focused browsers. In the first part of our work, we studied how much privacy protection browsers offer in their out-of-the-box state, but in the second part of our work we examine the most popular privacy extensions and how much protection they can provide.

4.1.1 List of Browsers Tested

Table 4.1 shows the browsers that we have tested in our project for the first research question along with the market share numbers to provide an idea about popularity of the tools.

Table 4.1. List of Browsers tested and their features.

Name of the Browser and Configuration	Chromium or Firefox Based?	Market Share (2019) ¹	Version Tested
Google Chrome	Chromium	65.40%	87.0.4280.88
Google Chrome (Block Third-Party Cookies)			
Firefox	Firefox	4.38%	83.0
Firefox (Strict Mode)			
Firefox (Custom Mode)			
Edge	Chromium	2.39%	87.0.664.66
Opera	Chromium ²	0.88%	73.0.3856.284
Brave	Chromium	-	1.18.75
Vivaldi	Chromium	-	3.4.2066.99
Iridium	Chromium	-	11.85.0
Epic	Chromium	-	84.0.4147.105
Ungoogled Chromium	Chromium	-	85.0.4183.121

1. **Google Chrome:** Google Chrome is the most well known browser on this list and accounts for 65.40% of the Market Share across devices in 2019 (NetMarketShare). Since Google Chrome is made by Google, all of the actions performed by the users are collected, recorded to their profile and used for advertising and analytics.
2. **Firefox:** Firefox is the first and the most popular browser that offers privacy and security features. Firefox is well regarded as an all around good browser, both for privacy and non-privacy-oriented users. Firefox has three levels of privacy - Standard, Strict or Custom - which our group tested. The out-of-the-box, standard mode is not the best for privacy, but the settings can be adjusted. Firefox allows users to choose what they want to

¹ The data is from “Browser Market Share” from NetMarketShare.

² Opera began to base its browser on Chromium starting with version 15.

block in the custom mode including cookies, tracking content, cryptominers and fingerprinters.

3. **Microsoft Edge:** Microsoft Edge is the newer browser from Microsoft, meant to replace Microsoft Internet Explorer after the company has stopped the support for the browser. Notably, unlike Internet Explorer, which was built on the Microsoft Trident engine, Edge is built on Chromium.
4. **Opera:** Opera used to be considered a privacy-oriented browser with good performance. In 2016 it was sold to Chinese consortium and its new privacy policy explains that the browser collects a lot of data on the users (Taylor, 2020).
5. **Brave:** Brave is a chromium-based browser which claims to be privacy-focused right out-of-the-box, unlike Firefox, which requires customization. By default, it will block ads and trackers, and it has several modes including more strict modes. It also has built in protections such as enforcing HTTPS connections, protections against fingerprinting with built in Tor and built in script blocker. Brave is fairly new so not a lot of research has been done investigating the practices and policies behind the company. Brave has received criticism and raised questions after launching its own ad program in April 2019. Brave decides which ads are acceptable and shows pre-vetted ads to their users and gives back a share of the revenue to the websites. Many called out the program since the browser is taking away the full revenue from the websites, and then gives itself and the website a fraction of the revenue if it deems the ads to be acceptable (Taylor, 2020).
6. **Vivaldi:** Vivaldi claims to be “fast, private and secure browser that blocks ads and trackers” (*Vivaldi Browser*, n.d.). Users have raised questions about their privacy policy which states that the browser assigns a unique user ID to the user that is stored on the computer. The company claims it is for the purpose of determining the total number of active users, but not all users believe the claim (Taylor, 2020).
7. **Iridium:** Iridium is a browser based on Chromium open source and it comes with numerous privacy enhancing features. A few notable changes are the use of “do-not-track” header, disabled autocomplete, blocking third-party cookies by default, keeping cookies only until the browser is closed, not storing passwords by default, and adding Qwant as a default search provider and DuckDuckGo search as an alternative (*Differences between Iridium and Chromium*, 2019).
8. **Epic:** Epic claims to be a privacy browser based on Chromium. The browser is not open source, and several users found it connects to Google upon startup (Taylor, 2020). It claims to block ads, trackers, fingerprinting and crypto mining and also a free VPN.
9. **Ungoogled Chromium:** Ungoogled Chromium is an open source project based on Chromium source code that strips away Google privacy issues from the source code and removes dependencies on Google web services. Unfortunately most of the privacy controls and modifications have to be manually activated (Taylor, 2020).

4.1.2 Evaluation

We evaluated which browsers listed above provide better privacy than others. All of the browsers tested were based either on Chromium or Firefox source code. We used the Selenium library and Python script we developed to drive browsers through Alexa Top 100 Sites in the US. We captured the HTTP/HTTPS traffic with Fiddler 4 as a proxy server and analyzed the amount of privacy protection each browser provided. We measured the amount of privacy protection offered, by assigning each domain a category based on the Ghostery filter list (such as Advertising, Analytics, etc) and we decided whether certain browsers block more cookies and scripts than others in their out-of-the-box state.

4.1.2.1 Logging Traffic

The cornerstone of our project was gathering HTTP and HTTPS traffic from different websites to determine the privacy risk associated with them. We started off by looking for tools available for free on the Internet that could fulfill our goal and inspect website traffic. We needed to inspect and decrypt HTTPS traffic because the prevalence of HTTPS traffic grew.

The first possible option we looked into was Burp Suite, which we still think is a good alternative especially since it is available on the range of platforms and machines. The most challenging part of using Burp suite is setting up the browser with the appropriate SSL certifications to allow Burp Suite to capture encrypted traffic. The process varies from browser to browser. It is possible to set it up on a variety of browsers including Firefox and Chrome (the browsers that we personally tested). The other option is using Burp's embedded browser, Chromium, eliminating the need to manually configure proxy settings. Using Chromium is a viable alternative for those that do not have the need to examine traffic from different browsers. For the purposes of our project, we were interested in testing a variety of browsers including Chrome, Firefox, Brave and others to see to what extent they block tracking cookies out-of-the-box.

Next we experimented with Telerik Fiddler Everywhere v1.0.2 which is a free tool available on a variety of platforms including Windows, OS and Linux and has an updated interface (*Fiddler Everywhere*, n.d.). There are a lot of benefits to using Fiddler Everywhere as a tool for traffic inspection and we were not the first group to use it for the purposes of studying Internet privacy. The best feature of Fiddler is the ability to easily set up Trust Root Certificate to inspect HTTPS traffic. Regardless of the browser used to conduct the tests, Fiddler needs to be granted permissions only once and will be able to inspect all HTTPS traffic from any browser. Fiddler Everywhere is also a great collaboration tool for teams since it allows for log sharing and import/export of gathered data, which allows groups to work together and examine the same gathered information.

Finally, we settled on Fiddler Classic which comes with a variety of options and allows for an installation of custom add-ons (*Fiddler Classic*, n.d.). Unfortunately, Fiddler Classic is available only on Windows so all of the research in this study was conducted on a Windows

machine. Initially, we added a Privacy Scanner extension to our Fiddler Classic, which “flags responses that set cookies and color codes based on P3P headers” (*Fiddler Extensions: Privacy Add On*, n.d.). The extension adds an additional “Privacy Info” column to the session list and flags HTTP/HTTPS responses that set cookies. The extension also color codes these flagged responses based on P3P statements, depending on whether they deemed the P3P policy satisfactory or not (*Fiddler - View Cookie Information*, n.d.). In the end, we did not end up using the add on to analyze cookies specifically, and instead examined all the traffic.

4.1.2.2 Driving Browsers

In order to drive the browsers to drive through all of the websites, we also needed a tool to automate the process. We began by examining Cypress as an option because we originally anticipated performing more interaction with the websites and given that Cypress is excellent for making specific keystrokes and behavioral testing, it was a logical first choice. We soon realized that we were not going to interact with websites a lot, but instead just visit them and collect traffic. The primary issue we ran into with Cypress is the fact that it couldn’t connect to proxies, so we would be able to only log HTTP but not HTTPS traffic.

Then, we moved to using Selenium library. Selenium is great for opening up the browser and monitoring the interaction for specific events, as well as sending keystrokes to those websites. Originally, we started with JavaScript because we assumed since JavaScript was the code for the browsers, it would be easier to manipulate the sites and get the “.dom” elements if needed to examine specific pieces of content. JavaScript is asynchronous, which meant that when we provided a list of browsers that we wanted to test and a list of 100 websites, the code would attempt to do that in parallel, instead of launching a specific browser, running through the list of websites, closing and proceeding onto the next browser.

We eventually landed on sticking with the Selenium library, but making it work with Python. Python is a scripting language, unlike JavaScript so the code runs one line and finishes it before starting another.

We downloaded all of the binaries needed for each web browser from their website. We found out that all of the browsers we wanted to test were either based on Chromium or Firefox. Initially, we intended to test more browsers than in the final list such as IceCat that were based on Firefox. In the final list only the Firefox itself was based on Firefox source code. In order to be able to instruct the code to interact with the actual binaries, we downloaded Chromedriver and Geckodriver. Chromedriver and Geckodriver and tools we used to talk to Chromium and Firefox based applications respectively. The drivers open up a server, and make calls to the binaries of the browsers, Selenium in turns instructs the server on what to do, in our case visit a specific list of websites.

4.1.2.3 Categorizing Domains

In our project, just like other works before us (Wills, 2016), we were primarily concerned with domains that would invade users' privacy and fall into the “trackers” category and not some other domains that seem to provide essential functions. In order to assess which domains would actually fall into the “trackers” category, we assigned a category to each domain we encountered and built our own database using a combination of existing filter lists from Ghostery, Disconnect, DuckDuckGo and Better extensions. As discussed in Section 2.4, Ghostery has updated its categories. We use updated categories established by Ghostery and classify the third-party cookie domains according to the filter list to gain an understanding of the types of cookies that are allowed by different extensions and those that are blocked.

First, we combined all of our Fiddler files into one and focused on the hostnames. From every hostname, we extracted the domain (turning a.doubleclick.net into doubleclick.net) using a line in our Python code shown in Figure 4.1 that split the hostname on the dot and took the last two elements of the address using the following lines.

```
join = join + line.split('\n')[0].split(".")[-2]
join = join + '.' + line.split('\n')[0].split(".")[-1]
```

Figure 4.1 Python code developed to split the hostname and extract just the domain name.

After we extracted the domains from the full host name, we removed all duplicate domains from the list leaving us with 573 unique domains. After we identified the unique domains, we came up with categories based on the Ghostery category labels as of October 2020. Ghostery is regarded as the best filter list and has been used by other works as the basis for categories (Wills, 2016). Compared to the work from 2016, Ghostery has added new categories labels that were not present in other works. The categories from ghostery are as follows: Advertising, Site Analytics, Customer Interaction, Social Media, Essential, Audio/Video, Adult Advertising, and Comments. We also added our own three categories: Unknown - for domains we were unable to identify, first-party - for domains that matched the first-party site and Extensions - for domains that were generated by installing the extension itself (ex. Adblock Plus making calls to install the extension).

Given that the Ghostery database is proprietary, we ran into the same issue that Uzunoglu did (Uzunoglu, 2016). Unlike other extensions, Ghostery does not make its database public so there is no easy way to tell which domain belongs to each category. One workaround that Uzunoglu used in 2016 is visiting each of the concerning sites individually and looking for the categories that Ghostery assigned to the third-party domains. When a website is visited, Ghostery extension flags the third-party domains it sees, and displays the domain name and the category it determines based on its proprietary filter list like the one shown in Figure 4.2 and there is no easy way to access Ghostery's central database.

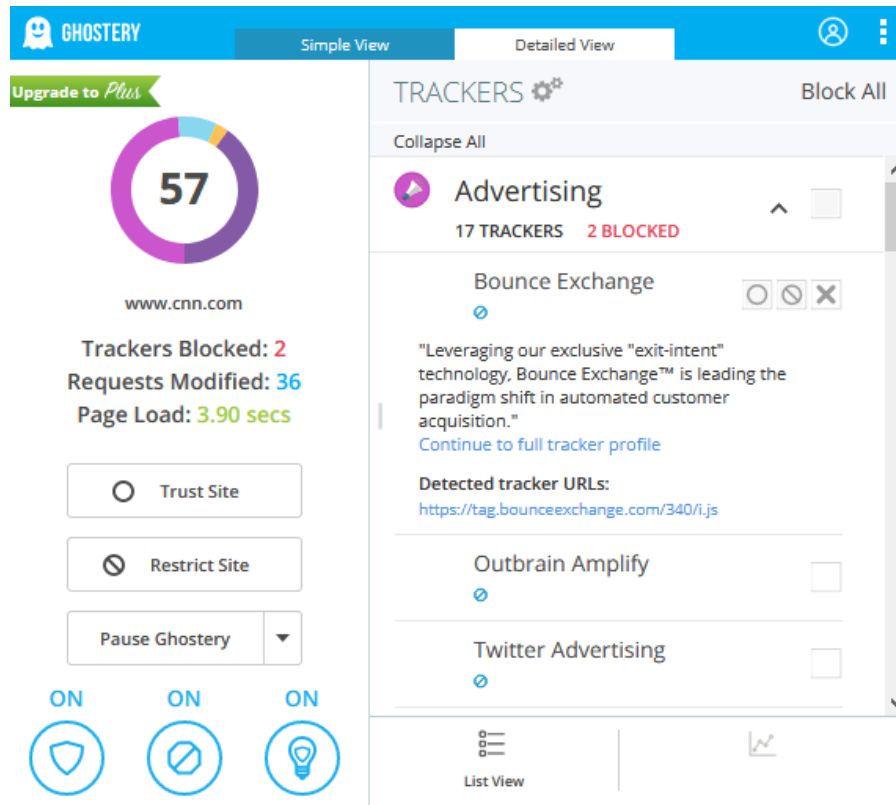


Figure 4.2. Screenshot of the Ghostery extension on Chrome, when visiting cnn.com, showing Bounce Exchange and detected tracker URL.

Our group developed a workaround: we extracted Ghostery's list by accessing the extension list and using Chrome inspect element feature to download the HTML code.

By clicking on the extension → Settings → Global Blocking, we were able access the Ghostery list with domain names and their categories to choose which blocking lists they wish to enable and disable as shown in Figure 4.3.

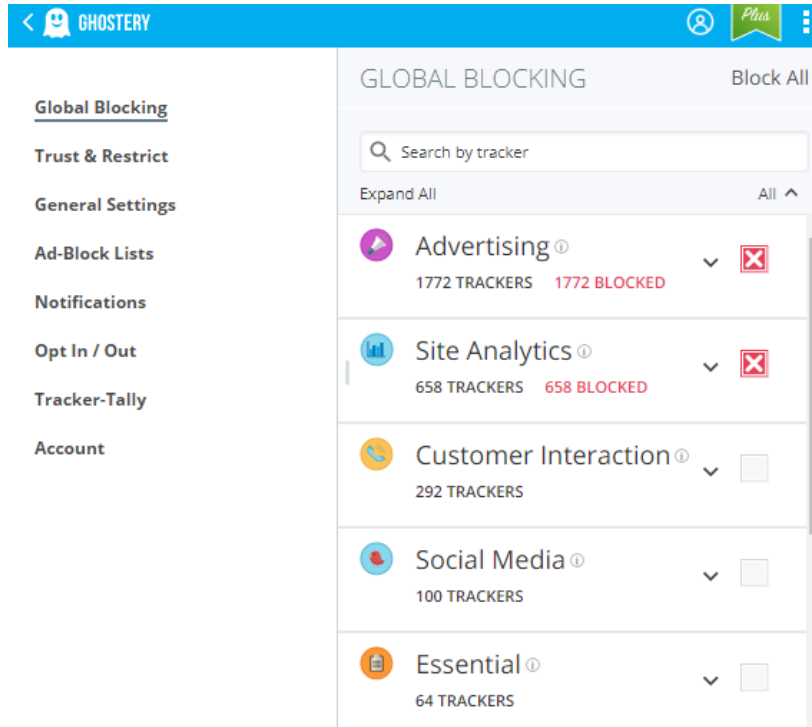


Figure 4.3. Screenshot showing Ghostery Global Blocking List and the categories inside.

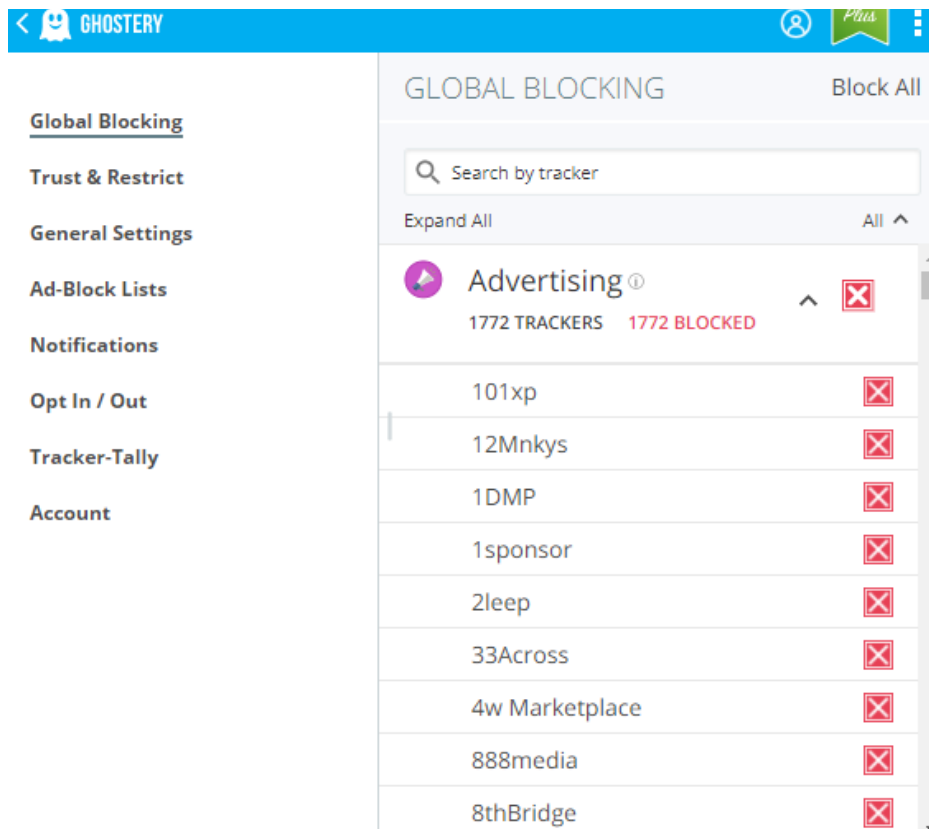
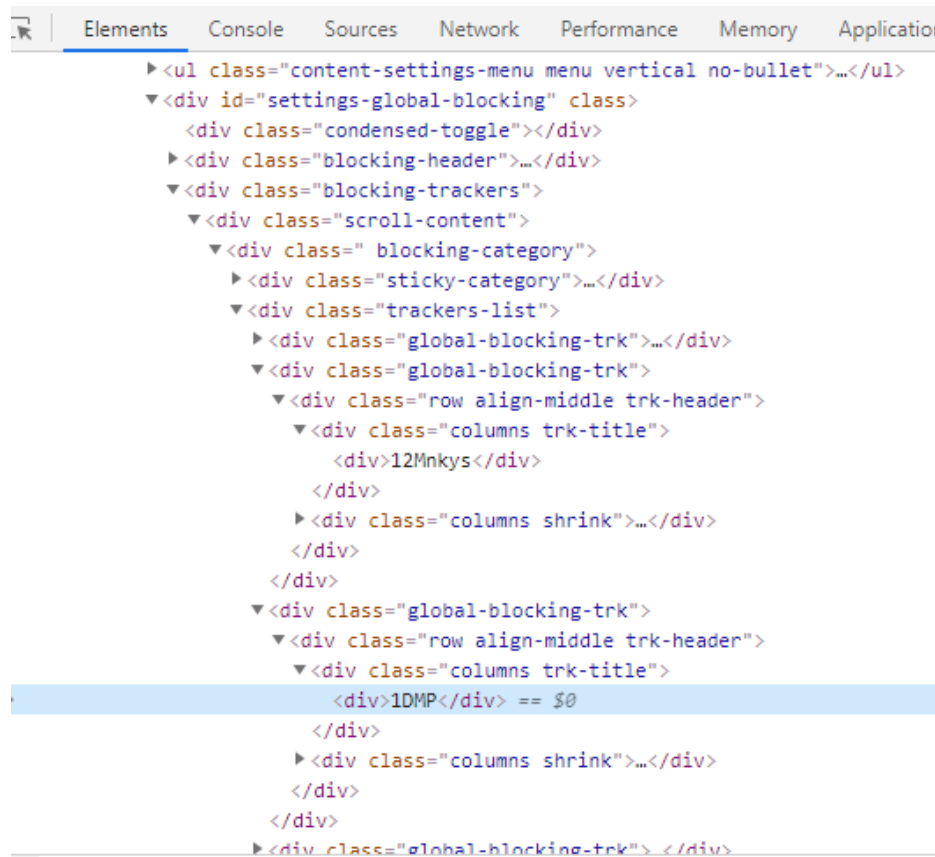


Figure 4.4. Screenshot showing Ghostery's Blocking list, specifically Advertising category.

Ghostery does not make it easy to copy the following list. In order to be able to copy the names from the categories, our group used the “Inspect” feature in Chrome to view the code behind the extension as shown in Figure 4.5.

We looked inside several div elements, specifically under “columns trk-title”, shown in Figure 4.5, and were able to extract all the names from each category. We copied the code from the Chrome code editor, saved the code as HTML files and turned to BeautifulSoup tool to parse through the elements of the web page. We used BeautifulSoup4, Python web scraping library, to parse through the elements of the page in order to look for the names inside the div elements, and extracted only the names and lists of Advertising, Site Analytics, Comments and e.t.c. We converted our HTML file into a JSON file in order to simplify the parsing process and ran the BeautifulSoup tool to get the names of the companies associated with each category.



```
Elements Console Sources Network Performance Memory Application
▶ <ul class="content-settings-menu menu vertical no-bullet">...</ul>
▼ <div id="settings-global-blocking" class=
  <div class="condensed-toggle"></div>
  ▶ <div class="blocking-header">...</div>
  ▼ <div class="blocking-trackers">
    ▼ <div class="scroll-content">
      ▼ <div class=" blocking-category">
        ▶ <div class="sticky-category">...</div>
        ▼ <div class="trackers-list">
          ▶ <div class="global-blocking-trk">...</div>
          ▼ <div class="global-blocking-trk">
            ▼ <div class="row align-middle trk-header">
              ▼ <div class="columns trk-title">
                <div>12Mnkys</div>
              </div>
            </div>
            ▶ <div class="columns shrink">...</div>
          </div>
        </div>
      </div>
    </div>
  </div>
  ▼ <div class="global-blocking-trk">
    ▼ <div class="row align-middle trk-header">
      ▼ <div class="columns trk-title">
        <div>1DMP</div> == $0
      </div>
    </div>
    ▶ <div class="columns shrink">...</div>
  </div>
</div>
▶ <div class="global-blocking-trk"> </div>
```

Figure 4.5. Screenshot showing Chrome code inspector with a snippet of code from the Ghostery extensions.

Even though we were able to extract company names and their mapping to the Ghostery category, we still did not have a mapping in terms of which company was behind which domain. Not all domains have easily recognizable names. For example doubleclick.net is a well known as Double Click if we saw doubleclick.net in our list of domains, we could quickly associate it with

Double Click name extracted from Ghostery and assign it a category, while something like “company.target.com” which maps to Demandbase is not immediately apparent.

Initially, we attempted to solve this problem using the “whois” command in our code. We found that there is a limit on how many requests one can make, which was actually stated in the terms of service for the tool so we ended up getting a warning message from the company that we were going against their policy by scripting it.

We came up with an alternative solution and relied on the Better FYI extension database that mapped trackers on Alexa Top 500 News Sites and mapped the names of the domains to the company names (*Trackers on Alexa Top 500 News sites*, n.d.). Better FYI dataset does not explicitly assign categories to trackers such as Advertising vs. Site Analytics, but rather only makes a distinction between a “tracker” and an essential call. So we did not rely on Better FYI dataset for the purposes of assigning categories, but rather only helped us establish a name mapping between our domain names and company names. We extracted the list from their trackers collection, created columns in Excel for 573 domains we encountered, used VLOOKUP function to assign them a name to the best of our ability based on the Better FYI dataset and then looked in the Ghostery Database to identify a category. Based on the combination of datasets from Ghostery and Better FYI we were able to generate our own database containing 573 entries with categories assigned that we used in our code to go through our Fiddler files and assign a category to each domain we encountered. Later, we used the summed up count for each category and represented those statistics on our diagrams as shown in the our results section.

4.2 Extensions

In addition to studying browsers, we repeated and built on previous studies that focused on the privacy protections offered by browser extensions. There have been many prior studies that looked into browser extensions and the additional privacy they provide to our users. We wanted to repeat the research in 2020, given that the extension landscape and the filters that they use changes year-to-year. We used a large number of the most popular extensions to determine which are the best for the users. The most similar and recent work related to our research question is by Mazel et al. from 2019. Their research tested a large number of extensions and received compelling results. We wanted to repeat the experiment to see how our findings compare and whether we come to similar results.

4.2.1 List of Extensions Tested

There are three categories of browser extensions we studied: AdBlockers, Privacy Tools and Miscellaneous. Although there is overlap between the tools, we want to make a distinction between tools that were aimed at simply blocking ads versus those that were more comprehensively designed to protect user’s privacy vs. tools such as NoScript and HTTPS everywhere that were not designed to protect a user's privacy per say. The goal of ad blockers is to prevent intrusive advertisements and to a certain extent, prevent user tracking by third-party

ads. The goal of the privacy-oriented tools is to block as much user tracking as possible and potentially block a portion of the ads if the code behind them is intrusive.

Also, previous work has established that the effectiveness of ad blocking tools depends on their underlying ruleset. As defined by the Mazel, rulesets can be divided into three categories: community-driven, centralized and algorithmic (Mazel, 2019) which are the categories we used for our categorizations.

For relative comparison of the popularity of the tools, as shown in Table 4.2 with the list of extensions we tested, the number of users according to Chrome Store and Firefox Store as of October 2020, the rulesets that it uses, the version we tested and which works we examined also tested the extension.

Table 4.2. List of browser extensions tested during the project and their features.

Name of the Extension	Ruleset	Users on Chrome Store ³ (October 2020)	Users on Firefox ⁴ (October 2020)	Version Tested	Works Studied
AdBlock	Community driven filter list	10,000,000	988,082	4.20.0	(Merzdovnik, 2017) (Uzunoglu, 2016)
AdBlock Plus	Community driven filter list	10,000,000	6,893,963	3.9.5	(Merzdovnik, 2017) (Mazel, 2019) (Uzunoglu, 2016)
uBlock	Community driven filter list	700,000	-	1.29.2	(Uzunoglu, 2016)
uBlock Origin	Community driven filter list	10,000,000	4,460,918	1.33.2	(Merzdovnik, 2017) (Mazel, 2019) (Uzunoglu, 2016)
AdGuard	Community driven filter list	7,000,000	352,255	3.5.12	(Uzunoglu, 2016)
Privacy Badger	Algorithmic	1,000,000	912,213	2020.8.25	(Merzdovnik, 2017) (Mazel, 2019)
MyTrackingChoices	Algorithmic	304	-	1.0.9	(Mazel, 2019)
Disconnect	Centralized filter list with categories	700,000	109,670	20.1.1	(Merzdovnik, 2017) (Mazel, 2019) (Uzunoglu, 2016)
Ghostery	Centralized filter list with categories; proprietary tracker database	2,000,000	1,026,460	8.5.2.1	(Merzdovnik, 2017) (Mazel, 2019) (Uzunoglu, 2016)
Blur	Centralized filter list with categories	100,000	19,688	8.1.2515	(Merzdovnik, 2017) (Mazel, 2019)
DuckDuckGo Essentials	Centralized filter list with categories	3,000,000	1,180,508	2020.8.12	
NoScript	Indiscriminate	100,000	392,178	11.0.46	(Mazel, 2019)
HTTPS Everywhere	Other	2,000,000	685,409	2020.8.13	(Mazel, 2019)

³ The numbers of users on Chrome Store were gathered in October 2020.

⁴ The numbers of users on Firefox were gathered in October 2020

Table 4.2 presents a list of the extensions we tested in our work. The following paragraphs review each extension in more detail, specify what is unique about it and why it was important to be tested in our work.

1. **AdBlock Plus:** AdBlock Plus is a free and open source browser extension that blocks advertisements. The goal of the tool is to block the most intrusive ads such as pop-ups, flashing banners and ads that might be malicious (*About Adblock Plus*, n.d.). An interesting feature of AdBlock Plus is their “Acceptable Ads” program. AdBlock Plus has a list of criteria that the advertisers must meet to be considered an acceptable ad and AdBlock Plus claims that participants cannot pay to avoid that criteria. Criteria for the acceptable ads includes placement, clear labeling, and size. According to AdBlock Plus, acceptable ads are those that are not intrusive or annoying, do not interfere with content, and are clearly labeled as an ad. Acceptable Ads are enabled by default when a user installs the extension since AdBlock Plus claims to want to support free websites that comply with their rules (*Allowing acceptable ads in Adblock Plus*, n.d.). This might be an appealing option to consumers that still want to support free websites and allow advertising as long as it is not invading their privacy. Users can at any time disable Acceptable Ads. AdBlock Plus is financed through their acceptable ads initiative, they charge large entities (those that gain more than 10 million ad impressions per month according to AdBlock’s definition) a fee for whitelisting their services. AdBlock claims that all the participants, paying and non-paying must abide by the same acceptable ads criteria and they do not whitelist companies that do not meet them. AdBlock Plus works based on a filter list, which is a set of rules that tells the browser which elements to block. AdBlock users can choose pre-made lists or make their own. By default AdBlock Plus contains a filter list that blocks ads, in their case EasyList, and an acceptable ads list which whitelists ads (*About Adblock Plus*, n.d.)
2. **AdBlock:** AdBlock and AdBlock plus get frequently confused. They are two different products. AdBlock was created by Michael Gundlach in 2009 and was originally created specifically for Chrome and titled “AdBlock for Chrome” when the AdBlock Plus team that originally designed their extension for Firefox was not interested in supporting Google Chrome. By now both extensions added support for additional browsers and ended up with similar names (*What’s the difference between AdBlock and Adblock Plus*, 2020). Although AdBlock Plus and AdBlock are different tools, they have the same goals and work similarly. AdBlock was also designed to give users control over what they see in their browser, prevent intrusive ads and block the advertisers that track people (*AdBlock*, n.d.). The company also believes in supporting free websites and they participate in the Acceptable Ads program which is run by non profit Acceptable Ads Committee. The company also receives a fee for whitelisting sites on the Acceptable Ads

list and allows the users to disable them if they wish, although it is turned on by default (*Acceptable Ads FAQ*, n.d.)

3. **uBlock and uBlock Origin:** Similar to the AdBlock vs. AdBlock Plus case above, there is a lot of confusion between uBlock vs. uBlock Origin. Although both products started off as the same product, the company got split into two creating uBlock and uBlock origin and now claim to be two separate products. Both extensions are free to use tools that work to block ads and many trackers used to keep track of user behavior. uBlock origin also makes use of the following filter lists: uBlock Origin Filter list, Easy List, Easy Privacy, Peter Lowe’s ad tracking, and Online Malicious URL Blocklist. As is with other tools, users can always add their own filter lists, modify existing filter lists or add their own custom rules (*uBlock FAQ*, n.d.) (*uBlock Origin*, n.d.)
4. **AdGuard:** AdGuard is an interesting example on our list for a few reasons. The first and the major reason is the number of entries on the filter list. AdGuard is a competitor to AdBlock and AdBlock Plus and works from a filter list of more than 10,000 rules. It claims to have one of the largest tracker filter lists on the Internet, larger than the database of Ghostery and Disconnect (*AdGuard AdBlocker*, n.d.). This claim is true, AdGuard has one of the most aggressive lists on the Internet. The filter list in AdGuard version 2.3 consolidated over 90 community created lists and then the final list was cleaned, sorted and duplicates were removed. AdGuard does mention “the list is very aggressive so please ensure that you add your own custom filtering rules for domains you want to allow” (*Filter List for AdGuard*, n.d.). It also claims to use half as much memory as other popular solutions such as AdBlock and AdBlock Plus (*AdGuard AdBlocker*, n.d.)
5. **Privacy Badger and MyTracking Choices:** Privacy Badger is a privacy extension that was developed by the Electronic Frontier Foundation. At the time of our research, Privacy Badger remains a unique privacy extension unlike the others that exist on the Internet. Most of the ad blocking and privacy extensions operate based on the filter list, such as AdBlock, AdBlock Plus, uBlock, Ghostery and Disconnect. The filter list contains a list of domains that are considered intrusive or malicious. The biggest disadvantage to the filter list method is that it takes a long time to manually curate and maintain the filter list. Filter lists also means that an individual or a group of individuals had to make judgement calls regarding third-party domains and classify them as intrusive or acceptable. The only other existing tool as far as we are aware of that functions similarly to Privacy Badger is an extension called MyTracking Choices (Mazel et. al, 2019). Just like Privacy Badger, instead of maintaining a filter list, the extension classifies a third-party domain as a tracker if it is present in three or more different domains. MyTracking Choices is a collaboration project that is aimed at designing an extension that can support ads and block the privacy intrusive ads. The team behind the extension operates under the assumption that “some categories of web pages (for example, related to health, religion, etc) are more privacy sensitive to users than others”

therefore they give users the option to block trackers on a specified set of categories of web pages (Achara et. al, 2016). Mazel, Garnier and Fukuda found that MyTracking Choices provides the users with a moderate amount of protection, but worse than even untrained Privacy Badger (Mazel et. al, 2019).

6. **Ghostery:** Ghostery is a well regarded privacy extension that has existed since 2009 and has been studied and given high marks by several papers (Mazel, 2019) (Wills & Uzunoglu, 2016) (Merzdovnik, 2017). The extension operates based on a comprehensive and extensive filter list that assigns classifications to different third-party trackers. Ghostery operates on a centralized filter list maintained by the company behind the extension, unlike AdBlock or AdBlock Plus which work based on community driven rulesets (Merzdovnik, 2017). Their filter list is well regarded as the 2016 work by Wills and Uzunoglu used their definitions of third-party trackers and assignments for third-party trackers for their work. As of 2020, Ghostery offers five products: Ghostery Browser Extension, Midnight, Insights, Start Tab and Ghostery Privacy Browser. Ghostery Browser Extension is the first and most popular tool covered by different research papers. Ghostery Midnight is a desktop application that functions similarly to the browser extension with an added VPN feature. Insights is a tracker analysis tool designed for teams that own their own websites and want to optimize their loading times, gather analytics about trackers and tags on their page and identify risks that may impact user security. Start Tab is an extension for Chrome that replaces the existing new tab page and provides Ghostery anonymous quick search feature.
7. **Disconnect:** Disconnect is a privacy tool that has been extensively studied by several papers over the years. It aims to help the users stop tracking by third-party sites and classifies them as Advertising, Analytics, Social or Content based on a pre-existing filter list. It also offers a Premium version with VPN and address masking features. They allow ad tracking websites that commit to respect user's Do Not Track (DNT) preferences and agree to comply with DNT as defined by Electronic Frontier Foundation (*Disconnect Help*, n.d.). It is similar to Ghostery and operates based on a centralized filter list maintained by the company, instead of community based lists.
8. **Blur:** Blur is a browser based privacy extension, although it can be viewed as a more comprehensive web privacy solution. It is made by the company called Abine that also makes Delete Me which is a tool that helps people remove their information from personal search engines. Blur operates on a company driven filter list like Ghostery and Disconnect (Merzdovnik, 2017). Blur has a free version and a paid version. Blur has a few additional features aimed for protecting the consumer's information. Blur can help users generate strong passwords and can function as a password manager, although it only stores them on a local machine. It also lets users mask their email address. Blur also provides masking for phone numbers and credit cards. Although Blur does not provide specific information about the details behind the extension, previous work that looked at

the extension (Mazel, 2019) (Merzdovnik, 2017) determined that it works similarly to other extensions such as Disconnect and Ghostery based on a proprietary filter list. The extension has become a lot less popular over the years, but we are still including it on our list in order to compare our findings with 2016 and 2019 studies.

9. **DuckDuckGo Essentials:** DuckDuckGo Essentials is a privacy-oriented extension that aims to block third-party trackers. A unique feature of the tool is the fact that the extension assigns almost every website a user visits a grade on a scale from “A” to “F” based on the number of trackers found, number of major tracker networks found and privacy practices. It gives a grade to the site privacy practices based on the “Terms of Service; Didn’t Read Project” which is an initiative that aims to rate and label website terms and privacy policies (*Terms of Service, Didn’t Read*, n.d.). DuckDuckGo Essentials extension also performs a similar function to HTTPS everywhere and attempts to direct a user to the encrypted HTTPS version of the site when possible. DuckDuckGo Essentials also primarily works based on a filter list. It looks at the trackers found on a web page, assigns them a category such as Advertising or Analytics. It is interesting that DuckDuckGo essentials does not block domains it considers to be associated with the first-party domain. So when a user visits *cnn.com*, DuckDuckGo does not block trackers such as “*agility.cnn.com*”, which is actually upon execution of a *nslookup* command is an alias for “*turner.edge.nc0.co*”. Even more concerning, DuckDuckGo does not block sites such as “*ib.adnxs.com*” which it classifies as “Analytics” (*DuckDuckGo Privacy Essentials*, n.d.).
10. **NoScript:** NoScript was determined by multiple studies (Wills & Uzunoglu, 2016) (Mazel, 2019) to be a particularly effective tool for blocking tracking and cookies. It was also determined to be the least intuitive tool due to the number of website elements that it breaks. This is due to the nature of the extension, unlike other extensions on the list, NoScript does not work based on a filter list. The extension works indiscriminately to block JavaScript hence providing the best protection, but also breaking the most sites (Mazel, 2019).
11. **HTTPS Everywhere:** HTTPS everywhere is a collaboration project between EFF and the Tor project that forces the browser to use HTTPS to encrypt its communication when possible. It is an extension for Chrome and Firefox and is automatically included in Tor and Brave. Many websites have inconsistent rules and policies regarding support for HTTPS over HTTP (*HTTPS Everywhere*, n.d.). HTTPS everywhere has two modes: regular and strict. Regular mode forces pages to use HTTPS when possible, but allows HTTP traffic. Strict mode blocks access to any page that does not use HTTPS. Even though HTTPS Everywhere is not an ad blocking or cookie blocking extension, we included it in our list due to its relative popularity in order to assess whether it provides any enhanced privacy, in addition to encrypting the connection.

4.2.2 Evaluation

For this research question, we analyzed the most popular privacy-oriented browser extensions listed above. We tested all of the extensions on Chrome and Firefox by downloading the extension files and loading them into our Python code, with Fiddler acting as a proxy, to obtain traffic information just like in the first research question in order to see which extensions provide the most privacy protection.

We used the same code and Fiddler set up that we created in the previous step, but instead of testing browsers we tested extensions with the browsers. The first step was extracting CRE files from the different extensions to load it into our code. In order to extract the extensions from Chrome and Firefox, we were able to use two quick tools to download them from the respective extension stores.

We also found out that there are only two major types of extension formats, CRX for Chromium based browsers and XPI for Firefox based browsers. Although we tested all of the extensions only on Chrome and Firefox, we would be able to easily test any of the extensions on any of the other listed browsers as long as they are a derivative of Chromium or Firefox.

The site <https://crxextractor.com> was used to extract the CRX files from the Chrome Extension webstore by taking in the URL of the CRX file. To our knowledge, no such site exists for the XPI files (Firefox extensions), so we were able to use the command ‘wget’ to extract the Firefox extensions by visiting Firefox add ons store and copying the link from the “Add to Firefox” button using the following arguments.

```
wget  
https://addons.mozilla.org/firefox/downloads/file/3719726/privacy\_badger-2021.2.2-an+fx.xpi -o PrivacyBadger.xpi
```

Once we had all of the extensions that we needed for both Firefox and Chrome, we then added them to the Selenium driver. For Chromium-based browsers, we only had to add in the “add_extension” function to our options in our code setup of the browser. For Mozilla, the alternative function was the “install_addon” method.

4.3 Website Degradation

Finally, we studied how these privacy tools compare not only in terms of privacy protection, but also ensuring minimal website degradation. We want to understand the amount of degradation which results from the use of each of these tools.

4.3.1 Evaluation

First, to determine the level of webpage degradation we picked out tools (browsers and extensions) from our testing in the previous sections that showed the most promise in terms of privacy protection and then drove through the same set of top 100 Alexa sites with those tools

and captured screenshots. We captured the screenshots using the same Python code we developed, and we took advantage of the screenshot feature method part of the Selenium library which is shown in Figure 4.6.

```
driver.save_screenshot("BreakageImages/" + "Chrome" +  
str(input[1]) + site.split(".")[1] + ".png")
```

Figure 4.6. Line of code responsible for capturing screenshots during web testing.

Then we manually analyzed the screenshots and assigned a value to them ranging from no webpage degradation, minor web page degradation such as soft warnings for ad blockers that a user can close or minor cosmetic changes and major webpage degradation such as hard stop warnings for ad blockers cannot be closed, webpage not displaying at all or not correctly, being unable to access elements of the page. Examples of minor and major degradation as presented in [Appendix A](#).

4.4 Summary

In order to answer the first three research questions, we configured Fiddler to capture HTTP and HTTPS traffic on a Windows machine, while browsing through Alexa Top 100 sites. We then analyzed the third-party domains that we saw and assigned a category to them based on the Ghostery extension filter list. The categories helped us establish whether the domains were essential to the site functionality or used for advertising and analytics purposes. The amount of advertising and analytics domains that each browser or extension reduced were the best performing tools. We also assessed the webpage degradation caused by each of the tools we tested by capturing screenshots of each of the 100 pages we visited and manually assigning a degradation rating. The following Section presents the results for our testing to answer the research questions.

5. Results for Browsers and Extensions Privacy Protections

Initially we tested the Alexa Top 50 sites based on September 2020 rankings and conducted testing in September and October of 2020. Once we fully developed our methodology and testing we extended our testing to include additional 50 sites that were not tested initially based on December 2020 rankings and conducted the testing in January 2021, so overall we tested all of the combinations against 100 sites. In order to represent our data testing as one unanimous dataset, we combined our results from September/October 2020 testing and January 2021 testing to represent them as comprehensive results for 100 sites. The results for browsers, extensions and degradation are represented below. The results for mobile and search engines are represented in later chapters.

5.1 Counting Third Parties Per First-party Site

In order to assess the performance of browsers, extensions and mobile applications in terms of privacy improvement we needed a consistent metric we would use. We initially began by summing up the number of third-party domains that appeared when we visited all 100 sites with each tool. We later realized that there were flaws with this metric and attempted to count the number of third parties that appear in our Fiddler files per each individual site. To accomplish that we needed to look inside our Fiddler files for each browser, extensions and mobile application and look between visited sites. For example, the first three sites we visited during our 100 site testing were google.com, youtube.com and amazon.com respectively. In order to find out how many third parties appeared when we visited google.com, we needed to look only at the section of the Fiddler traffic between google.com and youtube.com. To find out what third parties appear on youtube.com, we needed to look between youtube.com and amazon.com. The following is the array that was created that split up each fiddler file into 100 pieces based on the relevant visited site and calculate how many third parties appeared in each.

```

with open('FinalUniqueDomains.txt', 'r') as r:
    with open('Next50\\Domains_per_site\\browsers\\firefoxStrict.txt', 'w') as
t:
    lines = r.readlines()
    # For each host that we have
    for line in lines:
        line = line.split('\n')[0]
        domainArr[0] = line
    with open('Next50\\CSVs\\Browsers\\firefoxStrict.csv', 'r') as r:
        with open('topSites2.txt', 'r') as f:
            reader = csv.reader(r)
            readerTwo = csv.reader(r)
            siteLines = f.readlines()
            for site in siteLines:

domains.append(site.split('\n')[0].split('/')[1].lower())
                for lines in reader:
                    lineBuild = []
                    lineBuild.append(lines[0])
                    lineBuild.append(lines[3])
                    lineBuild.append(lines[4])
                    if(lineBuild[2] == '/' and 'www.target.com' in
lines[3]):
                        start = lineBuild[0]
                        # print(start)
                        if(lineBuild[2] == '/' and lineBuild[1] in
domains):
                            count = count + 1
                            arrInside = []
                            ArrOne.append(arrInside)

                f.close()
                r.close()
                bigAssArr = helper(start, domains, ArrOne)
                #count by
                for i in range(len(bigAssArr)):
                    #see if domainArr is inside
                    for j in range(len(bigAssArr[i])):
                        if((Arr[i][j][1].split('.')[2]+ '.' +
bigAssArr[i][j][1].split('.')[1]) in domainArr):
                            index =
domainArr.index((bigAssArr[i][j][1].split('.')[2]+ '.' +
bigAssArr[i][j][1].split('.')[1]))
                            domainArrCount[index] = domainArrCount[index] + 1
                            break
                t.write(str(domainArrCount[0]) + '\n')
                domainArrCount[0] = 0
                ArrOne = []

```

Figure 4.7. Code used to calculate the average number of third parties per firsty party.

```

def helper(start, domains, ArrOne):
    with open('Next50\\CSVs\\Browsers\\firefoxStrict.csv', 'r') as r:
        with open('topSites2.txt', 'r') as f:
            reader = csv.reader(r)
            count = 0
            for row in reader:
                lineBuild = []
                lineBuild.append(row[0])
                lineBuild.append(row[3])
                lineBuild.append(row[4])
                # print(lineBuild[2], start, lineBuild[1] in domains,
lineBuild[1]) FF opera and chrome
                if(lineBuild[2] == '/' and lineBuild[1] in domains):
                    count = count + 1
                if(lineBuild[2] != '/' and lineBuild[0] != '#'):
                    if(int(lineBuild[0]) > int(start)):
                        ArrOne[count - 1].append(lineBuild)
            return(ArrOne)

```

Figure 4.8. Code for the helper function used.

5.2 Browser Results

As described in Chapter 3, our first research question is to understand the effectiveness of browsers in their out-of-the-box state and assess whether extra features like “Block Third-Party” cookies on Chrome or Firefox “Strict” mode protect user privacy. The following section presents the result to answer previously mentioned research question.

In order to establish which out-of-the-box browsers provide better privacy protections than others, we examined nine browsers as described in Section 4.1. To compare different browsers we calculated the average number of third parties per visited site. Using the code we wrote in Chapter 5.1, we analyzed each of the Fiddler traffic logs and found the total number of third parties present in each visited site.

Figure 5.1 shows that Vivaldi performed the worst out of all the browsers, including Chrome which was used as a baseline. Vivaldi’s poor performance is especially interesting because as pointed out in Chapter 4.1.1, Vivaldi claims to be “fast, private and secure browser that blocks ads and trackers” (*Vivaldi Browser*; n.d.). As can be seen from Figure 5.1 Vivaldi does not appear to block any ads or trackers, and potentially allows extra trackers through Chrome without any protections configured. Opera performed slightly better than Chrome. We expected to find results like these since Opera was sold to a Chinese consortium and was rumored to gather data on the users. Firefox and Microsoft Edge performed better than Chrome

and were close to each other. These results were not expected by our group, given that Firefox has long been considered somewhat of a privacy browser, while Microsoft Edge is a relatively new player and has not made any claims regarding privacy of their users. Epic and Iridium are next in terms of performance and they are close together in terms of privacy protection. They perform better than Firefox and Microsoft Edge, given that they do claim to have some privacy enhancing features. Brave performed the best out all the browsers tested. Brave does advertise themselves as a tracker and advertising blocking browser and it matches with our results (Brave Browser).

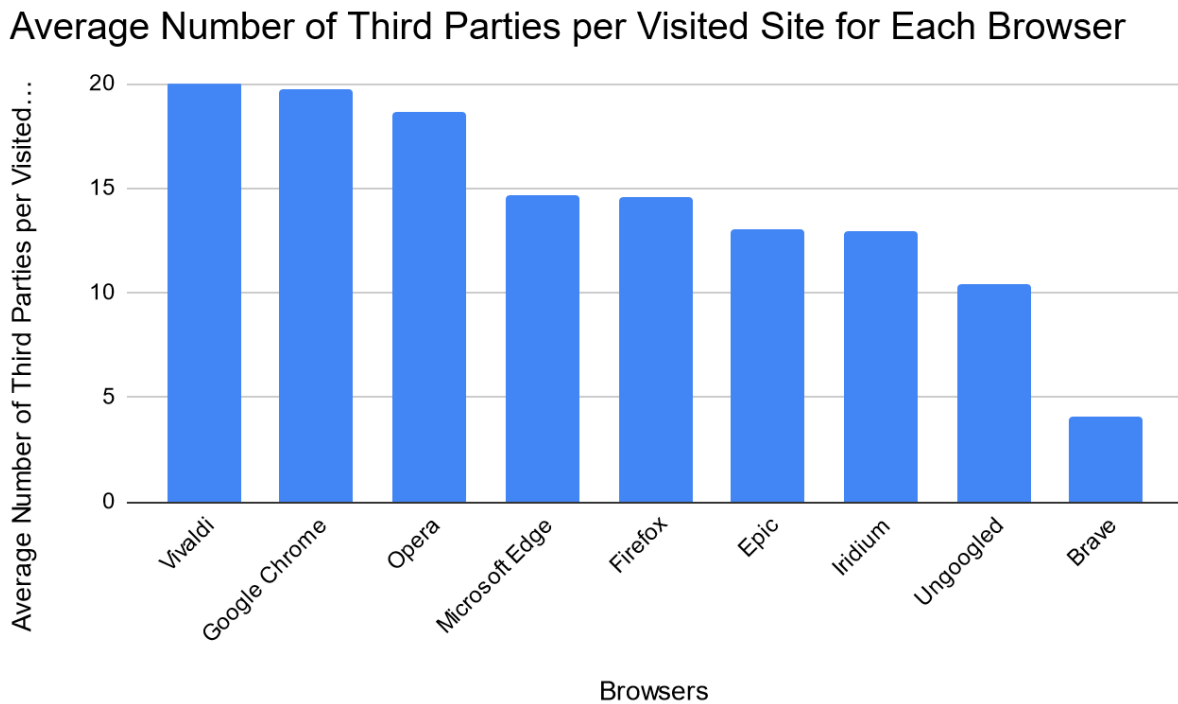


Figure 5.1. Average Number of Third Parties per Visited Site for Each Browser

We were also interested in how effective different browsers are at blocking certain categories of trackers. As discussed in Chapter 2, not all third-party cookies are meant to track the user. Some like the comments and audio/video third-party add ons can be useful to the sites. Figure 5.2 shows the same data presented in Figure 5.1, but third-party trackers down into categories of different trackers. As can be seen from the graph, advertising was the most popular category of third parties. Examining the performance of different browsers across categories is important, because while Vivaldi and Chrome perform better than Opera, Opera allows more advertising.

Average Number of Third Parties per Visited Site for Each Browser by Category

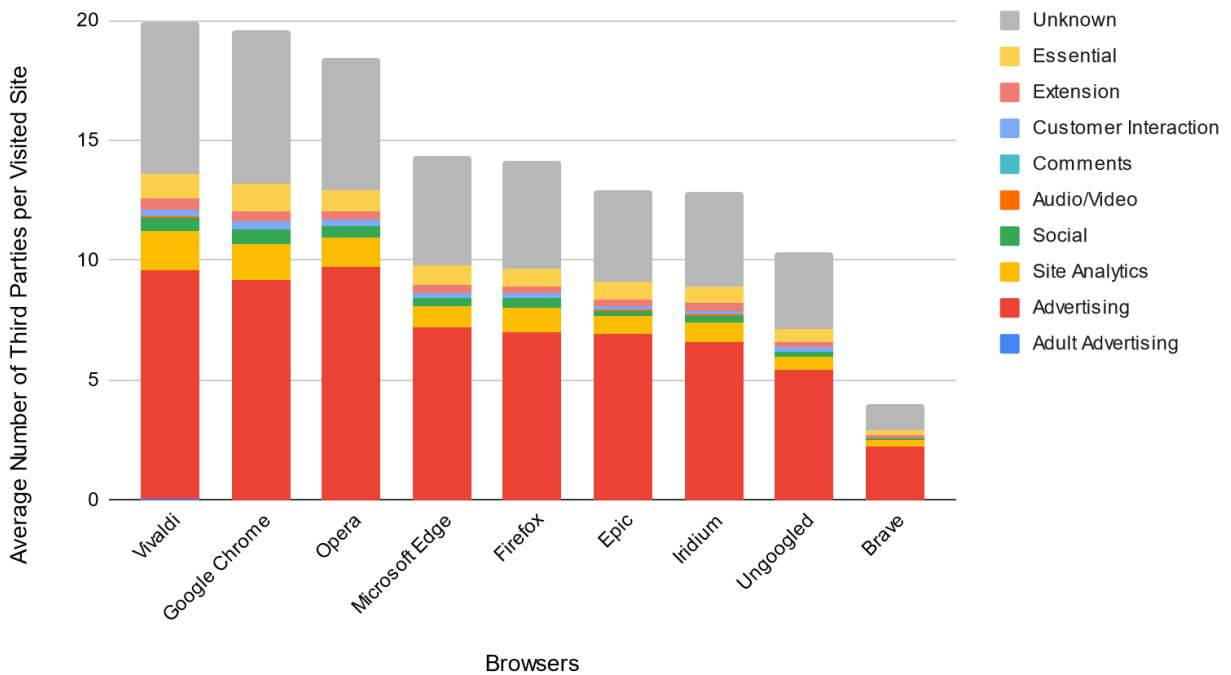


Figure 5.2. Average Number of Third Parties per Visited Site for Each Browser by Category.

In order to assess how good different browsers were at blocking trackers we studied three categories from Ghostery. The specific categories we examined were Advertising, Adult Advertising, and Site Analytics. We summed up the average number of third parties per first-party with those three categories. We turned the average into a percentage metric as compared to Google Chrome which was used as the baseline throughout our study. For example, Chrome was 100% in terms of trackers remaining as can be seen in Figure 5.3. As shown in Figure 5.3, the results were consistent with the results from Figure 5.1 and the order of how well browsers performed stayed the same. Figure 5.3 is simply used to represent the performance specifically in terms of trackers.

Percent Trackers Remaining for Each Browser

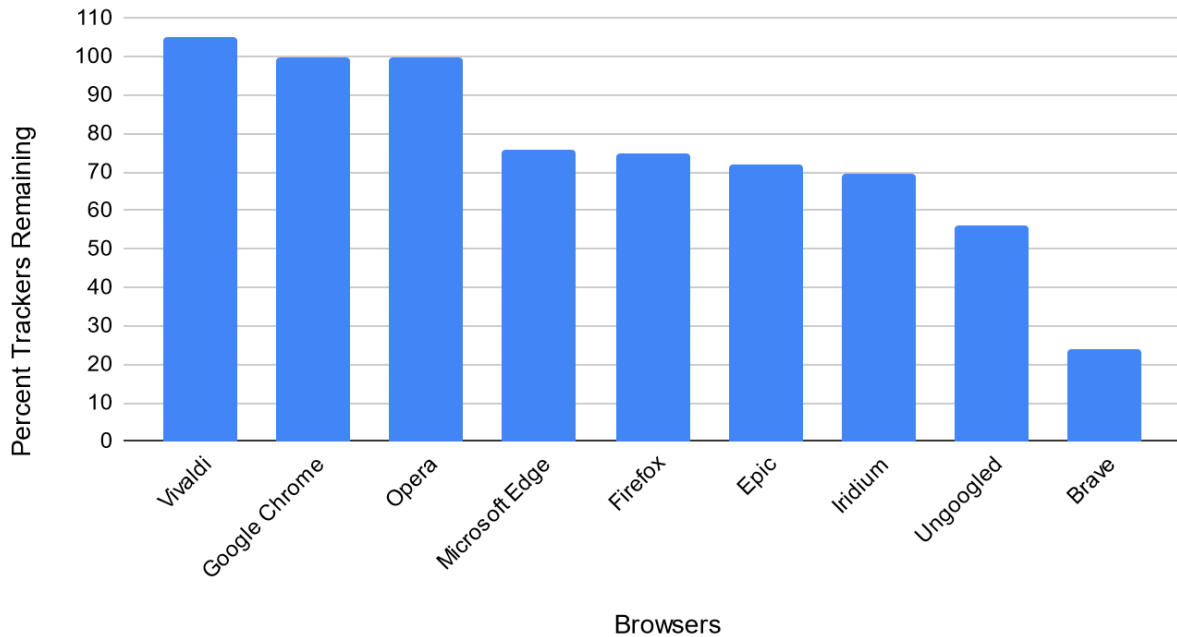


Figure 5.3. Percent Trackers Remaining for Each Browser

During our research, our group also expanded on our original question that was aimed to study browsers in their out-of-the-box state and studied some of the built in privacy features that the browsers had to offer. We looked at Google Chrome’s “Block third-party” feature as well as Firefox’s “Strict” and “Custom” modes in order to assess how much privacy protection they provide. As shown in Figure 5.4, Chrome’s “Block third-party” feature provided only small protection and performed worse than Firefox in it’s out-of-the-box state. Firefox’s features with “Strict” and “Custom” mode offered significant tracker reduction. Firefox’s Custom mode is supposed to be stricter than the Strict mode, so our group expected the Strict mode to offer less protection than the Custom mod. Both modes were close together and Strict mode offered 1% more tracker reduction than Custom Mode. Given how well Firefox’s built in features performed,

it is worth recommending them to the users.

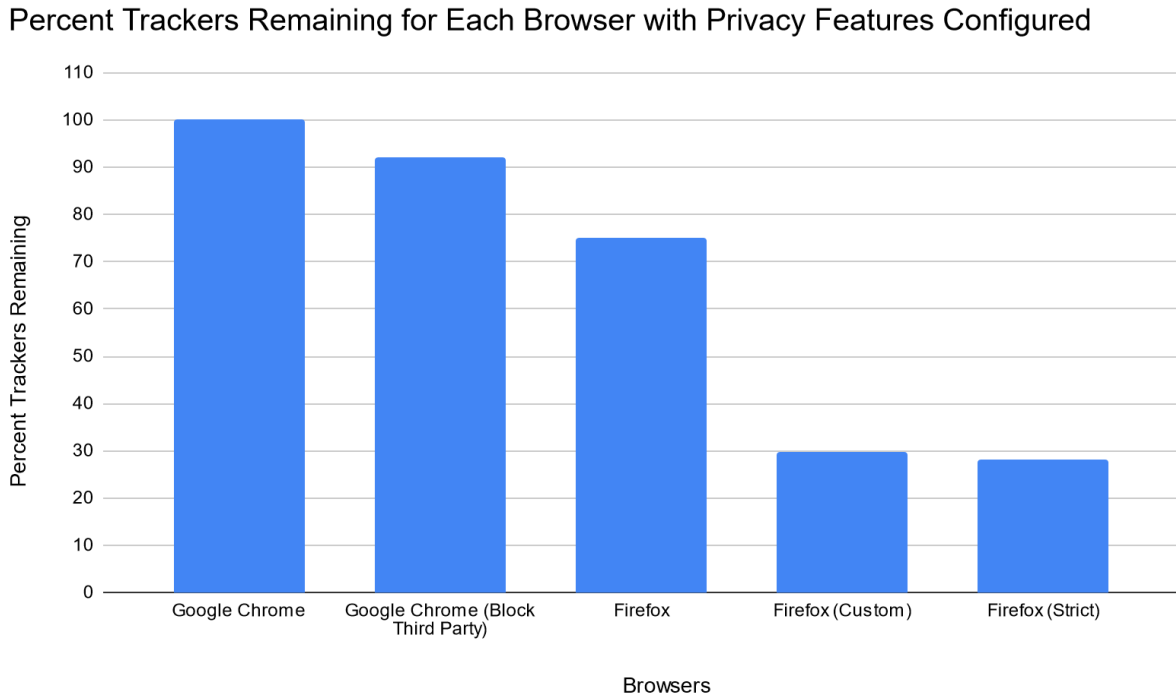


Figure 5.4. Percent Trackers Remaining for Each Browser with Privacy Features Configured.

5.3 Extensions Results

The next research question we sought to answer was regarding the effectiveness of browser extensions in protecting user privacy and which privacy tools - ad blockers, list based or heuristic - offer the best protection.

In order to establish which extensions provide better privacy protections than others, we examined thirteen browser extensions shown in Figure 5.5. We tested the extensions both on Chrome and Firefox, but two of the extensions - MyTrackingChoices and uBlock - were not available on Firefox, so they were only tested on Chrome.

Figure 5.5 shows performance on the different extensions both on Chrome and Firefox in terms of the average number of third parties per visited site just like in Chapter 5.2. HTTPS Everywhere offered the least protection given that it is not aimed to be a tracker or advertising reducing extension. MyTrackingChoices was next in terms of performance and although it was designed to be a tool similar to Privacy Badger, it performed significantly worse. Our group speculates that this might be due to the fact that the extension has different possible configurations that can reduce the allowed number of third parties, which the developer expects the user to take advantage of. Adblock, Blur and Adblock Plus performed the next best. Adblock and Adblock Plus were not designed as tracking reducing extensions, but are rather aimed at blocking intrusive ads, rather than all ads. The poor performance of the extensions is

likely attributed to their “acceptable ads program” discussed in Chapter 3 that allows a lot of advertising through. Blur’s not great performance is more surprising especially when the Wills, 2016 research is considered. The extension did become a lot less popular over the last five years as can be seen from Table 4.2, which indicates popularity numbers for each extension. AdGuard is next in terms of performance. AdGuard has an aggressive filter list, but it is still an ad blocking extension and not an all around privacy solution. It makes sense that it performs better than Adblock and Adblock Plus, but worse than extensions like Privacy Badger and Ghostery. AdGuard claims to use over 80 different ad blocking lists, including many that uBlock and uBlock Origin use so the difference in extension’s performance is not explained. Privacy Badger provides moderate protection, however it is meant to be trained. The extension is programmed to block trackers it sees more than three times and if the extension is tested on the same set of 100 sites more than once, the performance is much more improved. Ghostery’s great performance is not surprising, it has been determined to be the best extension by multiple previous studies (Merzdovnik, 2017) (Mazel, 2019) (Uzunoglu, 2016). The fact that DuckDuckGo and Disconnect are so close to Ghostery is promising and somewhat surprising. Although they also claim to be all around privacy extensions and both operate off a filter list, which grants them good performance. Good performance of uBlock and uBlock origin is interesting, especially that they perform better than Ghostery does. uBlock and uBlock Origin are meant to be only an ad blocking extension, however, based on their performance, users can use it both as an ad blocking and a privacy-enhancing extension. Lastly, the best performance out of all the extensions is NoScript, not surprising and consistent with the previous studies, usually not recommended given the amount of website degradation it causes as discussed in Chapter 5.4.

Average Number of Third Parties per Visited Site for Each Extension

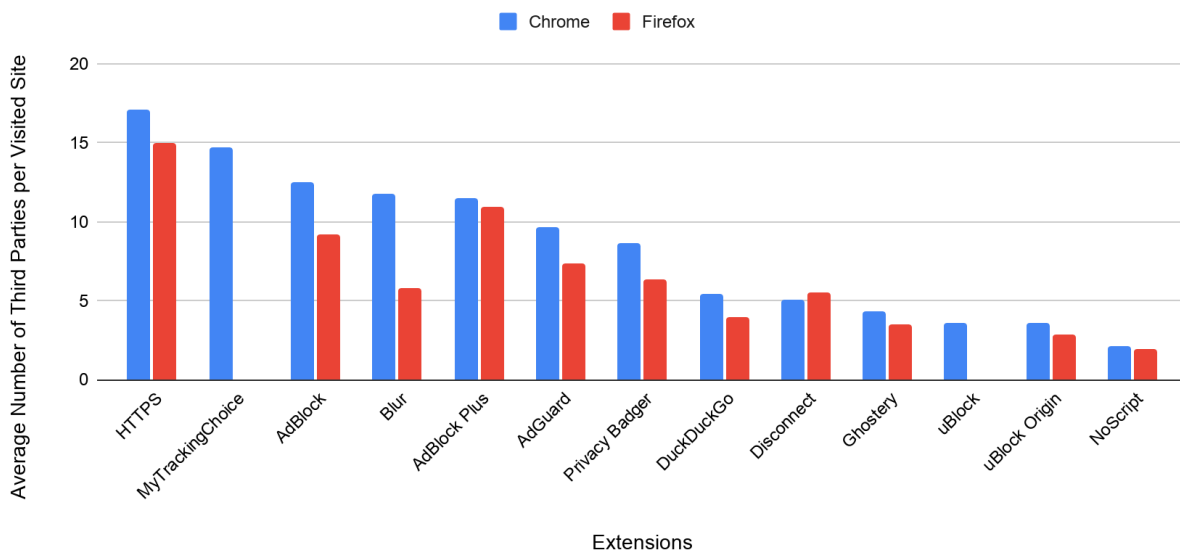


Figure 5.5. Average Number of Third Parties per Visited Site for Each Extension.

Figures 5.6 and 5.7 represent the same information about third-party trackers as Figure 5.5, but each third-party is colored-coded by category. The performance of the extensions remains consistent with Figure 5.5. It is helpful to see the graphs broken up given that Firefox in out-of-the-box mod provides privacy protections so the extensions that go into Firefox are working with existing Firefox blocking figures.

Average Number of Third Parties per Visited Site for Each Chrome Extension by Category

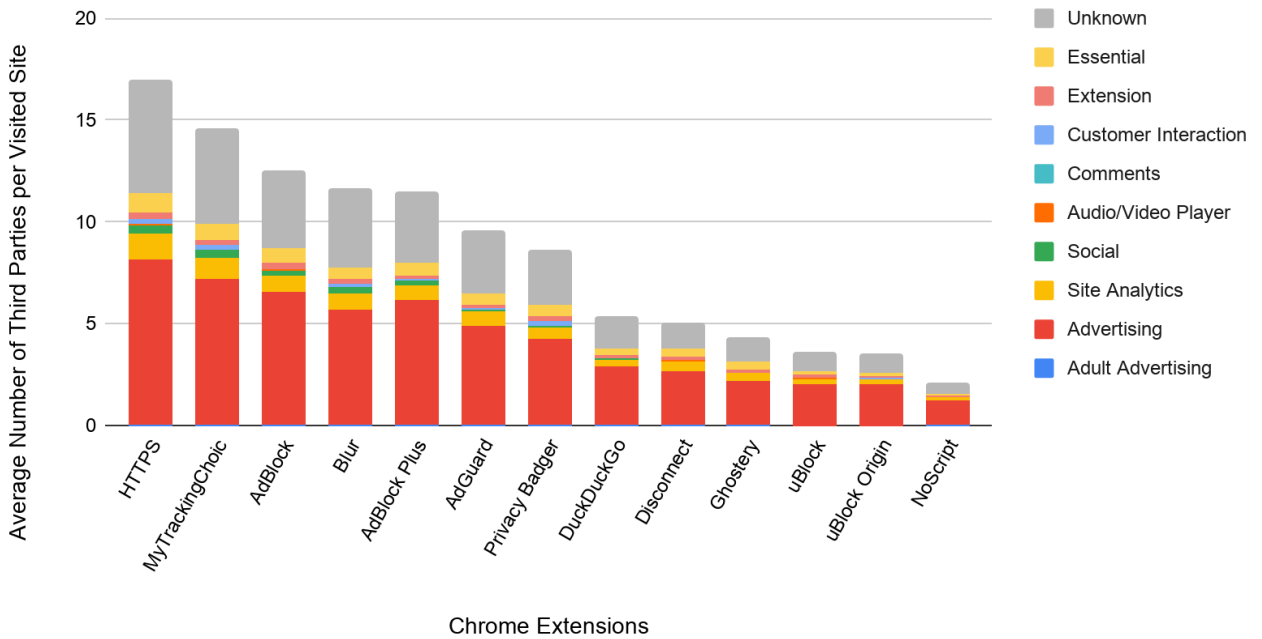


Figure 5.6. The average number of third parties per visited site for Each Chrome Extension by Category.

Average Number of Third Parties per Visited Site for Each Firefox Extension by Category

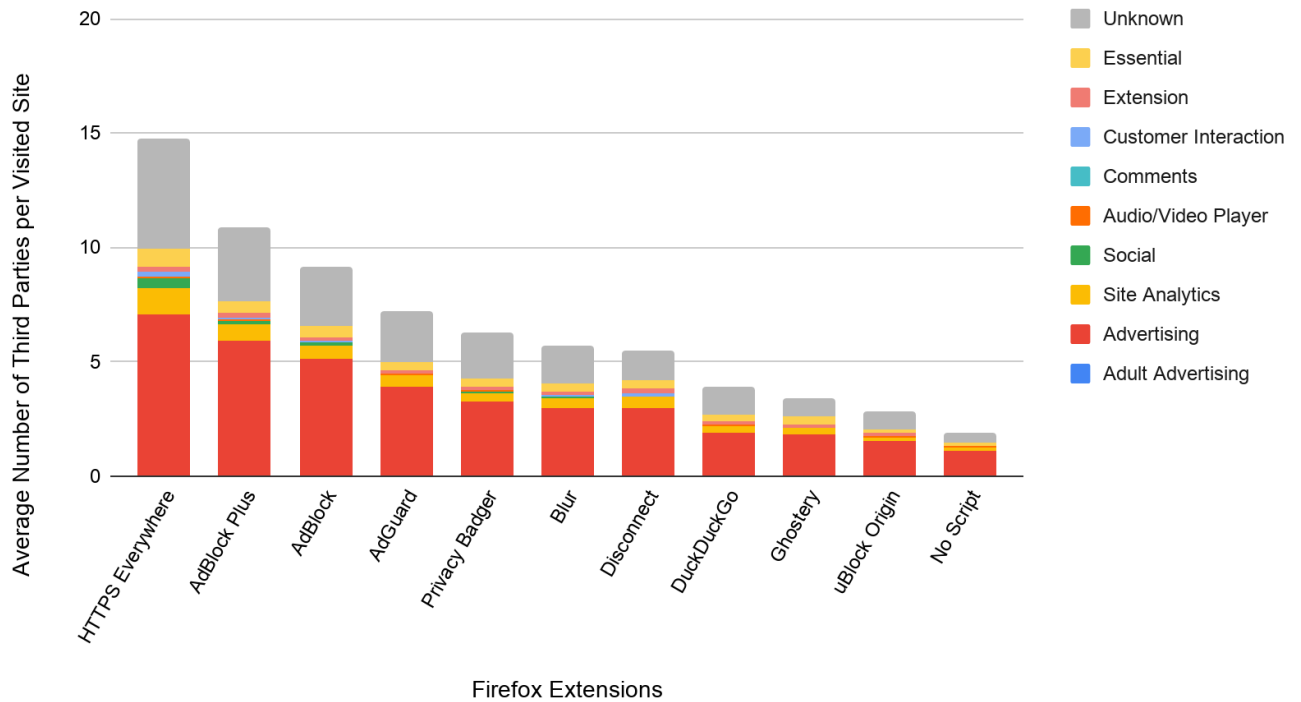


Figure 5.7. The average number of third parties per visited site for Each Firefox Extension by Category.

In order to assess how good different extensions were at blocking the categories of third parties that we were concerned with, which were Advertising, Adult Advertising, and Site Analytics, we repeated the same process from Chapter 5.2. We summed up the average number of third parties per first-party with those three categories and turned it into a percentage metric as compared to Google Chrome with no extensions which was used as the baseline throughout our study.

The calculation was as follows:

$$\text{Number of advertising, adult advertising and analytics third parties per first-party} / \text{Number of advertising, adult advertising and analytics third parties per first-party (on Chrome)} * 100\%$$

As shown in Figure 5.8, the results were consistent with the results from previous figures and the order of how well extensions performed stayed the same. Figure 5.8 is simply used to represent the performance specifically in terms of trackers.

Percent Trackers Remaining vs. Extensions

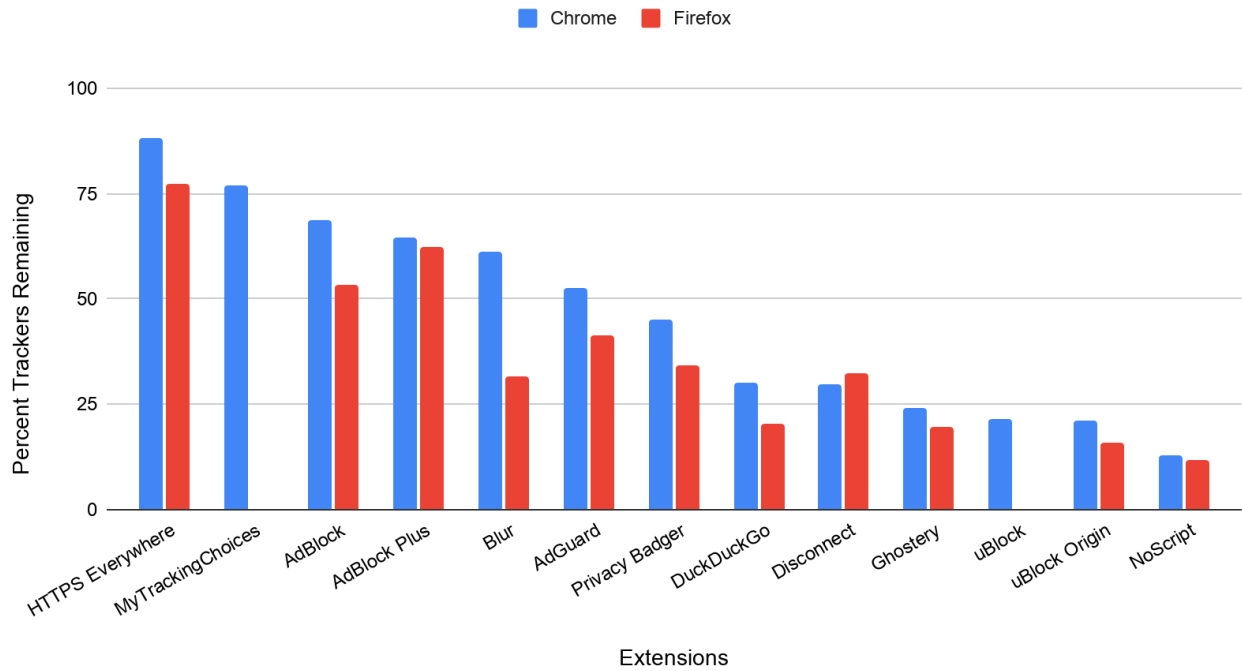


Figure 5.8. Percent Trackers Remaining for Each Extension.

Lastly, the ultimate goal was to plot the performance of browsers and extensions against website degradation which will be discussed in Chapter 5.4 so we needed a metric for an average between Chrome and Firefox extensions. We took the numbers from Figure 5.8 which shows the performance of each extension on Chrome and Firefox, averaged the two numbers and represented the average in Figure 5.9.

Percent Trackers Remaining Average on Chrome and Firefox for Each Extension

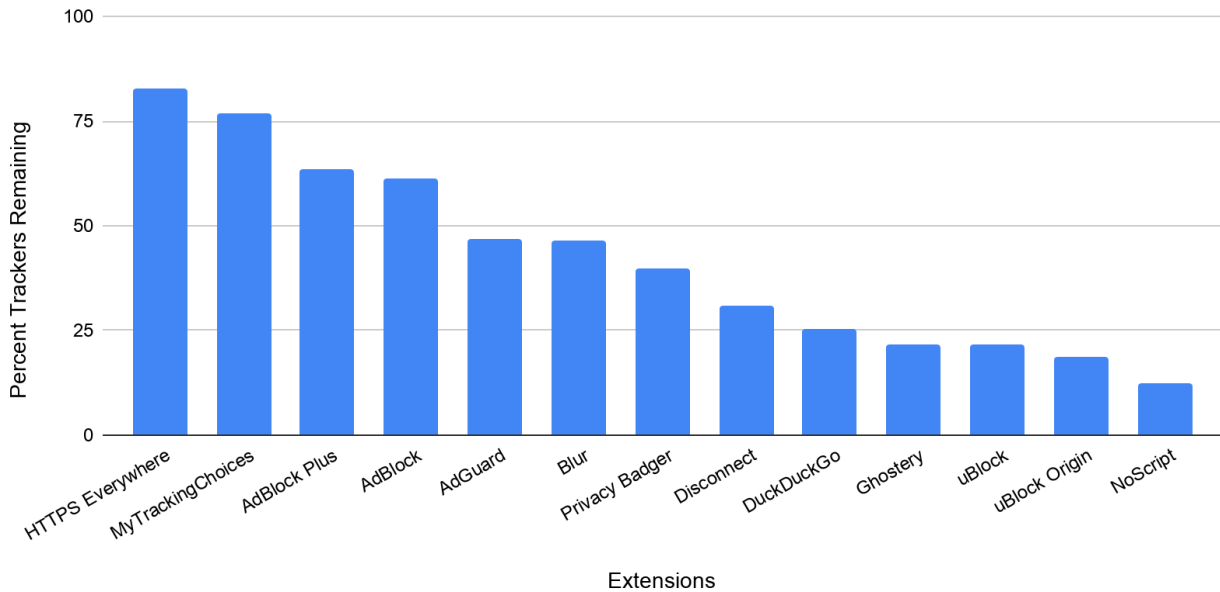


Figure 5.9. Percent Trackers Remaining Average on Chrome and Firefox for Each Extension.

5.4 Website Degradation Results

In order to answer our research question from Section 3.3, after conducting studies on the browsers and extensions, we also assessed the website degradation caused by the different tools by driving through the same set of Alexa 100 sites, capturing screenshots of the webpages and manually assigning them a degradation rating to plot the tools privacy improvement versus the website degradation and recommended the best tool. Below are the numeric results representing website degradation we experienced while using each privacy tool.

For both browsers and extensions, we assessed the website degradation that the tool caused as discussed in Chapter 4.3. Figure 5.10 and Figure 5.11 show the website degradation for browsers and extensions.

Out of all the browsers, Firefox in Custom mode caused the most degradation. This finding is not surprising given that in Custom mode our group blocked all the elements, including ones that might be essential.

Percent of Websites with Degradation for Each Browser by Severity

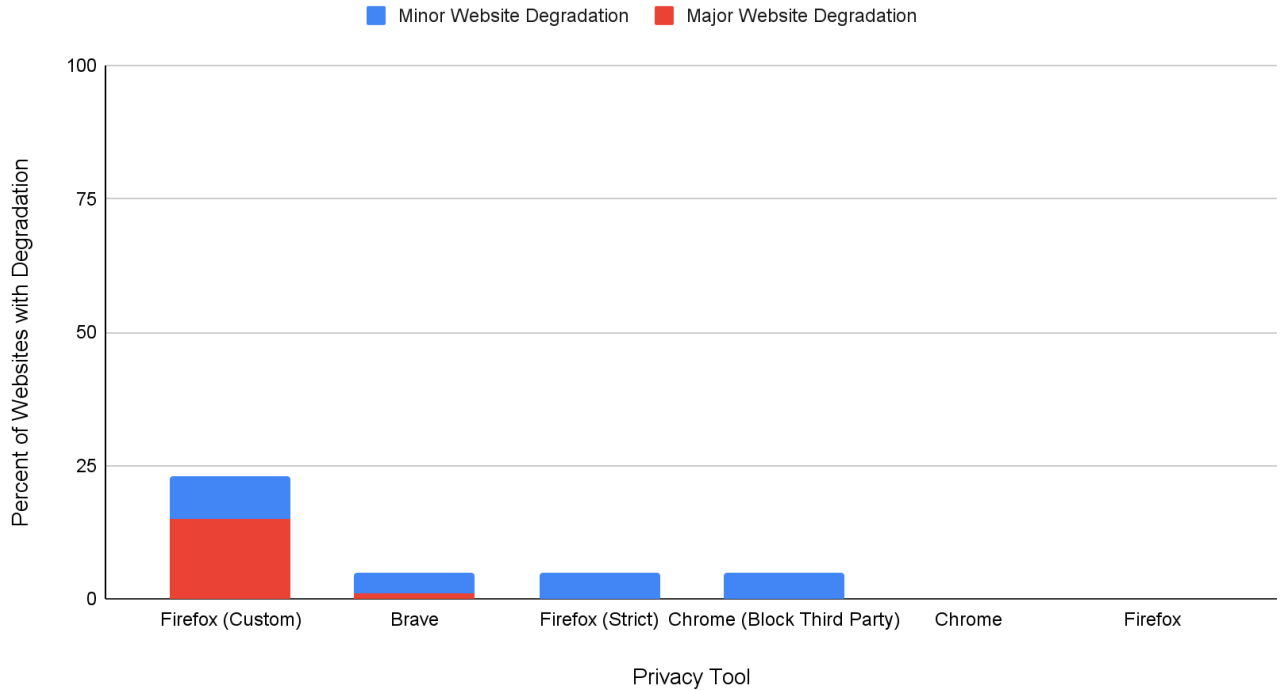


Figure 5.10. Percent of Websites with Degradation For Each Browser by Severity.

Out of all the extensions, unsurprisingly, NoScript caused the most degradation given that the majority of the websites require JavaScript to function and NoScript does not allow it. Adblock Plus and Adblock were the only extensions besides NoScript that caused major website degradation because they allowed variations of “Please disable Adblock” message to show through and the user was unable to exit, unlike with other tools like AdGuard and uBlock Origin.

Percent of Websites with Degradation for Each Extension by Severity

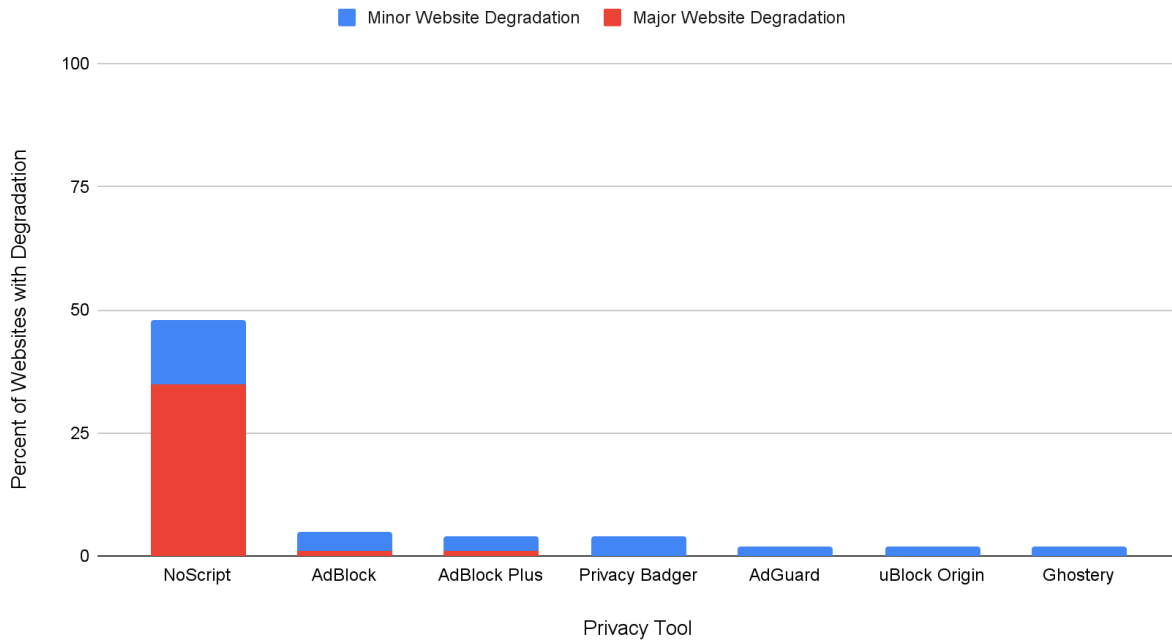


Figure 5.11. Percent of Websites with degradation for Each Extension by severity.

5.5 Website Degradation vs. Privacy Improvement

Finally, in order to make a comprehensive recommendation in terms of the best tool to recommend in our survey, we plotted the percent trackers remaining for browsers and extensions versus the amount of degradation caused. We were interested in finding browsers and extensions that had the lowest percent of trackers remaining and a lowest number of websites with degradation. As shown in Figure 5.12, uBlock Origin, Ghostery, Brave and Firefox in Strict mode show the most promise and will serve as the most effective privacy recommendations in our survey.

Percent Trackers Remaining vs. Percent of Websites with Degradation

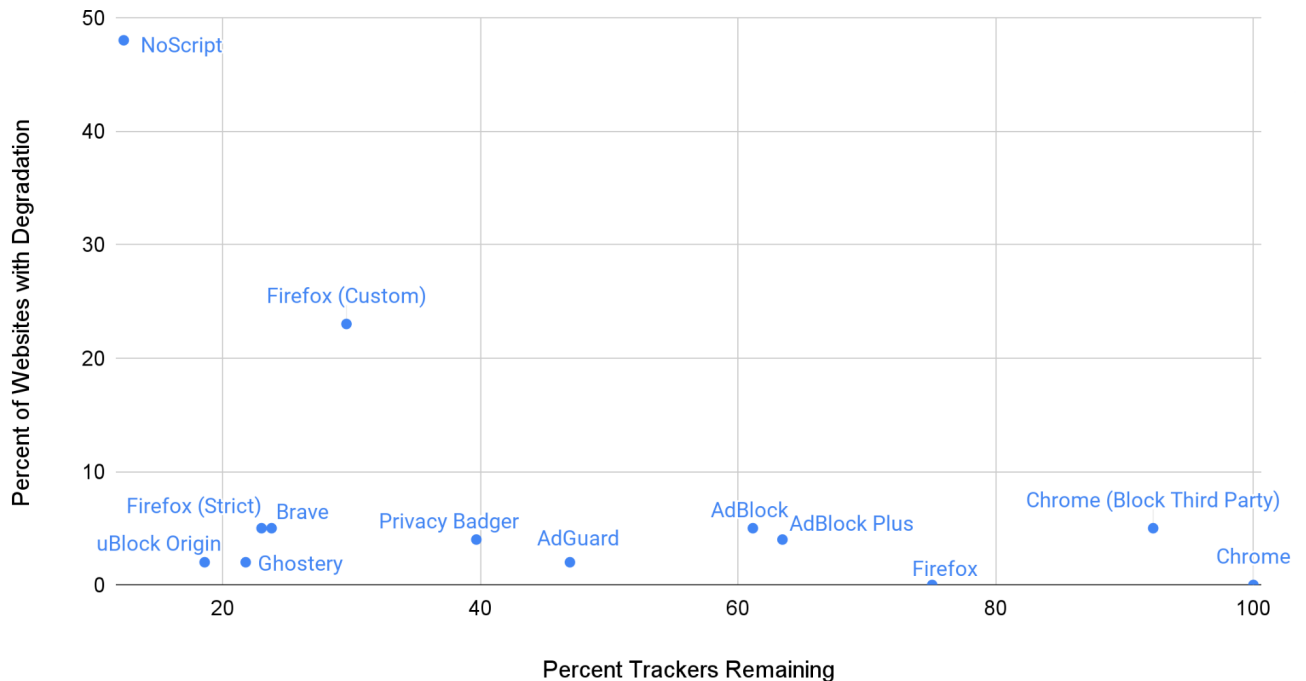


Figure 5.12 Percent Trackers Remaining vs. Percent Websites with Degradation.

5.6 Summary

The best performing browsers were Firefox in Strict mode and Brave out-of-the-box and should be recommended to the users seeking to obtain the best privacy protection. Chrome as established in the Chapter 4 remains the most used browser, so for users who are unwilling to change their default configuration, one potential option is to use Chrome with “Block Third-Party” option enabled which provides limited protections.

We evaluated two types of extensions: adblock and privacy ones. The best performance was shown by uBlock Origin and Ghostery. AdBlock and AdBlock Plus are another option for users who want to support “Acceptable Ads Program” because they provide moderate protection, while allowing through some ads supporting web revenue.

Extensions can be added on to browsers, which could provide added layers of protection. The best performance can be achieved from combining the best browsers with the best extensions. The balance between performance and website degradation is also important. Most of the tools do not cause major degradation, so that should not be too much of a factor when evaluating the best performing tools.

6. Mobile Applications Privacy Protections

In order to answer research question number four, regarding which mobile applications exist on the market that can help protect the users and how effective they are, we developed the following methodology.

The original goal for this research question was to automate the testing for the mobile devices using the webdriver Python code, similar to the method we used for browsers and extensions. Given the technical issues we experienced, we ended up performing manual testing on the smaller subset of sites, only the first Alexa top 50 sites and scanned mobile traffic setting up a proxy in Android Wi-Fi settings and routing the traffic to the Fiddler Classic Desktop Application.

6.1 Popular Mobile Privacy Applications

Originally, we started our research by identifying the most popular privacy enhancing applications for mobile. We identified several popular tools listed in Table 6.1 that we planned to test. All of the applications essentially are built on either Chrome for Android, Firefox for Android or Samsung Internet for Android source code. For example, an application like Ghostery does not add itself as an extension to an existing application, but rather installs its own application. That application however is based on the source code of another popular web browser, in Ghostery's case it is based on Firefox Fenix.

Table 6.1. List of the popular mobile applications tested in the project

Privacy Tool for Mobile	Downloads on Play Store	Ratings on Play Store⁵	Ratings on App Store⁶	Version
Google Chrome	5,000,000,000	24,886,115	88,900	88.0.4324.152
Firefox	100,000,000	3,641,534	13,600	85.1.3
Firefox Focus Browser	5,000,000	61,269	36,700	8.12.0
Brave Browser	10,000,000	238,508	61,800	1.20.103
Adblock Browser App ⁷	10,000,000	145,766	-	2.4.0
AdBlock for Samsung Internet by BetaFish ⁸	5,000,000+	13,406	1,300	2.5.0
AdGuard: Content Blocker for Samsung and Yandex	5,000,000	50,448	8,400	2.6.2
DuckDuckGo Privacy Browser	10,000,000	647,858	281,800	5.76.1
Ghostery	1,000,000	16,222	661	-
Disconnect for Samsung Internet Browser	1,000,000	7,208	-	2020.4

6.2 Methodology

The original goal for this research question was to automate the testing for the mobile devices using the webdriver Python code, similar to the method we used for browsers and extensions.

A linux terminal application exists for free in the Google Play store. Our idea was to use this app to download selenium and start up our webdriving on an Android device. During our testing, we realized that this would not be possible. For Selenium to be able to start its webdriving, it needs to open up its own server to start up a specific browser and give it directions. Essentially, there needs to be a server that gets started up for the browser to talk back and forth from. In addition, our code was designed for a Windows system and it would not run on an Android device.

⁵ As of October 2020

⁶ As of October 2020

⁷ This is the mobile equivalent of AdBlock Plus, even though it is called AdBlock. They do also make “Adblock Plus for Samsung Internet - Browse safe” but it is the less popular product of the two.

⁸ This is the mobile equivalent of the AdBlock extension.

Given that we experienced all of the issues above, we decided to perform manual testing on the smaller subset of sites, only the first Alexa top 50 sites and scanned mobile traffic setting up a proxy in Android Wi-Fi settings and routing the traffic to the Fiddler Classic Desktop Application.

6.2.1 Fiddler Proxy on Mobile

We began by configuring a Fiddler Proxy on an Android mobile device. The detailed steps with pictures can be found in the two articles that we referenced to find out the process for setting up the proxy (Velikov, 2019) (*Configure Fiddler for Android*, n.d.), but an overview of the steps can be found below:

1. Go to Fiddler Classic → Tools → Fiddler Options → Connections and check “Allow remote computers to connect”.
2. Note the port number listed under “Fiddler listens on port”. The default value is 8888.
3. Restart Fiddler.
4. Hover over the online indicator at the top right corner of the Fiddler application to find out the IP address of the Fiddler server. (This IP address will be used in the next steps on mobile as a proxy address).
5. On an Android device, open Wifi settings. (This varies from device to device, but generally can be found under Settings → Wifi).
6. Tap and hold on the current network to show network details and find advanced options.
7. Select “Manual” from the Proxy list.
8. Type in the IP address from step 4 and the port number from step 2 (the default is 8888).
9. Click Save to Apply changes.
10. In order to be able to capture and decrypt HTTPS traffic, install a security certificate from Fiddler. In any browser, go to <http://ipv4.fiddler:8888> and download the Fiddler certificate. Install the certificate to the device.

6.2.2 Certificate Pinning

The steps described above worked for extensions based on Chrome and Samsung source code, such as Chrome itself, Brave and Adblock and Samsung Internet and extensions that work for Samsung Internet. However, once our group tried routing the traffic from Firefox based applications, such as Firefox itself, Ghostery and DuckDuckGo, we ran into untrusted site problems.

Applications even a few years ago used to ignore SSL errors and allowed users to much more easily modify and intercept traffic which was incredibly useful for applications developers, but could be used by the attackers to intercept traffic (man in the middle attack). Most modern applications at minimum check for a valid certificate in order to avoid man-in-the-middle

attacks, which our process essentially is. Generally, this issue is solved by using root SSL certificate, which is why we added Fiddler certificate to our trusted certificates list both on desktop and mobile. By adding a trusted root certificate, the user lets the device and the application know that we accept responsibility for intercepting traffic and we trust the source collecting the data. However, even more modern apps that use higher API versions now commonly perform certificate pinning and if any information in the certificate does not match the web server we are connecting to, the app will fail the connection. While this does provide enhanced security and is of great benefit to the users, those who perform research and manipulate applications, run into problems and are unable to see traffic. There are about a dozen potential workarounds for this issue which are listed briefly below, along with the links to the articles they were sourced from and notes about the process.

6.2.2.1 Editing APK Code

First potential solution is, decompiling the APK code and changing targeted API level (Wass, 2018). Applications targeting versions higher than Android 6.0 and API level 23, do not as easily trust user added certificates by default and prefer to use secure connections through protocols like TLS and HTTPS. This potential workaround changes the main manifest file in the app source code in an attempt to increase the trust level for the user added certificate. It can be done by changing the line in the manifest.xml source code to contain

“platformBuildVersionCode = 23” and “PlatformBuildVersionName = 6.0” in AndroidManifest.xml file as shown in Figure 6.1.

```
<manifest xmlns:android="http://schemas.android.com/apk/res/android"
package="com.test.app" platformBuildVersionCode="23"
platformBuildVersionName="6.0">
```

Figure 6.1. Code change in the header of the AndroidManifest.xml file to lower the APK and API version.

Another solution which builds on the previous idea is decompiling the APK source code and adding a call to user added custom certificates in the source code (Wass, 2018). It can be done again by decompiling the APK and creating or modifying (if it already exists) “network_security_config.xml” file usually located in the “/res/xml/” path. An example of the code that needs to be added is shown in Figure 6.2. A call on line 5 “@raw/my_ca” is the location of the user defined folder where user created certificates can be placed and the application is supposed to trust them.

```
<?xml version="1.0" encoding="utf-8"?>
<network-security-config>
<base-config>
<trust-anchors>
<certificates src="@raw/my_ca"/>
</trust-anchors>
</base-config>
</network-security-config>
```

Figure 6.2 An example of code that needs can be added to define user added certificates.

6.2.2.2 Using an Android Emulator on Desktop

In order to avoid dealing with source code and decompiling APK's is using an emulator such as Genymotion or Android Studio emulator on desktop and capturing traffic on the desktop directly. Genymotion is an attractive option since it is an emulator that allows developers to quickly spin up a virtual simulation of hundreds of Android systems on a variety of devices. Unfortunately, our group ran into problems trying to capture traffic given that since both tools use the concepts of VirtualBox and create their own subnetworks to route traffic through, routing traffic back to the fiddler server presents the same issue as above with a need to create a proxy or a bridge, which gets flagged by the Firefox based applications as another potential man-in-the-middle attack.

6.2.2.3 Frida Hook and Objection

Another method that our group tried, but was unable to successfully execute, is using Frida Hook and Objection which is arguably the most complex method out of those listed (Wass, 2018) (Holding, 2018). If the previous steps do not work, it is most likely that the application is performing a kind of SSL pinning or additional SSL validation (which our group suspects is the case with Firefox). In order to bypass that, developers can hook the application code and interfere with the process itself which can be done using the Frida framework which is often used for mobile penetration testing. Frida is a framework that allows people to tamper with applications' code during runtime and can be done by injecting Frida Gadget into the target APK. The process is rather involved and the two articles references should be consulted in detail. In essence, users will need to extract the APK, insert the dynamic library, edit smali code and repackage the APK which requires installation of the Python 3.7.x, Pip3, Android SDK, apktool (Holding, 2018).

6.2.2.4 Root Android Device to Bypass SSL Pinning

The last two options involve rooting a device and manually adding Fiddler certificate to the root list and using third-party tools to bypass SSL pinning, similar to Frida and Objection.

By default, all downloaded and installed certificates by the user, like the Fiddler certificate, get added to the user certificates folder and there is no option to add the certificate to the rooted folder. However, on a rooted Android device, users can run a file management tool with root privileges and move the certificate from user folder to the root folder.

Rooting a device presents its own set of challenges and has become a lot more of a challenge over the last decade. Below is the process that our group ended up using for rooting Samsung S7 devices.

1. On the mobile device go to Settings → About Device → Software → Build Number, tap on it about 7 times until developer options are enabled.
2. Still on mobile, go to Settings → Developer Options → enabled USB debugging and OEM Unlock.
3. On the PC, go to androidfilehost.com and search for the root files compatible with the version of the device. For our group it was Samsung S7, Oreo Nougat ADB Advanced Root V12 (<https://androidfilehost.com/?fid=11410963190603904440>)
4. Download the file from step 3.
5. Unpack the downloaded file.
6. Open Windows PowerShell in the folder that the files were exported into.
7. Plug in the mobile device into the computer using a USB cable.
8. Run command “adb devices” in the Windows Power shell. Under the list of devices, there should be the device that was just plugged in. If the list is empty, disconnect the device, plug it in again and re-run the command.
9. Run the command “adb reboot download” to force the phone to download mode compatible with adb. The mobile device should reboot and start downloading the pushed file.
10. In the file folder with the image, open “Odin_Firmware” folder and run “Odin_313.exe” or equivalent version.
11. In the Odin Software interface that is opened, click “AP” and specify the path to the image downloaded. Pick the zip folder to download and install on the device. For our group it was a zip archive called “AP_SM_G930_OREO_ENG_BOOT.tar”.
12. Wait until the Odin log says “All threads completed”. The mobile should be rebooted.
13. Using the command line, run “root.bat” file in the original archive. In the terminal window that is titled “Install Oreo System S7 and S7 Edge Root” when prompted to make the decision, enter a number corresponding to the image that will be installed. For our group it was 2 for S7 Oreo Root.
14. Install SuperSU which is a superuser access management tool. Download the application and grant it necessary access.

Once a device has been rooted, users will have access to a number of file management apps available directly on the Play store that do require root privileges, but can move their

downloaded or self-created certificate into the root folder. Alternative, users can also try using Xposed Application which can be downloaded from the Play store and use “disable SSL pinning” add on to perform similar injection as the Frida and Objection attempt to do.

6.2.2.5 Edit Browser Configuration File

Firefox Nightly which is a version of Firefox, also gives users access to the “about:config” page and accessing it is as simple as typing it into the URL of the browser. Standard Firefox and applications in question such as DuckDuckGo and Ghostery do not give users access to the about:config menu on the mobile devices for security reasons, however firefox nightly does. Our group experimented with a variety of options concerning SSL pinning and certificates to try to see if it would make a difference. Ultimately, none of those options were able to disable ongoing SSL verification that Firefox must be performing. A list of the options that we have tried to disable in Firefox Nightly is below:

- security.ssl.enable_ocsp_stapling = false
 - Disable OCSP stapling
- network.stricttransportsecurity.preloadlist ⇒ false
 - To add an exception for self-signed certificates
- security.enterprise_roots.enabled = true
 - To accept signing certificates saved in Windows/Mac certificate store as valid authorities instead of going to Firefox own certificate store
- network.websocket.allowInsecureFromHTTPS = true
 - Allow insecure HTTPS traffic

6.2.2.6 End Result

We found that even after decompiling the APK and downgrading the API version, disabling SSL 2.0 pinning, rooting the device, adding fiddler certificate to the root folder, and disabling many security options from Firefox Nightly in the “about:config” file, we were unable to completely circumvent the issue. Websites like cnn.com do eventually load on the device, they only load partially which defeats the purpose of our research. Given the time constraints of the project and the fact that many of the applications we were going to test are available on the desktop and can be assumed to function based on the same filter lists and heuristics and their mobile counterparts, we did not investigate other ways of avoiding the SSL pinning issue on Firefox. We did however successfully test Google Chrome, AdBlock plus, Disconnect, AdGuard and Brave because those applications were based on Chrome and Samsung Internet source code which does not include as strict of protections against man in the middle as Firefox does. The results of those tests are available in the next section.

6.3 Mobile Results

The following section presents the results to answer question five from Chapter 3, regarding how effective different mobile applications are at protecting user privacy. As described in Section 6.2, given the technical difficulties and time constraints of the project, the following section only presents the results of applications that were based on Chrome or Samsung Internet source code that we were able to test.

As shown in Figure 6.3, the results of the mobile apps performance is fairly consistent with the results on the browsers and extensions on desktop. The only notable difference is that AdGuard performed better on mobile than it did on desktop. During our desktop testing, AdGuard performed slightly better than AdBlock and AdBlock Plus, but worse than Disconnect. During mobile testing, AdGuard performed better than Disconnect and was close to the performance of Brave. Brave was still the best application, on both desktop and mobile.

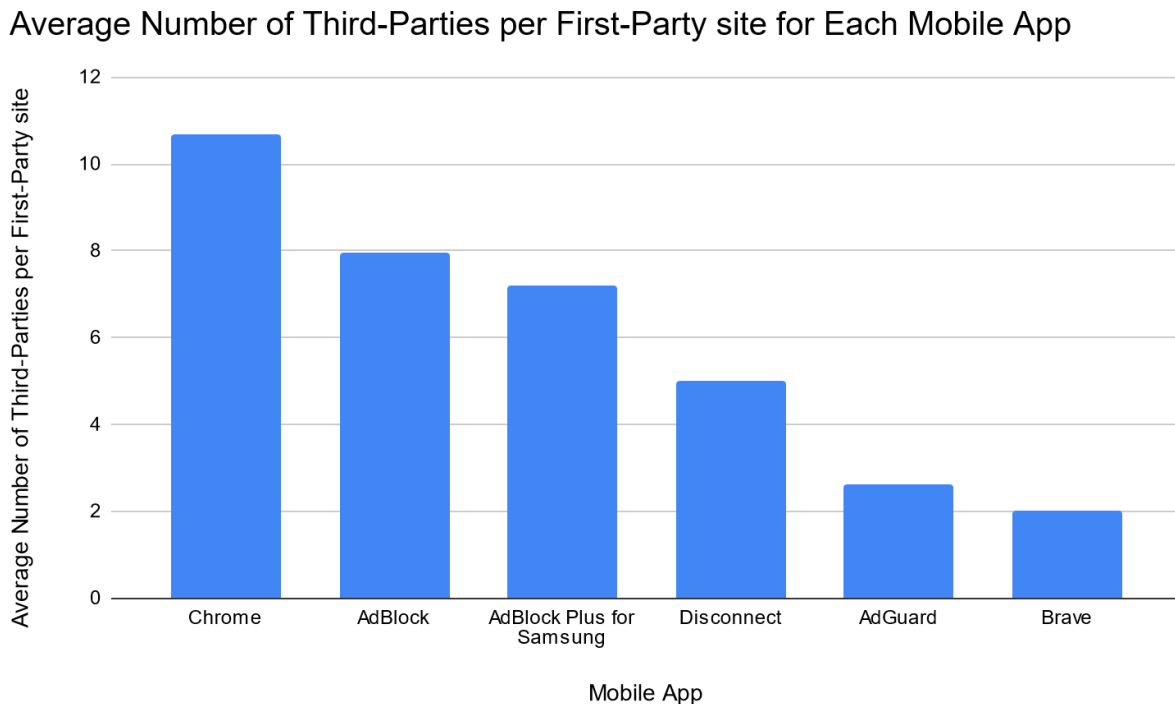


Figure 6.3. Average Number of Third-Parties Per first-party site for Each Mobile App.

Figure 6.4 shows the same information as Figure 6.3 regarding third parties per first-party site on mobile, but third parties are colored-coded up by category. The results in terms of the proportions of the categories and overall performance seem consistent with the desktop results.

Average Number of Third-Parties per First-Party site for Each Mobile App by Category

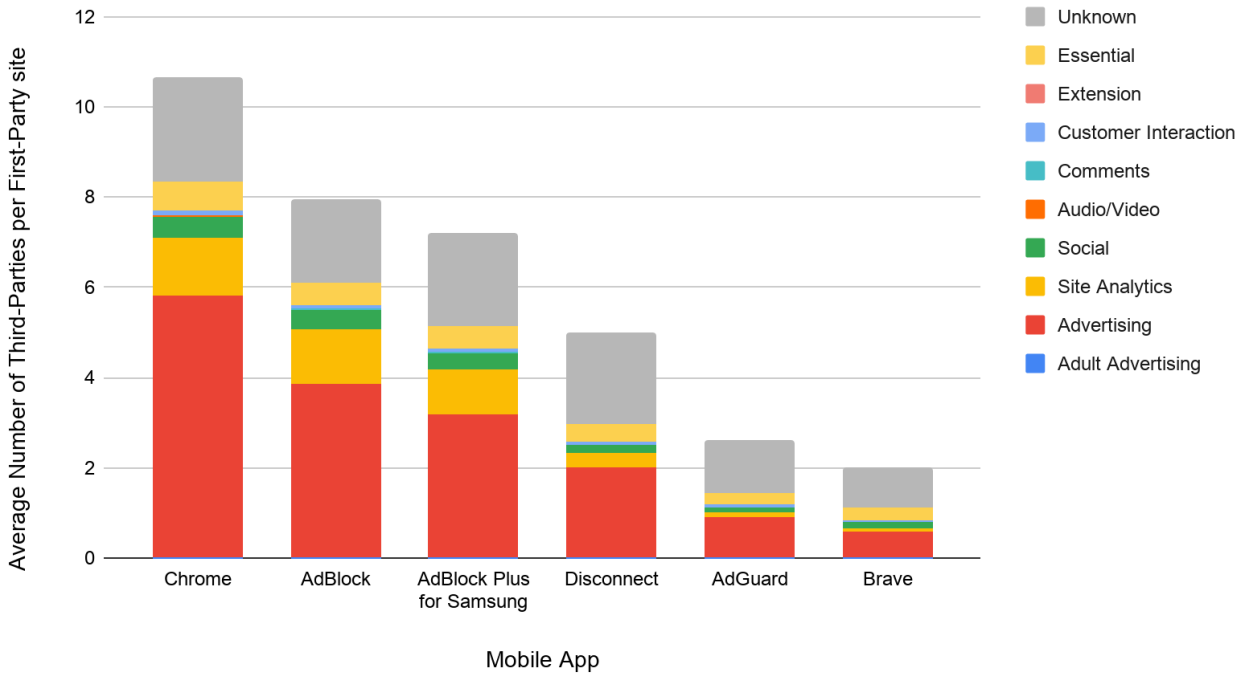


Figure 6.4. Average Number of Third Parties per first-party site for Each Mobile by Category.

Figure 6.5 shows percent of trackers remaining (adult and regular advertising, and site analytics using Chrome as 100%). The results are consistent with the previous figures and with the desktop results, except the Disconnect and AdGuard differences noted earlier.

Percent Trackers Remaining for Each Mobile App

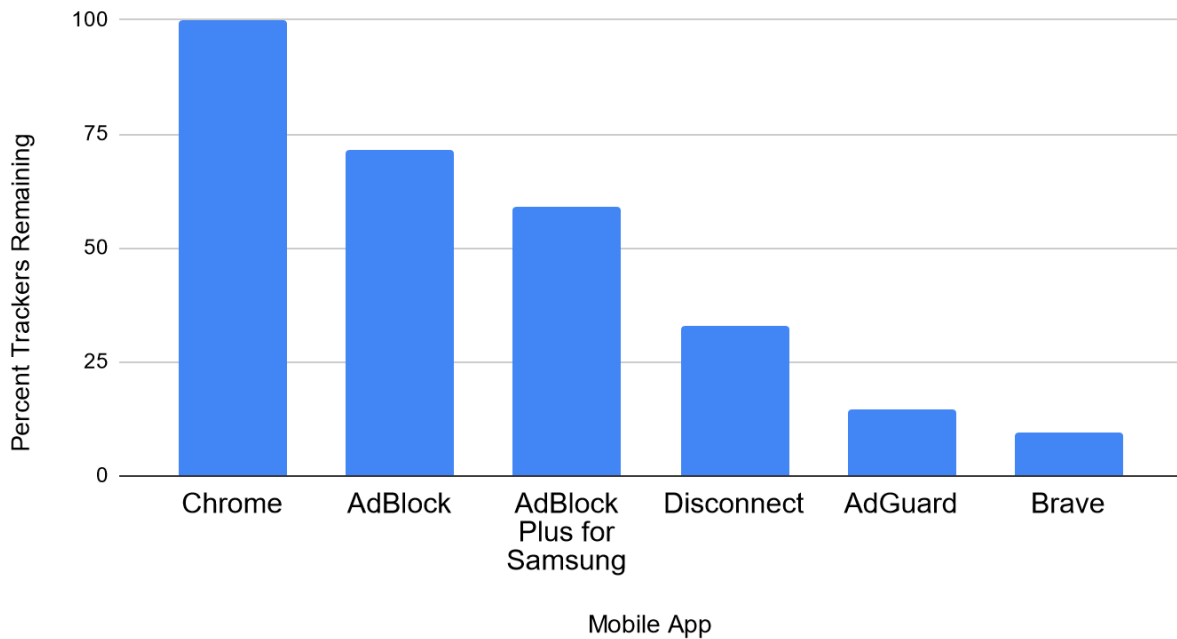


Figure 6.5. Percent Trackers Remaining for Each Mobile App.

6.4 Summary

We also repeated our entire process for mobile given the growing popularity of mobile devices. We evaluated most of the same tools on mobile as we did on Desktop using Fiddler capture and assigning categories to domains. Given the technical difficulties we experienced with testing on mobile, we decided to manually test Alexa Top 50 and were only able to analyze some of the tools. The overall results in the context of proportions of categories and overall performance was consistent with desktop results. The only notable difference is that AdGuard performed slightly better than Adblock and Adblock Plus, but worse than Disconnect.

7. Search Engines

In order to answer research question number five, regarding the effectiveness of search engines to protect user privacy, we examined six search engines. Google and Bing were our baselines while DuckDuckGo, Qwant, Privado and SwissCows were search engines that claimed to provide better privacy to the users. We developed our own methodology to induce interests in sensitive and non-sensitive terms and analyzed the amounts of advertisements we saw during our testing. Wills and Tatar work from 2012 that examined what web advertisers know about the users influenced and shaped our methodology for the search engine testing. In the context of search engines, better privacy means less or no targeted advertising. The researchers analyzed the ads shown to them during their controlled browsing tests and developed a set of induced interests for sensitive and non-sensitive interests. They visited a selection of sites that were equivalent to the induced interest and analyzed the type of advertisements that they saw in a controlled browsing environment. We used the tables as defined in their research with sensitive and non-sensitive interest, but instead of visiting sites that would correspond to the interests in the tables, we came up with sets of five queries to make the search engine algorithm think we were interested in a particular topic.

7.1 List of Search Engines Tested

In our research, we examined six search engines. Google and Bing were our baselines while DuckDuckGo, Qwant, Privado and SwissCows were search engines that claimed to provide better privacy to the users. In the context of search engines, better privacy means less or no targeted advertising.

Table 7.1. List of Search Engines Tested.

Name of the Search Engine	Market Share Data (2019) ⁹	Reason for selection
Google	82.08%	Baseline
Bing	5.72%	Baseline
DuckDuckGo	0.31%	Most popular privacy-oriented search engine based on the market share. Based in the US.
Qwant	0.02%	Second most popular privacy-oriented search engine. Based in France.
Privado	-	Additional privacy-oriented search engine.
SwissCows	-	Privacy search engine based in Switzerland.

1. **Google and Bing:** Google and Bing are the most popular, non-privacy focused search engines that make up the majority of the market share. They were used in our research as a baseline to determine how much tracking the search engines perform and whether their ads follow users around.
2. **DuckDuckGo:** According to Net Market Share, DuckDuckGo is the most popular privacy focused search engine as of 2020. However, compared to Google which accounts for 82.08% of the market share, DuckDuckGo accounts for only 0.31%. The search engine nearly doubled in popularity between 2018 and 2019, in 2018 it only stood at 0.18%. DuckDuckGo, like other privacy-oriented search engines on the list, makes profit by showing users ads during searches. Unlike traditional search engines, DuckDuckGo claims it does not collect any information to build a behavioral profile on the user and only shows them advertisements relevant to the search terms in the search bar. Designing and maintaining a unique web crawler and building a comprehensive web index is an expensive and time consuming endeavour. DuckDuckGo makes use of Bing Ads and Bing search results. DuckDuckGo operates its own web crawler called DuckDuckBot. It

⁹ The data is from “Search Engine Market Share” by Net Market Share.

also uses results from over 400 sources including Wikipedia, yandex, Bing and Yahoo. They say that they maintain their users' privacy by using a proxy call through their servers. When a search engine makes a call to any partner, the partner answers to the DuckDuckGo server and no personal user information is passed on (*DuckDuckGo Help - Sources*, n.d.).

There were a few studies were were able to find comparing features of DuckDuckGo to other search engines, like Google (Parsania, 2016), and several studies comparing performance (Negi, 2014) (Iqbal, 2016). Our group was not able to find any work related to the study of the privacy behind the engine. The 2016 study gave the engine high praise stating that all of its features were comparable to Google search and had a few advantages like the feature of infinite scrolling and its open source nature potentially makes it more trustworthy (Parsania, 2016). The study in 2014 by Negi found that Google came first, Bing came second and DuckDuckGo third in terms of natural language processing and understanding user queries (Negi, 2014). A 2016 study found that DuckDuckGo was still behind Google in terms of relevancy of the terms and the biggest disadvantage was the number of dead links, but overall it was comparable to Google (Iqbal, 2016).

3. **Qwant:** Qwant is a privacy focused engine developed in France. They market themselves by stating that they respect user privacy, do not record searches or use any personal data for advertising, and show unbiased results (no filter bubbles). Qwant makes money by showing ads to the users on the results page, without tailoring the results or tracking the users (*How does Qwant make money?*, 2017). They show ads only relevant to the search terms. Just like DuckDuckGo, they also work with Microsoft Bing for its ad infrastructure. They have their own crawler that indexes the web and have an agreement with Microsoft Bing to complement their own results with those from Bing (*How does Qwant index the web?*, n.d.). They also anonymize user queries by dissociating them and their IP address (*How does Qwant ensure my security?*, n.d.).
4. **Privado:** Privado is a privacy-oriented search engine that claims to not track users. The company claims that they create an anonymous ID for the user and encrypt the search term so it becomes unreadable in the browser history. They do store aggregated search information, however, they claim the data is anonymous and is collected for the purposes of improving the engine. Privado makes its revenue from search results, and shows advertisements to the user based on the search query. They claim the results are not biased on age, gender or any other personal data (*Privado - How It Works*, n.d.). Privado does not specify whether it runs its own search engine and web crawler or works with existing search engine giants.
5. **Swiss Cows:** Swiss Cows is a notable privacy focused search engine entry. Swisscows is based in Switzerland and not the US. They pride themselves on collecting absolutely no data from users. They have their own servers and do not work with third parties or cloud

services. They also have their datacenter in the Swiss Alps which they claim provide an extra layer of protection to user's data (*Swisscows - Our Datacenter*, n.d.). Its privacy policy states that they do not collect any of the following information: IP's, name of the browser used, system information or search terms. The only piece of information they store is a counter, for the amount of terms they receive per day to measure total traffic. Swisscow search results are sourced partially from Bing, however they also run their own Swisscow Crawler that claims to have innovative technology behind it (*Swisscows - Products*, n.d.). They also developed their own ad technology called "AdAnounce" to help show more relevant advertisements based on the search term, without tracking their users.

7.2 Methodology

To answer this research question, we examined privacy-oriented search engines and determined which ones live up to their privacy claims. Manual testing was conducted by entering queries into the search engines and the webpages were analyzed for the types of the ads displayed.

Below is the methodology we used to determine how many advertisements we saw related to the induced interest in the search engines.

1. Install Virtual Box.
2. Installed Ubuntu 18.04.5 Image onto Virtual Box.
3. Installed Firefox Version 82.0 (64 bit) (Mozilla Firefox for Ubuntu)
4. Installed NordVPN extension for Firefox.
5. Took a Snapshot of the virtual machine in the "clean" state with no searches that we would roll back to, to make sure no cookies were saved and reused across sessions.
6. Ran NordVPN extension and connected to the VPN.
7. Checked our public IP address and wrote it down in Table 7.4. We wanted to make sure we vary the IP address from test to test to make sure search engines and advertisements couldn't create an association between our searches.
8. Opened Google, Bing, DuckDuckGo, Qwant, Privado, and Swiss Cows and ran 5 searches in each based on Table 7.2.
9. Visited 13 popular news sites and Youtube to see what advertisements show up. Took screenshots of the ads that seemed related to the induced interest.
10. Counted and recorded the number of ads shown.
11. Reset the virtual box using the restore function and the previously taken snapshot of the machine.
12. Repeated steps 7-11 for the Table 7.2 and for Table 7.3 interests.

Our methodology and work has been heavily influenced by the work from 2012 by Wills and Tatar. First, we referenced Table 1 and Table 2 in the Wills and Tatar work to establish what we defined as “sensitive interests” vs. “nonsensitive interests” (Wills & Tatar, 2012, p.3-4). Based on the terms they used in their work, we came up with our own two tables listing the induced sensitive and non-sensitive interests that can be seen below in Table 7.2 and Table 7.3. Then we came up with five searches for each term that would be used to generate interest in the term.

Their work did not deal specifically with search engines and during their process they clicked on various sites and links related to the interest they were inducing. We purposely did not click any of the links that appeared during our searches since we wanted to analyze information leakage only from performing the search itself. We hypothesise that the information leakage would have been further increased if we clicked some links related to the induced interests since those sites are likely to contain their own site analytics and trackers as well that would further aggregate interest data.

Table 7.2. Induced Behavioral Interests

Induced Interest	Search terms
Cars	<ol style="list-style-type: none"> 1. Buy a car near me 2. Ford 3. Toyota 4. Tesla Model 3 5. Best Cars to Buy 2020
Dogs	<ol style="list-style-type: none"> 1. Dogs 2. Adopt a dog near me 3. Veterinarian near me 4. Best dog food to buy 5. Best dog breeds to adopt
Golf	<ol style="list-style-type: none"> 1. Golf 2. Best golf equipment 3. Learn to play golf 4. Virtual golf 5. Best golf clubs to buy
Investment	<ol style="list-style-type: none"> 1. Stocks 2. Stock Market 3. Best stocks to invest in 4. How to start investing 5. Best investing platforms
Florida	<ol style="list-style-type: none"> 1. Florida 2. Move to florida 3. Attractions in Florida 4. Buy apartment in florida 5. Plane tickets to florida
Tennis	<ol style="list-style-type: none"> 1. Tennis 2. Learn to play tennis 3. Tennis near me 4. Best tennis equipment to buy 5. Virtual tennis

Table 7.3: Induced Sensitive Interests.

Induced Interest	Searches:
Bankruptcy	<ol style="list-style-type: none"> 1. Bankruptcy 2. Chapter 7 3. Foreclosure 4. Debt 5. Tax Relief
Depression	<ol style="list-style-type: none"> 1. Depression 2. Mental health resources 3. Therapy near me 4. Depression medicine 5. How to determine if you have depression
Diabetes	<ol style="list-style-type: none"> 1. Diabetes 2. Foods for diabetes 3. What blood sugar levels are normal 4. Diabetes monitor 5. Diabetes medicine
Gay/lesbian	<ol style="list-style-type: none"> 1. LGBTQ+ community 2. Lesbian 3. Gay 4. Lesbian and gay dating apps 5. Gay marriage in the US
Pregnancy	<ol style="list-style-type: none"> 1. Pregnant 2. Pregnancy tests 3. Baby 4. Abortion 5. Pregnancy test
Skin cancer	<ol style="list-style-type: none"> 1. Skin cancer 2. How to know you have skin cancer 3. Sunscreen and skin cancer 4. Dermatologist near me 5. Skin cancer treatment

At the beginning of our testing, our group hypothesised that certain interests such as “cars” and “investing” will appear more often since there is likely more money involved in advertising cars and investing products, than something like golf. In order to avoid biasing our results and assigning a term to every search engine such as Google with “cars” term and DuckDuckGo with “investing” and so forth, we used all of the five terms with every search engine. The terms and the engine they were used with were rotated as can be seen in Table 7.4.

Table 7.4: A setup of interests generated in each search engine.

IP Address:	Google	Bing	DuckDuck Go	Qwant	Privado	SwissCows
184.170.253.68	cars	dogs	investing	golf	Florida	tennis
62.182.99.89	tennis	cars	dogs	investing	golf	florida
62.182.99.124	florida	tennis	cars	dogs	investing	golf
212.103.33.121	golf	florida	tennis	cars	dogs	investing
185.240.246.116	investing	golf	florida	tennis	cars	dogs
173.48.212.138	dogs	investing	golf	florida	tennis	cars
64.94.215.116	Bankruptcy	Depression	Diabetes	Gay/lesbian	Pregnancy	Skin Cancer
217.138.198.236	Skin Cancer	Bankruptcy	Depression	Diabetes	Gay/lesiban	Pregnancy
185.217.69.152	Gay/lesian	Skin Cancer	Bankruptcy	Depression	Diabetes	Pregnancy
104.140.52.123	Pregnancy	Gay/lesian	Skin Cancer	Bankruptcy	Depression	Diabetes
185.240.246.52	Diabetes	Pregnancy	Gay/lesian	Skin Cancer	Bankruptcy	Depression
62.182.99.67	Depression	Diabetes	Pregnancy	Gay/lesian	Skin Cancer	Bankruptcy

7.3 Search Engine Results

The following section presents the results in order to answer question five from Chapter 3, regarding the ability of search engines to protect user privacy.

In order to analyze which search engines offered better privacy we examined the number of ads shown during our testing as described in Chapter 7.2. First, we added up the number of advertisements shown after using each search engine. As shown in Figure 7.1,

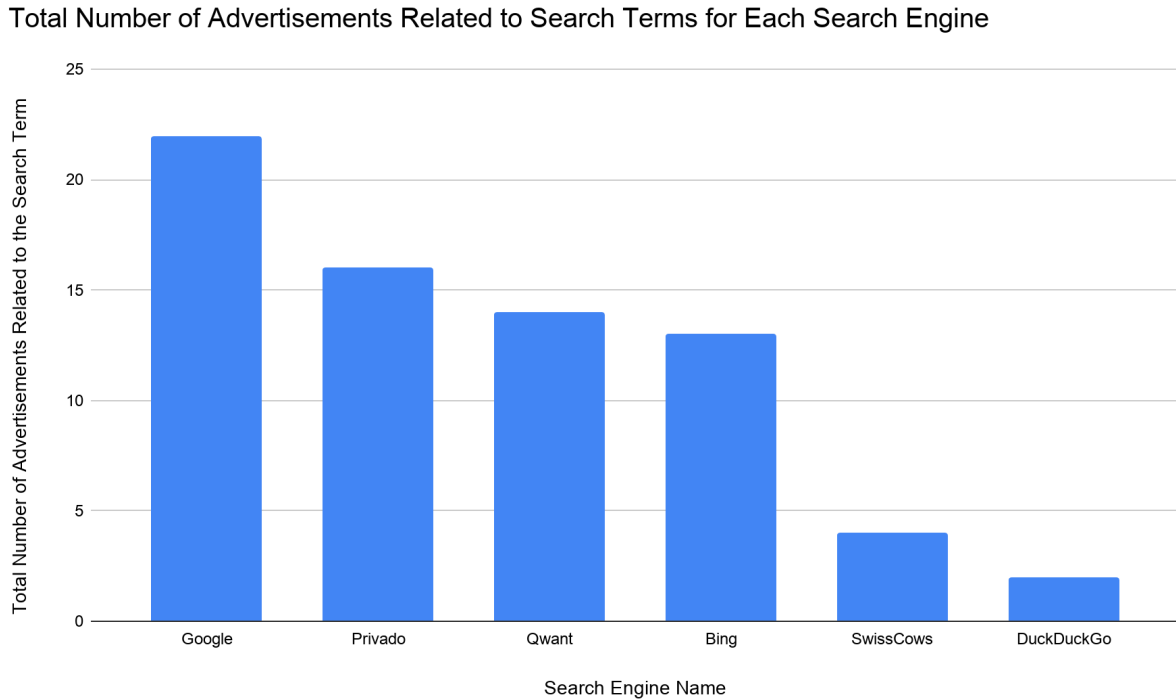


Figure 7.1. Total Number of Advertisements related to search terms for each search engine.

Total Number of Advertisements Related to Search Terms for Each Search Engine by Induced Interest

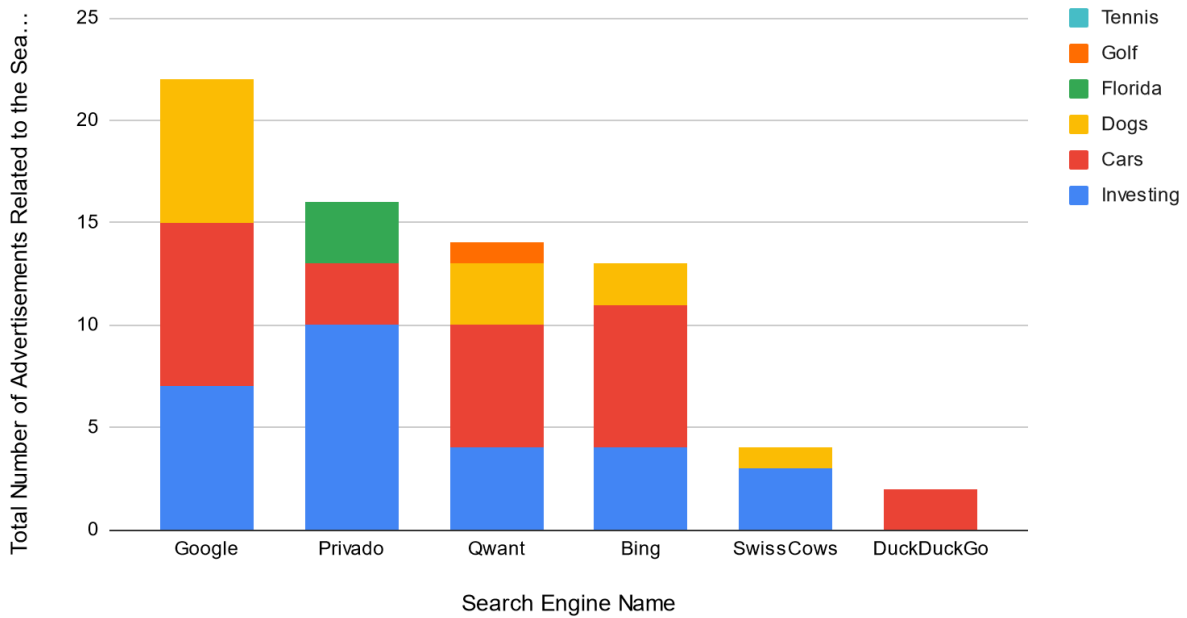


Figure 7.2. Total Number of Advertisements related to search terms for each search engine by Induced Interest.

Before starting search engine testing, we anticipated that certain terms such as cars and investing might have more ads associated with them. Our hypothesis was backed up with our results. As shown in Figure 7.3, investing and cars induced interests resulted in the most ads shown. Dogs, Florida and golf were the next three terms in order of decreasing popularity and we saw no advertisements related to tennis. Although we hypothesized that investing and cars will result in more ads shown, it is also probable that even without typing anything into the search bar, the websites we visited would have contained several advertisements related to the terms. If we were to repeat our methodology, we would add a baseline test and then subtract the baseline from the results to get a more accurate picture to what extent did search engines affect advertisements shown.

Induced Interest vs. Total Number of Ads Related to the Search Term Seen

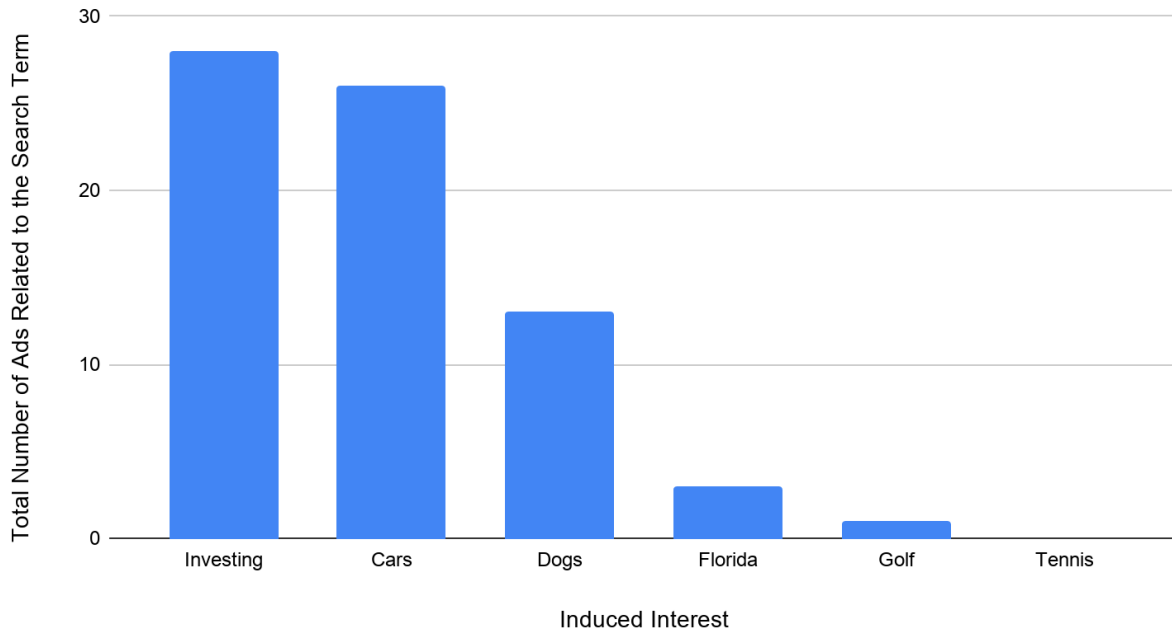


Figure 7.3. Induced Interest Term vs. the Total Number of Advertisements Shown Related to the Induced Interest.

7.4 Summary

According to our results, the best search engines in terms of respecting user privacy were DuckDuckGo and SwissCows. The finding is not surprising given that both of the search engines advertise themselves as privacy-oriented. Google and Privado performed the worst. The poor performance of Google is expected, but the same cannot be said for Privado. The fact that Privado performed only a little better than Google, but much worse than any of the other search engines, including Bing, is surprising and concerning given that they advertise themselves as another privacy-oriented search engine.

8. Recommendation Survey

Based on all of the information we gathered from our analysis of tools tested, we created a recommendation system for users. We wanted to give them a centralized location to see all of our results and how they could go about getting these products. Due to the growing concerns users have about internet privacy (*The Data Privacy Feedback Loop*, 2020, Wang, Lee, & Wang, 1998, Turow & Hennessy, 2007), this system was an essential part of our research.

8.1 Research Questions

Based on the data that we gathered from our previous analysis, we wanted to focus on a few key questions in terms of what participants perceived the extent of their privacy knowledge and protections to be. We were able to formulate the following research question from that aim.

8.1.1 Switching to Privacy Tools

Research question: Will viewing personal data - as compared to generic data - that Google has calculated influence participants' willingness to switch to more privacy-focused, browsing technologies than?

We initially sought to develop a recommendation system that would inform users of our findings and provide them with the list of tools we identified in our research. To accomplish this task, we created a survey-style questionnaire using the Qualtrics software. The survey gathers information of user intentions, knowledge, and perceptions of privacy to provide users with a personalized recommendation of tools they indicated being interested in. We extended this survey to include an experimental factor as a way to measure the attitudes and intentions of internet privacy and its protections.

8.1.2 Lack of Control vs. Google Synthesization

Research question: Would participants, who reported feeling a lack of being in control of the information websites can gather about them believe that Google would be able to more easily determine their hobbies?

It would make sense that if our users that report having a lack of control in the types, and amounts, of information sites gathered about them would report the same feelings for Google. We included questions about Google to show them that the sites that they use every day are tracking us in one way or another, especially if they are bigger technology companies. We wanted to see if the perceived lack of control would be similar to the perceived accuracy of Google as they gather information on users.

8.1.3 Lack of Control vs. Seeing Invasive Ads

Research question: Would participants, who reported feeling a lack of being in control of the information websites can gather about them report seeing more invasive ads than those that felt more in control?

Using the same measurement, the amount of control over information surrendered to websites, we wanted to see if there was any association to the amount of invasive advertisements the participants saw. We needed to first make sure, however, to first measure the validity of the term ‘invasive’ by asking our participants questions about their opinions on what characteristics of invasive advertisements are. We then can measure any association between the two variables.

8.1.4 Societal Issue vs. Seeing More Invasive

Research question: Do those that see invasive advertisements as a societal issue report seeing more invasive ads?

Those that see invasive advertisements as a societal issue, we hypothesized, might be the ones that experience more advertisements. We created questions that would focus on different ways for participants to report seeing invasive advertisements and compare that against how much people see this topic as a broader issue. Measuring these two variables will give us insight into how much participants might be paranoid over invasive advertisements. If those that see these ads as a problem report seeing more advertisements in general, it will lead us to believe that they feel as though this type of recommendation system is in demand to them and want to start protecting their data. Once again, we will first have to rely on our validation of any variables that measure the word ‘invasive’ and its meaning.

8.2 Methodology

Now that we had our roadmap of the questions that we wanted to answer, we moved into creating our survey and researching how to specially ask the right questions in the right ways in order to get valuable information to analyze later.

8.2.1 Participants

We had two different sources of where we pulled participants from. In total, we received 146 responses over both surveys, which we will discuss later in Section 8.2.7. Our core survey had 69 responses (48 being female) with a mean age of 21.4. All of these participants came from the WPI Sona System participant pool. The Sona System is a way of gathering participants for studies by offering them class credit for their participation. Almost every major was represented in a variety. As for the revised survey, which will be discussed, we received 77 responses.

8.2.2 Conducting Survey Research

We first began by conducting research on how to run survey research efficiently. We turned to a book written by Fink to be able to help us understand the steps that we need to take in order to create all of our questions in an appropriate manner (Fink, 2009). Fink's book goes into detail on creating surveys that will allow us to gain meaningful data to analyze while still allowing for the participants to be comfortable and informed with answering all the questions being asked. Some of our questions will be asking participants to answer questions that they might not have enough knowledge on to make a final decision, Fink stresses the importances of leaving an option for them not to answer or claim that they are unsure of their response. Any ambiguity in our instruction, she claims, should be spelled out for the user so they do not attempt to offer their own interpretation. We were able to accomplish this task by giving our participants specific instructions on where to go to obtain the information we were looking for. We also took into account her note of adding in familiarity to the beginning of each section by adding a priming introductory statement, informing our participants about the questions they will be asked in the section. Finally, we created questions that aligned with her claim that surveys need to be valid. We created questions that asked participants how accurate certain information was to the participants as well as what they believed were characteristics of "invasive advertisements". This was to ensure the validity of our questions, assuring that we had the same working definitions as our participants.

8.2.3 Creation of Questions

To begin creating our questionnaire, we first needed to develop an exhaustive flowchart that contains all of the possible outcomes of the survey. This flowchart goes into detail about all the questions that are needed to assess the current setup of the user, their current satisfaction of performance, desires, and intentions in the context of privacy tools. A section of the chart can be seen in Figure 8.1. In this small section of the chart, we can see the logic flow of answers if a participant chose Google Chrome as their primary browser and indicated that they are using an ad block extension already.

In addition to the flowchart, we brainstormed relevant questions that we could think of to ask. As we came up with more questions, we labeled each question into a specific category. We started with six different classifications of questions that got more narrow as we refined our list. These six categories asked the user's technological setup, their knowledge of privacy, their current privacy behaviors, their attitudes on privacy, potential perceived risks and benefits of privacy protections - or lack thereof - and their opinions on the severity of privacy.

We went through three iterations in order to narrow down the specific topics we wanted to focus on. Some of the main topics ended up joining together to have three categories survive the final refinement. The final groups included the user's technological setup, their knowledge of

privacy, and their potential future intentions of using privacy-focused tools. Once we finalized our categories, we then worked on increasing the quantifiability of each question in order to not leave a lot of qualitative coding to be done at the end. One question was left to be qualitatively coded due to the fact that there were too many ranges of answers that a participant could answer. This option is necessary at times to allow users to have freedom and flexibility to

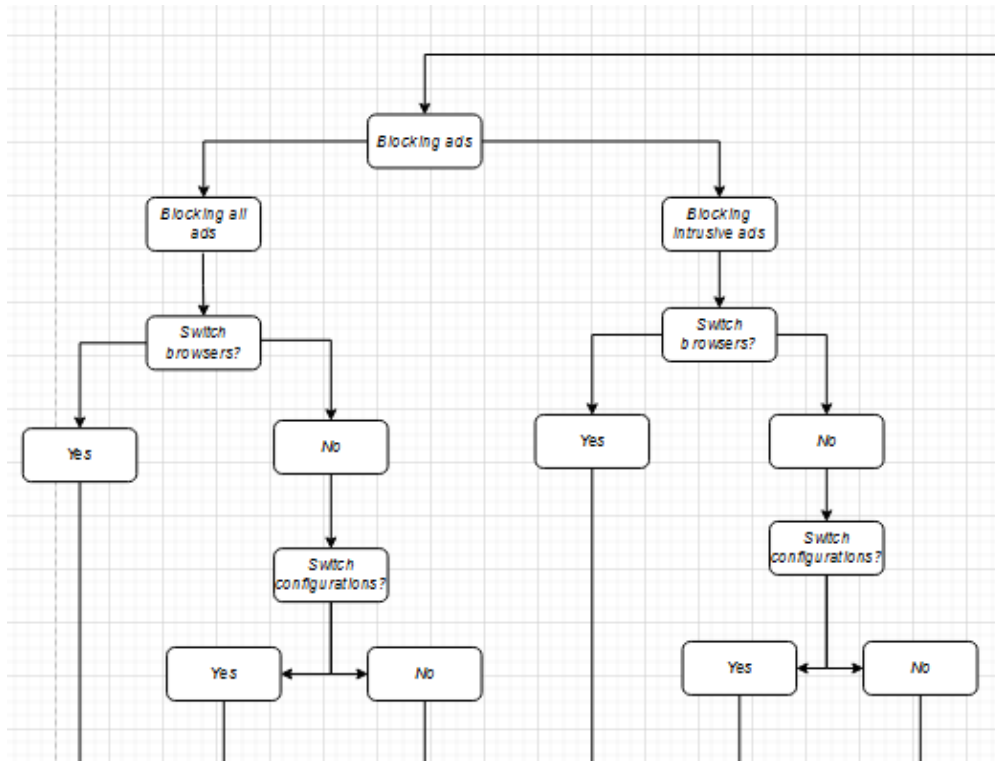


Figure 8.1 Snippet of flowchart containing answers for our survey recommendation questions

explain topics in their own words that we can group together later in analysis. A list of all final questions can be found in Appendix B.

8.2.4 Experimental and Control Groups

After all of the questions were put into our survey software, our group noticed that we could take advantage of our survey to continue the work done in 2019 (Bashir, et al., 2019). We could further assess users' opinions on internet privacy by adding an experimental factor: showing users human profiling in action as discussed in Section 2.2.2. To do further assessment, we utilized a website that is publicly available, hosted by Google. Google hosts a site that allows anyone to view all of the information that they have gathered about a person based on their

search history, products viewed, and other sites visited. The URL of the site is 'adssettings.google.com'. Using this information, they create a human profile, calculating certain characteristics such as age range, gender, income, marital status, and hobbies. Our study's participants were given specific instructions to visit this site and view all of their own personal data gathered by Google. A suggested time frame was given in order to maintain some control over the survey. We would not be able to know if the participants would want to go off and explore their data in certain ways, therefore, we provided them with a time frame to allow us to maintain a tighter control on their actions.

In order to account for the possible results that the experimental group would yield, we manipulated a data set that was presented to our control group. We sent out a request to 8 students, with a variety of majors, who volunteered to visit the same site as the experimental group. They were asked to send us their list of their hobbies that were listed by Google. We then took the top 30 most common hobbies that were shown by most, if not all, of the 8 students' data and edited them into one image. The image was presented to the control group in the place of visiting the site itself. The control group was given the same, detailed instructions with the difference in description. We explained that this image was what the site looked like for a typical student at WPI. An image of this image can be seen in Figure 8.2.

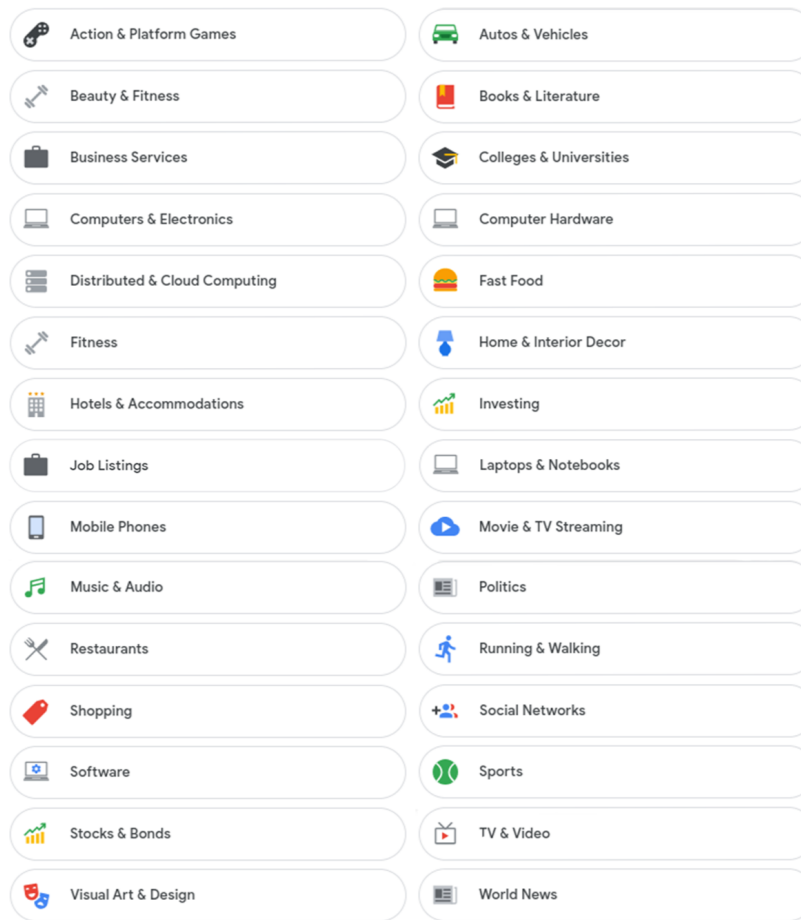


Figure 8.2. A manipulated list of the interests of “the common WPI student” for the control group of our experimental survey

Adding the experimental factor to our survey required us to add in a notification at the end to inform our participants of the true meaning behind the actions that we had them do. The addition requires a debriefing form to be added to the end of the survey. The debrief discussed the differences between the experimental and control groups and the data that each saw.

8.2.5 Survey Design

Our survey was a between-participants design with two conditions. Random assignment was utilized to place participants in either our experimental or control groups. The implementation of this design was done by enabling the random flow generator though Qualtrics itself. There was one exception, for participants to participate in the experimental condition, they needed to have an active Google account that they use while browsing the internet. If

participants reported not having an active account, they were put into the control group. We assumed that not having an active Google account would be a rare occasion, therefore we decided that this solution would be a reasonable way of handling the edge case.

The different parts of the survey that survived the refinement includes information gathering of thoughts and opinions, privacy background/education and current setup, and possible technology switching intentions. Additionally, there were questions that were part of the experimental factor of the survey, along with the recommendations and the demographics. We decided to include the information gathering section first, in order to prime our participants with the topic of internet privacy and get them ready to answer questions regarding their current setup and willingness to switch.

When asking participants about their current setup, we broke down their current technologies into two different sections, browsers and extensions. When asking about browsers, we decided to make Brave its own category. Based on it placing so high on our charts - as well as it being semi-common as a primary browser - if users are currently using Brave, there would really be no reason to suggest anything else for them. It would be evident, also, that they are aware of the benefits of using Brave and are currently using the browser for its ability to protect personal data. For the other browsers and extensions, we provided detailed instructions, along with links and occasionally pictures, for participants to follow in order to find their information and report it to us. These details are helpful for any user who might not be familiar in finding the information we asked.

All of these questions led up to a recommendation section that participants had an opportunity to email to themselves if desired. The recommendation has links and descriptions of the personalized tools that we believe would strongly benefit them if they cared about keeping their data more private. As part of our testing, we made sure that the links persisted through email and would allow the participants to refer back to the email if they wished to switch browsers or extensions at a later date.

While designing our survey, we gave our participants specific directions in an attempt to obtain the maximum amount of control as described in Section 8.2.4. There were slight limitations to these directions at some points. For example, when asking the experimental group to visit the Google-hosted site, we are unable to know how much time they spent viewing their data. The only way to account for this limitation would be to design our own website and log the amount of viewing time per participant along with other website characteristics.

8.2.6 IRB Approval

At this step in our research, we applied for an IRB approval from the school. Appendix C is the IRB letter of approval we received to continue with our study. We received an exempt letter from further review due the fact that we were causing little to no risk to our participants. A copy of our informed consent can be found at the beginning of Appendix B and our debrief form for our participants can be found at the last part of Appendix B.

8.2.7 Revised Survey

Our survey project was initially made to serve as a recommendation system to educate users on their own privacy. However, with the current state of the survey, we would not be able to have it be used widely. We wanted to take steps towards the delivery, distribution, and utilization of our tool to the real world. For this reason, we decided to make an abridged version of our survey. The new version would have a few key pieces missing, including the experimental and control conditions, visiting a Google website, and any supplementing questions.

There are certainly limitations to the way that we collected responses. Using the Sona system limits the responses we recorded to a population of more educated, and younger than the representative person. We decided that we wanted our tool to reach people of a broader audience. There exists an 'opt-in' mailing list that many of the faculty and staff are a part of. The main purpose of the mailing list is to purchase, sell, or trade objects such as furniture or printer parts for offices. Occasionally, however, there will be miscellaneous tools - such as ours - that are shared to increase awareness for certain topics. This list appeared to be a good opportunity to get additional answers for our survey.

The email group comes with its own limitations. As compared to the Sona participant pool, we gather responses from an audience that is closer in age to the typical person. We still, however, have to acknowledge the fact that most of the faculty and staff on the list have higher incomes and are more educated than the common person. Despite these limitations, we sent out a modified version of our survey in order to get some additional responses on the general questions that did not pertain to the experimental factor.

From the new survey, we had to take out the demographics section of the design. The questions in the demographics we had previously discussed participants' characteristics about them being students, which would not apply to the new population. A slight oversight was not changing the demographics to align with faculty and staff, therefore no stats can be reported. Despite this oversight, however, we do know roughly the demographics of those on the email list as previously mentioned. We also eliminated the experimental factor which was geared towards the 'common WPI student' as described in the wording of the questions. Not only would these questions not apply to the new population, but we would also lose all of the control we would have over who is seeing what questions. There could be participants who followed the links to the Google site and ignored it to get to the end of the survey, or have those who looked at the new site and never returned.

As a result, we will be able to report the questions that were similar between the two surveys. For example, we can focus on what invasive ads look like between groups and compare and contrast. Although we can not provide specific demographic information, we will still be able to gather information on the privacy knowledge and concerns of this group to compare them to a younger population.

8.2.8 Survey Logic

Extensive logic flow was built into our survey to direct participants to their personalized recommendations. The logic followed the same direction of our flowchart that was created to make the recommendations tailored to users. There was additional logic that was also added in other main sections to account for options where participants could express their uncertainty in intentions. It is reasonable to believe that some participants would not necessarily know if they would be willing to switch to certain tools or not within the time it takes to complete our survey. Logic was built in order to allow users to see the types of options that could be available to them before making decisions.

The logic flow would eventually lead users to their recommendation that was specifically crafted for them. We had in total, nine different recommendations that a user could fall into depending on which type of technology they indicated in changing. For some of the options, we still suggested using at least a different configuration or a privacy extension in order to educate users with tools they can use in the future if they wish. Each of the nine combinations are explained in Table 8.1. You can see the combinations as they are presented in the survey starting on page 132 of Appendix B.

Table 8.1: List of Survey Results.

Condition	Recommendation
1	<ul style="list-style-type: none"> ● Google Chrome <ul style="list-style-type: none"> ○ Suggested switch to “Block 3rd Party Cookies” ● Privacy Extension <ul style="list-style-type: none"> ○ Suggested use of Ghostery ● AdBlocker <ul style="list-style-type: none"> ○ Suggested use of uBlock Origin
2	<ul style="list-style-type: none"> ● Google Chrome <ul style="list-style-type: none"> ○ Suggested switch to “Block 3rd Party Cookies” ● Privacy Extension <ul style="list-style-type: none"> ○ Suggested use of Ghostery ● AdBlocker <ul style="list-style-type: none"> ○ Suggested use of AdBlock
3	<ul style="list-style-type: none"> ● Google Chrome <ul style="list-style-type: none"> ○ Suggested switch to “Block 3rd Party Cookies” ● Privacy Extension <ul style="list-style-type: none"> ○ Suggested use of Ghostery
4	<ul style="list-style-type: none"> ● Firefox <ul style="list-style-type: none"> ○ Suggested switch to “Strict” ● Privacy Extension <ul style="list-style-type: none"> ○ Suggested use of Ghostery ● AdBlocker <ul style="list-style-type: none"> ○ Suggested use of uBlock Origin
5	<ul style="list-style-type: none"> ● Firefox <ul style="list-style-type: none"> ○ Suggested switch to “Strict” ● Privacy Extension <ul style="list-style-type: none"> ○ Suggested use of Ghostery ● AdBlocker <ul style="list-style-type: none"> ○ Suggested use of AdBlock
6	<ul style="list-style-type: none"> ● Firefox <ul style="list-style-type: none"> ○ Suggested switch to “Strict” ● Privacy Extension <ul style="list-style-type: none"> ○ Suggested use of Ghostery
7	<ul style="list-style-type: none"> ● Brave ● Privacy Extension <ul style="list-style-type: none"> ○ Suggested use of Ghostery ● AdBlocker <ul style="list-style-type: none"> ○ Suggested use of uBlock Origin
8	<ul style="list-style-type: none"> ● Brave ● Privacy Extension <ul style="list-style-type: none"> ○ Suggested use of Ghostery ● AdBlocker <ul style="list-style-type: none"> ○ Suggested use of AdBlock
9	<ul style="list-style-type: none"> ● Brave ● Privacy Extension <ul style="list-style-type: none"> ○ Suggested use of Ghostery

Users could report if they felt comfortable and willing to switch not only their browsers and extensions, but also their current browser configurations as well. If a user did not fall into any of these categories, we had default ‘no change’ options as well, where we discussed that we acknowledge the participant might not have wanted to change their setup, however, we provided them with our general recommendations if they were to want to change in the future. This option, once again, could be emailed to the participant if desired. At the end of the survey, everyone was presented with our debriefing form so they were able to see what exactly was being measured.

8.2.9 Survey Testing

The survey was sent out to 12 people for pre-testing, making sure that grammar, survey flow, and timing were all correct and accurate. We received feedback from each individual that indicated clear changes in certain parts of the survey that had been overlooked during development. The first being the timing of the survey. In our informed consent, we initially had thought that the survey would take longer than it actually had when being tested. To address this concern, we consulted with the testers and reported a more accurate time based on the average reports. Secondly, there were a few syntax and grammatical errors that we were able to fix. Some of these issues were located in the informed consent, which caused some testers confusion in the subject of some of the terms. Finally, we fixed the ordering of how some parts of the survey was presented. All of the logic can be seen in the output of the Qualtric survey in Appendix B.

8.3 Summary

We created two surveys to supply our users with recommendations for the tools that we identified in our prior research. We were able to hide an experimental factor in our original survey in order to get responses from two different populations to understand what users think invasive ads look like and understand the level of concern they have of these ads. There are clear demographic differences between the two audiences of the different surveys, each with their own limitations that have to be taken into consideration. Due to these known limitations, we are unable to analyze the results from the one lacking the experimental factor.

9. Survey Results

Our survey was available on the WPI Sona System from March 12 to April 15. We recorded 69 participant responses from that source and 77 from our alternative survey. Our alternative survey responses were gathered from April 7 to April 19. From all of the responses we were able to run tests to compare mean responses and compare and contrast the results from the two different populations.

9.1 Experimental Survey Results

We will now be discussing the results for the experimental survey. All of our participants reported having an active Google account, which was helpful in placing everyone evenly into our experimental and control conditions. The main piece of analysis we were interested in was seeing if the different conditions would have any significant effect on the rates that participants would report being willing to switch to more privacy-focused tools.

We first asked about what invasive ads looked like to our participants. It was important for us to know that we had the same working definition of what intrusive and invasive ads are. If they were different, there would be a lack of validity behind their responses indicating that they believe these types of ads are a problem. Our questions regarding the specific characteristics tried to focus on the physical appearance of the ads as well as the content of the ads. One question was left as an open response for participants that required qualitative coding to analyze.

Four main responses appeared from both this question and a previous question regarding other common characteristics of ads. Those being identifying information (such as age and gender), personal interests (such as hobbies), location, and psychological health. 82% of all responses mentioned location, 76% mentioned identifiable information, 63% touched upon personal interests and 37% identified psychological states as being an issue. In order to find these numbers, we added all of our responses to an Excel spreadsheet and manually began to see which words stood out. From there, we created separate categories with those words we identified and labeled each response with a 1 in each of the new columns created based on what the response mentioned. Finally, we divided the sum of each column against the number of response we received to obtain each percentage.

Now that we know that we have similar working definitions, we were able to analyze questions pertaining to the overall levels at which users believe that these ads are of real concern. The Data Privacy Feedback Loop back in 2020 reported that many believed that privacy protections for users is a big concern that companies need to focus on and the government puts effort into keeping users safe while online (*The Data Privacy Feedback Loop*, 2020). We reached similar conclusions with our findings, many of our users also claimed to have a high level concern for their own privacy while browsing.

9.1.1 Switching to Privacy Tools Per Condition

First, we looked at the set of results for our main research question. We analyzed the data that pertained to switching tools per condition. In order to accomplish this task, we had to come up with a metric to be able to measure the amount of tools that a participant would switch to and compare this number between the two conditions. We decided to make a ‘switch’ metric that was a single integer in the range of 0 to 2 to represent either no change, change in either a browser or extension, or a change in both. We ran an independent T-test to try to find an association between the amount at which groups reported they would switch based on this new calculated measure. The control group ($M = 1.24$, $SD = 1.136$), unfortunately, showed to not have a significant difference to the experimental group ($M = 1.27$, $SD = .827$) when comparing the average responses, $t(67) = -.115$, $p = .91$.

The results that we gathered shows that there is no association between condition and willingness to switch. There seems to be a ceiling effect that has occurred with our data. In our sample, it seems that we have had each participant indicate that they would be willing to switch, regardless of condition. These results mean that we can not draw conclusions either way about our research question. For our specific sample, we had a ceiling effect present, however, we do not believe that would always be the case. We took samples of undergraduate students at a science and engineering school, it makes sense that these students might be a bit more concerned with their privacy than those at other types of universities. The questions asked in the survey, however, are only asking about behavioral intent of their actions. Their follow-through behavior is beyond the scope of our project.

9.1.2 Level of Control vs. Google Accuracy

In both conditions, we asked participants about the accuracy of Google in calculating their age, gender, and hobbies. Participants reported that for both age and gender, Google would or were able (depending on condition) for the experimental condition, accurate in coming to conclusions on these characteristics. Our participants reported an accuracy of 6.12 for age and 6.57 for gender in the experimental group. These reports are gathered from responses on a 7-point likert scale. We saw similar results for the control group, 5.9 for age and 6.61 for gender. These responses make sense in context of the responses to our main analysis. Both conditions’ participants reported high numbers in accuracy from Google and also both reported high rates of being willing to switch to different technologies. These two measures could be associated.

To supplement these reports, we also found a significant set of results when analyzing how much our participants reported not being able to have control over the information sites can gather about them and comparing the responses to how much they believed Google would be able to gather information about their hobbies. We asked participants the level at which they believed that they were able to control what kinds of information the sites they visited could synthesize about them. To follow this question, in both the experimental question and the control

groups, we asked a question regarding how much they believe that Google did or could guess their hobbies correctly. For those that reported having less control ($M = 2$, $SD = 1.109$), they also responded feeling as though Google was or could be more accurate at gathering information on them. Comparatively, those that had a higher perceived level of control ($M = 2.79$, $SD = 1.409$) claimed that Google would be or was less accurate in their calculations, $t(67) = 2.076$, $p = .046$.

9.1.3 Level of Control vs. Presence of Invasive Ads

Another set of analyses that we were able to significantly measure was comparing the same measure of control to questions regarding the current level of invasive ads that our participants see. We asked two questions that aimed to assess the amount of which users currently see invasive advertisements. These two questions were how many times in the past 6 months have our participants stopped using certain websites and to what degree do our participants believe the ads they currently see are too invasive. We used a 4-point scale to measure the number of sites a user has stopped using. Each point indicated either 0, 1-2, 3-5, or 6 and higher sites that they have terminated use of. We compared this number to the 7-point likert scale for our questions of amount of control as well as current invasiveness. We found that participants who reported the termination of a larger number of sites due to invasive ads ($M = 2.1$, $SD = 1.21$) had higher levels of control in their browsing as compared to those who claimed to stop using a larger number of sites ($M = 3$, $SD = 1.388$). These control levels were negatively associated with the amount of sites that they stop using $t(67) = 2.389$, $p = .021$.

We tried to replicate these results for the level at which our participants felt that the current ads they see now are too invasive, however we did not achieve a low enough significant level at an alpha level of .05. Those that reported a higher level of control tended to report seeing a lower level of invasiveness in advertisements ($M = 4.8$, $SD = 1.105$) than those who have less control ($M = 5.46$, $SD = 1.401$), $t(67) = 1.835$, $p = .073$. However, at an alpha level of .05, this statistic proves to not be significant enough to make an accurate statement. This difference in findings suggest that these participants do not necessarily see advertisements to be invasive until they are forced to stop using sites due to presence of the advertisements.

9.1.4 Invasive Ads vs. Societal Concerns

After running all of our analysis, we were unable to find any significant associations between the amount of invasive advertisements present and increased levels of societal concern. Unlike the previous report, our test statistics had significant levels that were much higher than our alpha of 0.05, and therefore is not worth reporting.

9.2 Revised Survey Results

Although there were differences between the two different surveys that got sent out, we still were able to get some meaningful data. We carried out the same coding procedure for the

qualitative metric to see what our new participants thought invasive ads were. The responses were similar compared to the findings of our core survey.

Once again, we repeated the same qualitative coding process for the questions regarding what invasive ads looked like. The breakdown was almost the same for both the experimental and alternative surveys. An interesting deviation is that psychological state did not appear as frequently as did financial situation and political views. 79% of responses mentioned location, 81% mentioned identifiable information, 72% addressed personal interests, 57% discussed financial situations and 31% touched upon political affiliation. This difference might go to show what exactly is the key difference in the concerns facing the two sets of populations. These results also included the data from the other question regarding common characteristics of invasive ads.

9.2.1 Level of Control vs. Presence of Invasive Ads

Since these results indicated that, once again, we had the same working definition as our respondents, we could carry on with the rest of our analysis. This survey had no data pertaining to the Google site we asked our experimental group to visit, therefore we have to focus on data relating to questions such as sense of control, characteristics of current advertisements and site usage terminations.

We analyzed the results comparing the means of the level of control users have while browsing versus the amount of invasive ads that users see. Using the same questions and methods as before, we were able to see that those who stopped using a smaller number of sites ($M = .73$, $SD = .45$) had a higher perceived level of control than those who reported the termination of a larger number of sites ($M = 2.55$, $SD = 1.01$), $t(75) = 2.264$, $p = .013$. An interesting fact about these findings is that the means and spreads are drastically different from the original survey. The data suggests that the faculty and staff population seems to be more drastic with their responses. Either they feel in control and do not stop using any sites or they feel a strong lack of control and stop using many sites. The standard deviations suggest little responses in between. One confound that might account for this variation is the level of understanding of modern technology for the Potpourri population. Those that understand technology better might feel like they have more control over what sites can gather about them. This idea might be a great idea for future work to look at.

It came as a surprise to us that we replicated our results when comparing levels of control with seeing more invasive advertisements currently. Those that reported a higher level of control seemed to not report seeing a lower level of invasiveness in advertisements ($M = 4.11$, $SD = 1.678$) than those who have less control ($M = 4.67$, $SD = 1.122$), $t(75) = 1.259$, $p = .106$. Once again, these results could just be the product of the levels at which the Potpourri population understand modern technology. It could be a possibility that those who see more ads and have perceived understanding of technology might appraise the advertisements differently in a way

that makes them less invasive. This idea could also be another topic that could be explored more in depth.

9.2.2 Invasive Ads vs. Societal Concern

An interesting comparison arose from this difference as well. After running more independent T-tests to make mean comparisons, we were able to see that those who indicated financial situations or political affiliation ads to be invasive reported higher levels of believing that invasive ads were a problem to society ($M = 6.33$, $SD = 1.21$) as compared to those who did not indicate those characteristics to be invasive ($M = 3.671$, $SD = 1.57$), $t(75) = 3.302$, $p = .0007$. These results also support the fact that there are key factors in which older populations see ads to be an issue when judging their invasiveness.

9.2.3 Possible Response Bias

One main, known, limitation that we must discuss as part of our revised survey is the possibility for response and non-response bias. We sent our survey out to an email list to faculty and staff as previously discussed. Anyone on this list has the option to either ignore the email, open the survey and not respond, open the survey and answer part of it, or see the email and wait until they have time to take it. They are not receiving any credit, nor are we requiring any sign-up to take our abridged survey. Those who have completed the survey on time are those that most likely find the topic to be important or interesting. For this reason, as well as the lack of manipulation, we are unable to say for certain that any of the results that we gathered for our revised survey are results that can be applied more generally. More research needs to be done in this area for that to occur.

We do not have to worry about response bias for the subject pool. Every participant from the subject pool is taking the survey for credit and receiving the same benefit. Participants have an incentive to take every survey honestly to ensure they are not docked points on their final grade. There also is a time limit for the survey that the subjects are aware of; they will not receive the credit if they do not complete the survey before a certain time limit of our choosing.

9.3 Discussion and Limitations

Some additional limitations that we ran into along the way include self-report bias and representation issues. We had to rely on our participants to give us accurate and true answers. We also had to take into consideration our own confidence that we were asking appropriate questions based on the level of knowledge and experience of our participants which would lead us to believe that the participants will be able to accurately understand and respond to our questions. We have no way of preventing or indicating lying by our participants. This being said, we did provide guaranteed anonymity in our informed consent, and therefore, we have little reason to believe that there would be any incorrect responses.

Additionally, we had an issue of participants not being representative of our population. WPI is a technology school that prides itself on innovation. Those that work or learn here are clearly not representative of the general population of those the same age. A survey, with a topic such as ours, needs to have a wider audience and representation from other demographics. For this reason, as stated in Section 9.2.3, we can not say that our survey has high generalization levels.

Besides these two limitations, we ran into the issues previously stated when creating our survey. We lacked any demographics data for the abridged survey due to the initial demographics being geared towards students. We were also unable to specifically see what our participants were doing when visiting the Google site. Since we were not hosting our own, we have no way of knowing how much time participants spent exploring their own data. Some of the participants could have briefly glanced at the site while others took time to scroll through and observe the data more in-depth. This extra analysis of their data might have skewed the data as they found more data that did not specifically align with their interests as much as others on the list. Finally, we might have had possible response bias occur as discussed in Section 9.2.3.

9.4 Summary

The results of our two surveys were similar with some clear deviations from each other. There was one main difference between what the two different sets of populations determine is invasive, that being psychological state for undergraduates versus financial status and political affiliation for the faculty and staff population. Going more in depth on this finding might be a great place to start for future work. Our main research question did not have a concrete association, however, these findings do not mean that we can say one does not exist. We are unable to comment on if the presentation with personalized data would affect participants willingness to switch to privacy-protection technologies. We found some significant results between other measures that possibly suggest that the level of control users have while browsing might be a great place to carry on with future research. Associations between some variables in the faculty and staff population might be able to be mediated by the levels at which adults understand modern technology, however, future work would need to be conducted in order to make this claim is certain.

10. Conclusion

We started off by conducting research from previous surveys and projects and found that users are increasingly concerned with their own privacy on the Internet. Many websites have built in tracking technologies, such as cookies, that record what users click on and the searches they make. Many sites then build human profiles on their users which companies use as part of their business models. To combat this issue, there are many different tools that exist that make ‘privacy-protection’ claims, promising to protect their users. Our goal was to evaluate these claims to see which of these tools live up to their name. We provided a recommendation to the users that consists of a combination of several tools that do not break the main website features and prevent sites from gathering information about people in ways depending on their privacy concerns determined by a survey we developed.

We evaluated whether different browsers and extensions provide better privacy than others. We used the Selenium library and Python script to drive through 100 sites with a proxy and captured all traffic. We measured the amount of privacy protection offered, by assigning each domain a category and we decided whether certain browsers and extensions block more cookies and scripts than others.

Upon the conclusion of our testing, we found that Brave and Firefox in Strict mode are the best browsers in terms of tradeoff between percent of websites with degradation versus percent trackers remaining. uBlock Origin, Ghostery and Privacy Badger are the best browser extensions in terms of the same tradeoff. Most of the tools that we studied with the exception of NoScript and Firefox in Custom mode did not cause a lot of webpage degradation. Adblock and Adblock Plus are also acceptable options for users that do want to support ad whitelisting programs, although they perform significantly worse in terms of reducing the amount of trackers.

To make a more comprehensive recommendation, we extended our research to mobile applications and search engines. Since we were unable to find similar work in the field, we attempted to pioneer our own methodology to test privacy claims of mobile applications and search engines based on the methodology of related topics described in the main body of the work. For mobile applications, we did not find significant differences between desktop versions and mobile versions of the same apps. Brave performed the best, followed by AdGuard and Disconnect, with Adblock and Adblock Plus in the middle and Chrome performing the worst. For search engines, DuckDuckGo showed the best performance, followed by SwissCows. Excellent performance of DuckDuckGo is not surprising given that the platform makes a lot of privacy claims.

We originally tried to measure the validity of the privacy claims each tool makes. We found that while extensions performed mostly consistent with their privacy claims, the browsers and search engines did not. The worst offender on our list was Vivaldi which claims to be “fast, private and secure browser that blocks ads and trackers” (*Vivaldi Browser*, n.d.). Based on our

research, it performed worse than Chrome, allowing the largest number of trackers through. Opera also performed rather poorly, however, after being sold to Chinese consortium in 2016, they do explain in their privacy policy that the browser collects a lot of data on the users and only safety claims it makes are in regards to the VPN they offer (Taylor, 2020). Search engine wise, Privado and Qwant did not perform well, performing worse than Bing which makes no privacy claims. Privado's performance is especially concerning, given as we have discussed before it makes a lot of privacy claims. Worth noting though that like discussed before, our methodology for search engine testing has a lot of limitations so further testing is needed to confirm Privado's and Qwant's performance.

The main body of this work lays groundwork for the survey portion of study. We wanted to make use of Google AdSetting website in order to begin to evaluate participants' perceptions of privacy protections that they have while browsing the Internet. The main research question for the survey portion was to study whether viewing personal data causes participants to switch to our recommendations. While this question proved to show a ceiling effect - almost everyone indicated the willingness to switch technologies - other analyses proved to be significant. Our results suggest that those that reported having a lower level of control was associated with two different findings. The first being that they reported feeling as though Google would be more accurate with synthesizing personal information about them. The second proved that the level appears to be associated with the amount of sites they have stopped using due to the presence of invasive advertisements. Both of our validity checks for Google and characteristics of invasive advertisements turned out to be successful, revealing a key difference between groups. We were able to explore that difference further to show a possible mediation in the faculty and staff population for what types of advertisements they found invasive on the level at which they believed invasive ads play a role in being detrimental to society. We unfortunately, however, can not say this claim for certain given the lack of an experimental manipulation required to make such assumptions.

11. Future Work

Given the rapidly changing landscape of privacy on the Internet and the growing user concerns about privacy, it is likely that more and more tools will emerge on the market that claim to be privacy-oriented. We speculate that while some tools such as uBlock Origin, Ghostery, Firefox, Privacy Badger and other established tools will remain on the market, newer competitor tools will emerge as well, as we have seen with Brave for example which came out in 2019. The established tools will need to be tested again as they are updated to see whether they continue to offer the best protection. New tools will require the same type of testing we conducted in order to validate privacy claims. This testing might require the sample size of the number of websites tested to be increased from the Alexa Top 100, which would gain more confidence in the results. We also believe that further work also needs to be done in regards to categorizing third-party objects. It is not sufficient to say that a certain tool blocks a certain percentage of traffic; we need to have a breakdown by category and the effectiveness of tools be analyzed based on its ability to prevent advertising and tracking, while maintaining website integrity. Automated processes need to be developed for website degradation testing since it is not sufficient to say how effective a particular tool is at blocking traffic; we need to know the impact it causes on usability and best tools need to be picked based on the tradeoff between effectiveness and impact as we conducted in our project. Perhaps with the improvement of artificial intelligence and machine learning, the process of analyzing the website to determine degradation automatically could be achieved, however, more developments are needed in that area of technology.

A dedicated project could be run to study privacy tools available on mobile. Based on our group's research, there is a significant lack of data available that studies the effectiveness of privacy tools on mobile platforms. The lack of data poses a problem given how widespread and popular mobile and portable devices are becoming. While we did not find significant differences between the tools of the same name that were available on desktop and mobile, further testing needs to be done to confirm that theory with more data and make sure it stays the same if it is true. Mobile research was not part of our main study, so the methodology behind it was not as comprehensive as the study regarding browsers and browser extensions. We do recommend a dedicated project because of the increasing difficulty of performing traffic capture on mobile platforms. For future groups that do want to undertake that work, it would help them to first establish whether they can run traffic capture tools on a mobile device and capture HTTPS traffic. Then the remaining portion of the project could be spent trying to automate testing and write code specifically to test applications on mobile.

We believe there is an urgent need for a dedicated project studying privacy claims of search engines. We were not able to find any work specifically related to how much data search engines collect on users, but we got our idea from a research paper by Wills and Tatar from 2012.

Wills and Tatar attempted to study what ads web advertisers show to users based on sites visited and we performed a similar study, except only based on what users searched for in search engines. Given the lack of existing methodologies in other research papers regarding search engines, we designed our methodology almost from scratch, only relying on theoretical ideas previously proposed (Wills & Tatar, 2012). During our testing, we ran into many issues that should have been more thoroughly thought through and accounted for in the methodology. Therefore, our confidence in our findings is not as high as it could be with further testing and better methodology. Given what we were working with, however, we still believe our work lays important groundwork for future research. Automation needs to be applied to this process as well and the number of websites tested needs to be increased significantly before concrete conclusions can be drawn. Same as with degradation testing, perhaps with improvement in machine learning and artificial intelligence, the process of analyzing which ads are related to which terms could be automated.

One limitation to our work is our focus on Windows on desktop and Android on mobile. This limitation was largely due to the significant challenges that MacOS and iOS pose in terms of testing. Many of the tools that we used for testing in our project would either be completely unavailable on Apple or severely limited. Chrome Drivers that we used in our Python code for Selenium are only available for Windows. Fiddler Classic that we ended up using because of its rich functionality is also not an option, and researchers would have to use Fiddler Everywhere which is a good option, but more limited. Future work could study the effectiveness of Safari to block intrusive third-party elements and study a set of Apple exclusive extensions.

There are some steps that we could have done differently in terms of carrying out our survey research that would have been able to give us more reliable data to measure. If we were able to replicate the Google AdSetting website, we would be able to measure specifically how long participants took in the site and record what they interacted with and for how long. This site would give us more insight to what specifically about their own data users would care about. Instead, we had to give specific directions to our participants, giving us control, with a tradeoff of freedom for participants to give us indications of their perceptions.

If our surveys were replicated, there would be a need to rerun the revised version with a true demographics section in order to be able to confidently give statistics of the alternative demographics surveyed. Additionally, this change would allow for future work to focus more on differences in what types of characteristics of invasive advertisements that both groups identified. That key finding needs to be explored to see if it can be replicated - with a population that represents more of the common adult and not faculty and staff at a university - and take steps to concluding why these differences in findings exist.

The levels at which our faculty and staff participants understand modern technology might have accounted for some of the results that we found in the analysis of our revised survey. These results should be looked at more in depth to see if these findings are true. Having this

information could help developers make informed decisions when creating tools for the general public to use and give us key insights to increasing general levels of privacy for all users.

We finally suggest diving deeper into the discrepancy of the representativeness of our participants. WPI students and faculty are not representative of all the general population of the ages we were targeting. If the study were to be replicated, this limitation needs to be explored further. The abridged survey will continue to be active for anyone who wants to make use of it. Our work will be continued to be distributed to help educate anyone, who is willing to take the time, on their privacy. Our team has already received emails and questions from the faculty and staff about the tools we recommended in our survey. It is clear that this system is in demand and is already beginning to aid others in protecting their information.

References

1. *About Adblock Plus*. (n.d.). Adblock Plus. <https://adblockplus.org/en/about#monetization>
2. Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., & Diaz, C. (2014). The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. *In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 674–689. https://securehomes.esat.kuleuven.be/~gacar/persistent/the_web_never_forgets.pdf
3. *Acceptable Ads FAQ*. (n.d.). Adblock. <https://getadblock.com/acceptable-ads-faq/>
4. Achara, J., Parra-Arnau, J., & Castelluccia, C. (2016). *MyTrackingChoices: Pacifying the Ad-Block War by Enforcing User Privacy Preferences*. <https://hal.inria.fr/hal-01302613/file/paper.pdf>
5. *AdBlock*. (n.d.). Adblock. <https://getadblock.com/>
6. *AdGuard AdBlocker*. (n.d.). <https://chrome.google.com/webstore/detail/adguard-adblocker/bgnkhhnamicmpeenaelnjfhikgbklg>
7. *Allowing acceptable ads in Adblock Plus*. (n.d.). Adblock Plus. <https://adblockplus.org/en/acceptable-ads>
8. Anglim, C. (2016). *Privacy Rights in the Digital Age*, Grey House Publishing. ProQuest Ebook Central. <http://ebookcentral.proquest.com/lib/wpi/detail.action?docID=4454671>.
9. Bashir, M., Farooq, U., Shahid, M., Zaffar, M., & Wilson, C. (2019). Quantity vs. Quality: Evaluating User Interest Profiles Using Ad Preference Managers. *Network and Distributed Systems Security (NDSS) Symposium 2019*. <https://www.ahmadbashir.com/static/pdf/bashir-ndss19.pdf>
10. Bauer, C., & Strauss, C. (2016). Location-based advertising on mobile devices. *Management Review Quarterly*, 66, 159–194. <https://link-springer-com.ezpxy-web-p-u01.wpi.edu/article/10.1007/s11301-015-0118-z>
11. Borgolte, K., & Feamster, N. (2020). Understanding the Performance Costs and Benefits of Privacy-focused Browser Extensions. *Proceedings Of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan. ACM, New York, NY, USA,*. <https://kevin.borgolte.me/files/pdf/www2020-privacy-extensions.pdf>
12. *Configure Fiddler for Android*. (n.d.). Telerik Fiddler. <https://docs.telerik.com/fiddler/configure-fiddler/tasks/ConfigureForAndroid>
13. Cyphers, B. (2019, July 17). *Sharpening Our Claws: Teaching Privacy Badger to Fight More Third-Party Trackers*. Electronic Frontier Foundation. <https://www.eff.org/deeplinks/2019/07/sharpening-our-claws-teaching-privacy-badger-fight-more-third-party-trackers>

14. *Differences between Iridium and Chromium*. (2019, June 7). GitHub.
<https://github.com/iridium-browser/tracker/wiki/Differences-between-Iridium-and-Chromium>
15. *Disconnect Help*. (n.d.). Disconnect. <https://disconnect.me/help>
16. *DuckDuckGo Help - Sources*. (n.d.). DuckDuckGo.
<https://help.duckduckgo.com/duckduckgo-help-pages/results/sources/>
17. *DuckDuckGo Privacy Essentials*. (n.d.). DuckDuckGo. <https://duckduckgo.com/app>
18. Englehardt, S. (2014). The hidden perils of cookie syncing. *Princeton's Center for Information Technology Policy*.
<https://freedom-to-tinker.com/2014/08/07/the-hidden-perils-of-cookie-syncing/>
19. *Fiddler Classic*. (n.d.). Telerik Fiddler. <https://www.telerik.com/download/fiddler>
20. *Fiddler Everywhere*. (n.d.). Telerik Fiddler.
<https://www.telerik.com/download/fiddler-everywhere>
21. *Fiddler Extensions: Privacy Add On*. (n.d.). Telerik Fiddler.
<https://www.telerik.com/fiddler/add-ons>
22. *Fiddler - View Cookie Information*. (n.d.). Telerik Fiddler.
<https://docs.telerik.com/fiddler/Observe-Traffic/Tasks/ScanCookies>
23. *Filter List for AdGuard*. (n.d.). Github.
<https://github.com/hl2guide/Filterlist-for-AdGuard>
24. Fink, A. (2009). *How to conduct surveys: A step-by-step guide*. Los Angeles: SAGE.
25. Hanson, J., Wei, M., Veys, S., Kugler, M., Strahilevitz, L., & Ur, B. (2020). Taking Data Out of Context to Hyper-Personalize Ads: Crowdworkers' Privacy Perceptions and Decisions to Disclose Private Information. *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.
<https://www.blaseur.com/papers/robotext-full.pdf>
26. Harris, S., & Maymi, F. (2018). *CISSP All-in-One Exam Guide, Eighth Edition (8th ed.)*. McGraw-Hill Education.
27. Hill, K. (2016, March 31). How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did. Retrieved October 13, 2020, from
<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>
28. Holding, J. (2018, September 5). Bypass Certificate Pinning on Android. *Jamie Holding Blog*. <https://blog.jamie.holdings/2018/09/05/bypass-certificate-pinning-on-android/>
29. *How does Qwant ensure my security?* (n.d.). Retrieved October 14, 2020, from
<https://help.qwant.com/help/overview/security/how-does-qwant-ensure-my-security/>
30. *How does Qwant index the web ?* (n.d.). Qwant. Retrieved October 14, 2020, from
<https://help.qwant.com/help/overview/how-does-qwant-index-the-web/>
31. *How does Qwant make money?* (2017, October 1). Qwant.
<https://help.qwant.com/help/overview/how-does-qwant-make-money/>

32. Hrapsky, C. (2019, January 27). *The Target app price switch: What you need to know*. Kare 11.
<https://www.kare11.com/article/money/consumer/the-target-app-price-switch-what-you-need-to-know/89-9ef4106a-895d-4522-8a00-c15cff0a0514>
33. *HTTPS Everywhere*. (n.d.). Electronic Frontier Foundation.
<https://www.eff.org/https-everywhere>
34. Iqbal, P., Baba, A., & Bashir, A. (2016). Comprehensiveness, Dead Links and Duplicacy of Select Major Search Engines in the Field of Library and Information Science. *International Journal of Library Science and Research*, 6(4), 1–10.
<https://www.scribd.com/document/324897589/1-IJLSR-Comprehensiveness-Dead-Links-and-Duplicacy-of-Select-Major1>
35. Klosowki, T. (2013, January 7). *How Web Sites Vary Prices Based on Your Information (and What You Can Do About It)*. LifeHacker.
<https://lifel hacker.com/how-web-sites-vary-prices-based-on-your-information-an-5973689>
36. Mattioli, D. (2012, August 23). *On Orbitz, Mac Users Steered to Pricier Hotels*. The Wall Street Journal.
<https://www.wsj.com/articles/SB10001424052702304458604577488822667325882>
37. Mazel, J., Garnier, R., & Fukuda, K. (2019). A comparison of web privacy protection techniques. *Computer Communications*, 144, 162–174.
<https://doi.org/10.1016/j.comcom.2019.04.005>
38. Merzdovnik, G., Huber, M., Buhov, D., Nikiforakis, N., Neuner, S., Schmiedecker, M., & Weippl, E. (2017). Block Me If You Can: A Large-Scale Study of Tracker-Blocking Tools. *2017 IEEE European Symposium on Security and Privacy*, 319–333.
<https://ieeexplore-ieee-org.ezpxy-web-p-u01.wpi.edu/document/7961988>
39. Miyazaki, A. D., & Fernandez, A. (2000). Internet Privacy and Security: An Examination of Online Retailer Disclosures. *Journal of Public Policy & Marketing*, 19(1), 54–61.
<https://doi.org/10.1509/jppm.19.1.54.16942>
40. Negi, Y., & Kumar, S. (2014). A Comparative Analysis of Keyword- and Semantic-Based Search Engines. *Intelligent Computing, Networking, and Informatics. Advances in Intelligent Systems and Computing*, 243, 727–736.
https://link-springer-com.ezpxy-web-p-u01.wpi.edu/chapter/10.1007/978-81-322-1665-0_73
41. *Panoptlick 3.0*. (n.d.). Panoptlick 3.0. <https://panoptlick.eff.org/about>
42. Parsania, V., Kalyani, F., & Kamani, K. (2016). A Comparative Analysis: DuckDuckGo Vs. Google Search Engine. *Global Research and Development Journal for Engineering*, 2(1).
https://www.researchgate.net/publication/312626988_A_Comparative_Analysis_DuckDuckGo_Vs_Google_Search_Engine

43. Parser, E. (2011). *Beware online filter bubbles*. TED.
https://www.ted.com/talks/eli_parser_beware_online_filter_bubbles/transcript?referrer=playlist-how_to_pop_our_filter_bubbles
44. Peterson, A. (2015, January 7). *Zombie cookies: How Verizon Wireless's "supercookies" make it even harder to avoid being tracked online* [Washington Post].
https://link-gale-com.ezpxy-web-p-u01.wpi.edu/apps/doc/A397834365/AONE?u=mclin_c_worpoly&sid=AONE&xid=68b9c6f3
45. *Privado - How It Works*. (n.d.). Privado. Retrieved October 14, 2020, from
<https://www.privado.com/about/how-it-works/>
46. Reczek, R. W., Summers, C., & Smith, R. (2016, April 4). *Targeted Ads do not Just Make You More Likely to Buy — They Can Change How You Think About Yourself*. Harvard Business Review.
<https://hbr.org/2016/04/targeted-ads-dont-just-make-you-more-likely-to-buy-they-can-change-how-you-think-about-yourself>
47. Soltani, A., Canty, S., Mayo, Q., Thomas, L., & Hoofnagle, C. (2009). *Flash Cookies and Privacy*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1446862
48. Sorensen, O. (2013). *Zombie-Cookies: Case Studies and Mitigation*. *The 8th International Conference for Internet Technology and Secured Transactions (ICITST-2013)*. <https://ieeexplore-ieee-org.ezpxy-web-p-u01.wpi.edu/document/6750214>
49. Susser, D., Roessler, B., & Nissenbaum, H. (2018). *Online Manipulation: Hidden Influences in a Digital World*. *Georgetown Law Technology Review 1 (2019)*.
https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3306006
50. *Swisscows - Our Datacenter*. (n.d.). Swisscows.
<https://company.swisscows.ch/en/about/datacenter>
51. *Swisscows - Products*. (n.d.). Swisscows. <https://company.swisscows.ch/en/products>
52. Taylor, S. (2020, May 27). *Secure Browsers That Protect Your Privacy*. Restore Privacy.
<https://restoreprivacy.com/browser/secure/>
53. *Terms of Service, Didn't Read*. (n.d.). <https://tosdr.org/>
54. *The Data Privacy Feedback Loop 2020*. (2020). Transcend.
<https://www.datocms-assets.com/16414/1597336087-transcenddataprivacyfeedbackloop20201.pdf>
55. *Trackers on Alexa Top 500 News sites*. (n.d.). Better FYI. Retrieved October 14, 2020, from <https://better.fyi/trackers/>
56. Turow, J., & Hennessy, M. (2007). Internet privacy and institutional trust. *New Media & Society*, 9(2), 300-318. doi:10.1177/1461444807072219
57. Turow, J., Feldman, L., & Meltzer, K. (2005). Open to Exploitation: American Shoppers Online and Offline. *Annenberg Public Policy Center of the University of Pennsylvania*.
https://cdn.annenbergpublicpolicycenter.org/Downloads/Information_And_Society/Turow_APPC_Report_WEB_FINAL.pdf

58. *uBlock FAQ*. (n.d.). UBlock. <https://ublock.org/faq/>
59. *uBlock Origin*. (n.d.). UBlock Origin. <https://ublockorigin.com/>
60. Uzunoglu, D. (2016). *Understanding Ad Blockers*. *WPI IQP*.
<https://digitalcommons.wpi.edu/cgi/viewcontent.cgi?article=4268&context=mqp-all>
61. Velikov, K. (2019, January 15). How to: Capture Android Traffic with Fiddler. *Progress Telerik*. <https://www.telerik.com/blogs/how-to-capture-android-traffic-with-fiddler>
62. *Vivaldi Browser*. (n.d.). Vivaldi. <https://vivaldi.com/>
63. Wang, Huaiqing, Matthew K. Lee, and Chen Wang (1998), “Consumer privacy concerns about Internet marketing,” *Communications of the ACM*, 41 (3), 63–70.
64. Wass, C. (2018, January 9). Four Ways to Bypass Android SSL Verification and Certificate Pinning. *NETSPI*.
<https://blog.netspi.com/four-ways-bypass-android-ssl-verification-certificate-pinning/>
65. *What’s the difference between AdBlock and Adblock Plus (ABP)?* (2020, January 17). AdBlock.
<https://help.getadblock.com/support/solutions/articles/6000087894-what-s-the-difference-between-adblock-and-adblock-plus-abp->
66. Wills, C., & Tatar, C. (2012). Understanding What They Do with What They Know (Short Paper). *WPES’12*. <https://web.cs.wpi.edu/~cew/papers/wpes12.pdf>
67. Wills, C., & Uzunoglu, D. (2016). *What Ad Blockers Are (and Are Not) Doing*.
<https://web.cs.wpi.edu/~cew/papers/hotweb16.pdf>
68. Wlosik, M., & Sweeney, M. (2020, November 25). *What’s the Difference Between First-Party and Third-Party Cookies?* Clearcode.
<https://clearcode.cc/blog/difference-between-first-party-third-party-cookies/>
69. Zeljkovic, M. (2010). Privacy Awareness of Web Users. *WPI IQP*.
<https://digitalcommons.wpi.edu/cgi/viewcontent.cgi?article=2074&context=mqp-all>

Appendix A: Examples of Minor and Major Website Degradation

Types of minor degradation included:

1. Popup banner asking to allow ads that can be closed



Figure A.1. Example of minor degradation, Adblock on fox.com

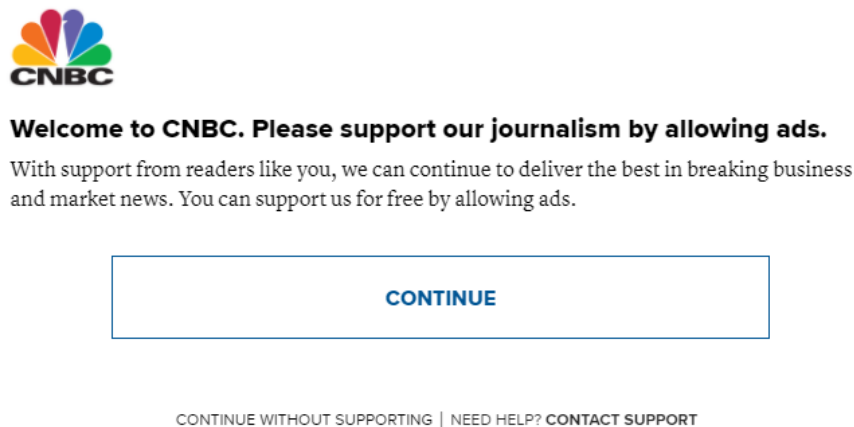


Figure A.2. Example of minor degradation, Adblock on CNBC

2. Empty spaces where ads would have been



Figure A.3. Example of minor degradation, Firefox Custom on washingtonpost.com

3. Minor cosmetic differences (different fonts, different image sizes)

Types of major website degradation include:

1. Completely blank page (in our testing only happens with NoScript and Firefox with all cookies blocked)
2. Mostly blank page with the an error message stating “JavaScript is disabled”

We've detected that JavaScript is disabled in your browser. Would you like to proceed to legacy Twitter?



Figure A.4. Example of major webpage degradation, NoScript on twitter.com

3. Missing major webpage elements that severely affect the visual display of the page

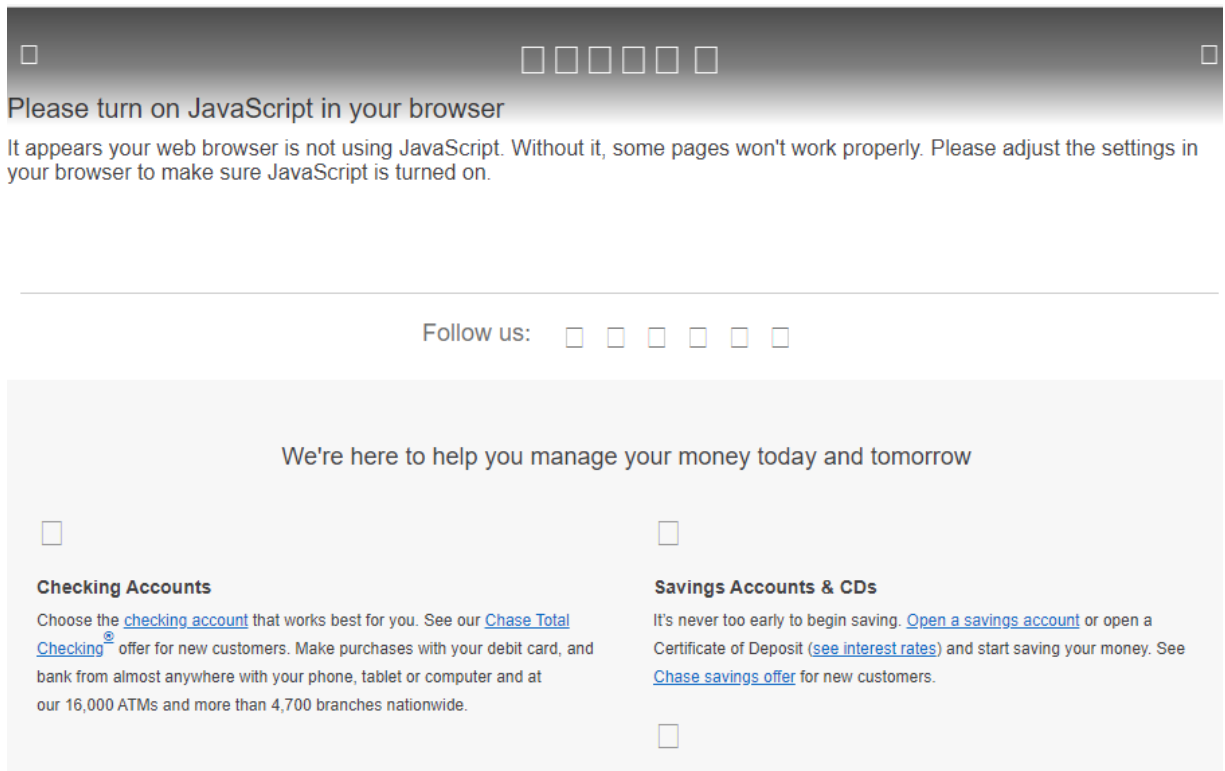


Figure A.5. Major website degradation example, NoScript on chase.com

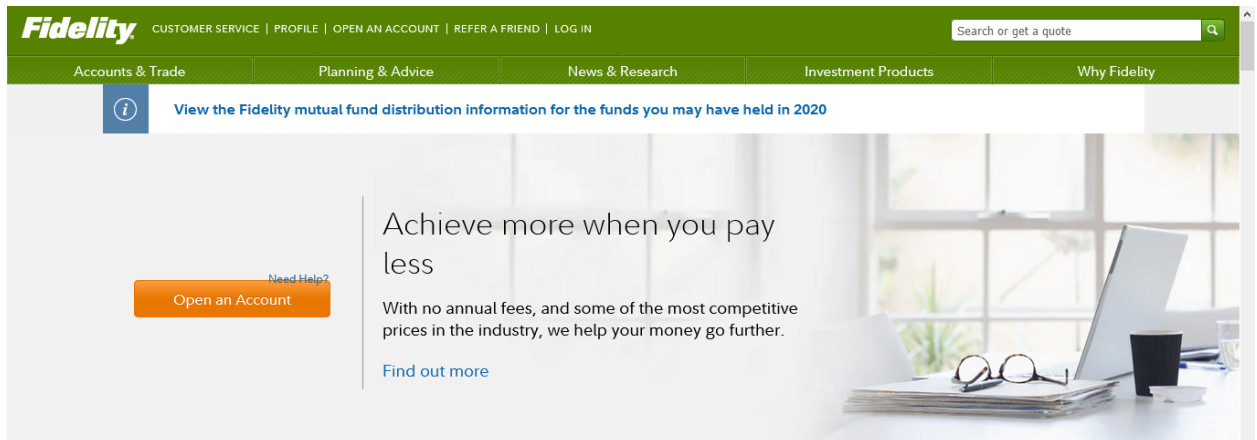


Figure A.6. Example of major degradation, Firefox in Struct mode on fidelity.com missing login fields

4. Popup banner asking to allow the ads that cannot be closed

BUSINESS
INSIDER

It looks like you're using an ad-blocker

Business Insider is an advertising-supported site. Here are two ways you can keep reading:

SUBSCRIBE FOR \$1

TURN OFF YOUR AD-BLOCKER

Have an account? [Log In](#) [LEARN MORE >](#)

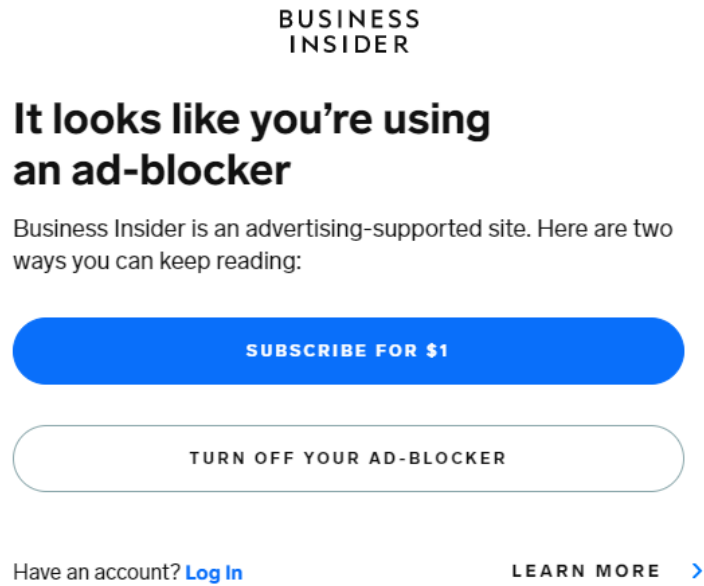
A screenshot of a Business Insider website banner. At the top, the Business Insider logo is displayed in a stacked format. Below the logo, the main heading reads "It looks like you're using an ad-blocker". Underneath this heading, a line of text explains that Business Insider is an advertising-supported site and offers two ways to continue reading. The first option is a prominent blue button with rounded ends that says "SUBSCRIBE FOR \$1". The second option is a white button with rounded ends and a thin grey border that says "TURN OFF YOUR AD-BLOCKER". At the bottom of the banner, there are two links: "Have an account? Log In" and "LEARN MORE >".

Figure A.7. Example of major degradation, AdBlock on businessinsider.com

Appendix B: List of Survey Questions

Internet Privacy Survey

Start of Block: Consent

Q1 Informed Consent Agreement

Primary researcher: Jeffrey Harnois (jharnois@wpi.edu)

Purpose of study: The purpose of this study is to gather information on what you know about privacy while browsing the internet and provide information on how to improve your privacy by adapting or changing your technical setup to utilize more privacy-centered tools.

Procedures to be followed: Please follow all of the directions given to you in this study. Once all of the questions are answered then we will provide you with a recommendation of privacy-centered tools. If applicable, we may ask you to open sites in a new tab during the study. Please view the information in the site we provide and continue with the study.

Time required: You will spend approximately 15 minutes in this study.

Risks to study participants: There are no physical or psychological risks beyond those in everyday life.

Benefits to research participants and others: At the end of this study, there will be a recommendation of tools that we have identified in our testing to be more focused on privacy-protections. You can start to use these tools that we suggest to you and incorporate them while browsing the internet.

Confidentiality: The information that you give will be handled anonymously and confidentially. Your identifying evidence to your responses (i.e your name) will not be used in any report. Only the primary researcher and the faculty advisor will have access to the responses.

Voluntary participation: There will be no payment for this study. Those participating for a class requirement, however, via the Psychology Participant Pool will earn .5 experiment credit for this study. Your participation in this study is completely voluntary. You may stop answering questions at any time and close this tab.

For more information about this research, contact: James Doyle, Department of Social Science, WPI, 100 Institute Rd, Worcester, MA 01609

Phone: (508) 831-5000 x5583

Email: doyle@wpi.edu.

For more information about this research or about the rights of research participants contact:
IRB Manager Ruth McKeogh, Tel. 508 831-6699, Email: irb@wpi.edu and the Human Protection
Administrator Gabriel Johnson, Tel. 508-831-4989, Email: gjohnson@wpi.edu.

Agreement: By clicking "I Agree", you are indicating that you agree to participate in the studies
described above.

Q2 After reading the Informed Consent, do you agree to participate?

I agree (1)

I DO NOT agree (2)

Skip To: End of Survey If After reading the Informed Consent, do you agree to participate? = I DO NOT agree

End of Block: Consent

Start of Block: Gathering info

Q3

The following questions will ask you about your general thoughts on the topic of internet privacy
and what you consider to be an 'invasive advertisement'.

Q9 To what extent are you concerned about invasive advertisements posing a societal issue in
terms of being a violation of individuals' privacy while online?

No concern at all (1)	(2)	(3)	Neutral (4)	(5)	(6)	A very high level of concern (7)
-----------------------------	-----	-----	----------------	-----	-----	--

Level of concern (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-------------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Q4 To what degree do you believe the current online advertisements that you currently see are too invasive?

	Not concerned at all (19)	(20)	(21)	Neutral (22)	(23)	(24)	Extremely concerned (25)
Level of concern (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q5 What characteristics of online advertisement do you find to be invasive?

	Not invasive at all (1)	(2)	(3)	Neutral (4)	(5)	(6)	Extremely invasive (7)
Ads on sites that show you information based on previous searches you made (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ads on sites that show you information based on previous searches you made on other sites (6)



Ads on sites that show you information based on any previous search you made after deleting cookies from your history (9)



Video ads before watching online content (7)



Ads on social media (8)



Pop-up ads that open a new window (10)



Ads that
open an
alert
window
(11)

Q6 In the past 6 months, how many times have you stopped using a website due to an invasive advertisement that you encountered?

- 0 (1)
- 1-2 (2)
- 3-5 (4)
- 6+ (5)

Q7 What types of information do you feel certain websites can gather about you based on your browsing and searches?

Q8 How much control do you think you currently have over what types of information sites can gather about you?

	No control at all (1)	(2)	(3)	Neutral (4)	(5)	(6)	A very high level of control (7)
Level of control (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: Gathering info

Start of Block: Background

Q10 In this section you will answer questions relating to your devices and the technology that you use on a daily basis

Q11 Which of the following devices do you regularly use?

- Smart phone (1)
- Laptop (2)
- Desktop (3)
- Tablet (4)
- Smart watch (5)

Q12 What is your main operating system?

- Windows (1)
- MacOS (2)
- Linux (3)
- Other (4) _____

Q13 What is your main search engine?

- Google (1)
- Bing (2)
- DuckDuckGo (3)
- Other (4) _____

End of Block: Background

Start of Block: Google

Q14 Do you currently have a Google account that you actively use?

To check, visit accounts.google.com

- Yes (1)

No (3)

I'm not sure (4)

End of Block: Google

Start of Block: Control

Q15

The following image contains interests that Google has synthesized about the average WPI student.

Take a 2-3 minutes to observe the interests listed.

Q16 How accurate do these interests seem to be in relation to yours?

	Not accurate at all (1)	(2)	(3)	Neutral (4)	(5)	(6)	Extremely accurate (7)
Level of accuracy (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q17 How accurate do you think Google would be in identifying the following information about you?

	Not accurate at all (1)	(2)	(3)	Neutral (4)	(5)	(6)	Extremely accurate (7)

Age Range (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gender (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: Control

Start of Block: Experiment

Q18

In a **new tab**, please visit the site below and follow the directions below:

<https://adssettings.google.com/authenticated?hl=en>

1. Login with your Google account when prompted
2. If you answered 'yes' to the next question, observe the first two bubbles that appear about gender and age range

Q19 Do you have ad personalization enabled, indicated by the slider saying 'Ad personalization is ON'

Yes (4)

No (5)

Display This Question:

If Do you have ad personalization enabled, indicated by the slider saying 'Ad personalization is ON' = Yes

Q129

Observe the first two bubbles that appear about gender and age range.

Take an additional 2-3 minutes to scroll through the information provided in the bubble subsequent that are related to hobbies.

Once you are finished, come back and answer the following questions below.

Display This Question:

If Do you have ad personalization enabled, indicated by the slider saying 'Ad personalization is ON' = Yes

Q20 Was Google correct in identifying your age range located in the first bubble?

Yes (1)

No (2)

Display This Question:

If Do you have ad personalization enabled, indicated by the slider saying 'Ad personalization is ON' = Yes

Q21 Was Google correct in identifying your gender located in the second bubble?

Yes (1)

No (2)

I do not identify as either binary gender (3)

Display This Question:

If Do you have ad personalization enabled, indicated by the slider saying 'Ad personalization is ON' = Yes

Q22 To what extent do you believe that Google is accurate in the information they have gathered on your hobbies/interests?

	Not accurate at all (1)	(2)	(4)	Neutral (5)	(6)	(7)	Extremely accurate (8)
Level of Accuracy (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: Experiment

Start of Block: Browsers

Q23 You will now be asked questions about your default browser that you usually use to browse the internet with.

Q24 What is your default browser?

- Chrome (1)
- Firefox (2)
- Brave (3)
- Other (4) _____

Display This Question:

If What is your default browser? != Brave

Q25 Are you willing to switch your default browser in order for you to obtain better internet privacy protections?

- Yes (1)
- No (2)
- Not sure (3)

Display This Question:

If What is your default browser? = Firefox

Q26

In Firefox, there are different privacy settings that you can select that help protect against some intrusive ads and trackers that websites are running.

The image below is an example of Firefox settings that show what type of tracking protections that are available.

View your settings at here:

`about:preferences#privacy`

Do you run your browser with settings that are different than 'standard'?

- Yes (1)
- No (2)

Display This Question:

If What is your default browser? = Chrome

Q27 In Chrome, there are different privacy settings that you can select that help protect against some intrusive ads and trackers that websites are running.

The image below is an example of Chrome settings that show what type of tracking protections that are available.

View your settings here:

`chrome://settings/cookies`

Do you run your browser with settings that are different than 'Allow all cookies'?

Yes (1)

No (2)

Display This Question:

If In Firefox, there are different privacy settings that you can select that help protect against so... = Yes

Q28 What settings do you run?

Strict (1)

Block headers (2)

Block images (3)

Custom (4)

Display This Question:

If In Chrome, there are different privacy settings that you can select that help protect against som... =
Yes

Q29 What settings do you run?

- Block third-party cookies (1)
- Block all cookies (2)

End of Block: Browsers

Start of Block: Extensions

Q30

In this section, you will be asked about extensions that you might have running on your primary browser. To check your extensions, follow these steps

Chrome

1. Visit chrome://extensions in a new tab
2. View all of your current active extensions

Firefox

1. Visit about:addons in a new tab
2. Click on the 'extensions' tab on the left side menu
3. View all of the 'active' extensions

Some common examples of extensions include:

Adblock
Dark Reader
Screencastify
Grammarly

Q31 How many extensions do you usually run that are related to privacy and ad-blocking?

- 0 (1)
- 1-2 (2)
- 3-5 (3)
- 6+ (4)

Display This Question:

If How many extensions do you usually run that are related to privacy and ad-blocking? = 0

Q32 Would you be willing to start using certain extensions on your browser to help give you specific internet privacy protections?

- Yes (1)
- No (2)
- I'm not sure (5)

Display This Question:

If Would you be willing to start using certain extensions on your browser to help give you specific... = Yes

Or Would you be willing to start using certain extensions on your browser to help give you specific... = I'm not sure

Q33

Which of these topics would you care more about and would be willing to use an extension that provided these protections?

Ad blocking extensions help to mitigate the ads that appear on a page and prevent them from appearing.

Privacy protection extensions help reduce website's cookies from tracking you from site to site and stop websites from targeting you based on your traffic and search results.

- Ad-blocking (1)
- Privacy protection (2)
- Both (3)

Display This Question:

If Which of these topics would you care more about and would be willing to use an extension that pro... = Both

Or Which of these topics would you care more about and would be willing to use an extension that pro... = Ad-blocking

Q34 Which types of ads would you be more inclined to block?

- All ads in general (1)
- Only intrusive ads (2)

Display This Question:

If How many extensions do you usually run that are related to privacy and ad-blocking? != 0

Q35 What type of extensions do you run?

- Ad-blocking (1)

Privacy protection (2)

Both (3)

Display This Question:

If What type of extensions do you run? = Both

Or What type of extensions do you run? = Ad-blocking

Q36 How satisfied are you with how much your extension blocks ads?

Extremel
y
dissatisfi
ed

Neither
satisfied
nor
dissatisfi
ed

Extreme
ly
satisfied

0 1 2 3 4 5 6 7

Level of satisfaction ()

Display This Question:

If How satisfied are you with how much your extension blocks ads? [Level of satisfaction] >= 5

Q37 What types of ads does your extension block?

All ads in general (1)

Only blocks intrusive ads (2)

Display This Question:

If How satisfied are you with how much your extension blocks ads? [Level of satisfaction] < 5

Q38 What is something that you wish your extension did better?

- Block more ads in general (1)
- Block more intrusive ads (2)

Display This Question:

If Would you be willing to start using certain extensions on your browser to help give you specific... = I'm not sure

Q39 Would you be willing to switch to the type of extension that you indicated caring about?

- Yes (1)
- No (3)

End of Block: Extensions

Start of Block: Intentions

Display This Question:

If How many extensions do you usually run that are related to privacy and ad-blocking? != 0

Q40 Are you willing to switch the extensions?

- Yes (1)
- No (2)
- I'm still not sure (4)

Display This Question:

If What settings do you run? = Block all cookies

Or In Chrome, there are different privacy settings that you can select that help protect against som...
= No

And If

In Firefox, there are different privacy settings that you can select that help protect against so... = No

Or What settings do you run? != Strict

Q41 Are you willing to switch the configuration on the browser that you are using?

Yes (1)

No (2)

I'm still not sure (4)

End of Block: Intentions

Start of Block: Demographics

Q126

These next questions focus on your demographics

Q61 What gender do you identify as?

Male (1)

Non-binary (5)

Female (2)

Prefer not to say (4)

Self-describe (3) _____

Q62

How old are you?

18 19 20 22 23 24 25 26 28 29 30

	Age ()	
--	--------	--

Q63 What is your projected graduation year?

2021 (1)

2022 (2)

2023 (3)

2024 or later (4)

Q64 What is your major?

Mechanical Engineering/Aerospace Engineering (4)

- Biomedical Engineering (12)
- Biology/Bio Tech (11)
- Biochemistry (13)
- Buisness (9)
- Chemistry/Chemical Engineering (5)
- Civil/Architectural engineering (18)
- Computational Biology (17)
- Computer/Data Science (6)
- Electrical/Computer Engineering (7)
- Humanities (19)
- Industrial Engineering (20)
- Mathematical Sciences (14)
- Physics (16)
- Psychological Science (21)

- Robotics Engineering (15)
- Other Social Science (8)
- Other (10) _____

Q127 Are you Hispanic, Latino/a, or Spanish origin?

- Yes, Hispanic, Latino/a, or Spanish origin (4)
- No, not of Hispanic, Latino/a, or Spanish origin (5)

Display This Question:

If Are you Hispanic, Latino/a, or Spanish origin? = Yes, Hispanic, Latino/a, or Spanish origin

Q128 Which group best describes you?

- Mexican, Mexican American, Chicano (1)
- Puerto Rican (2)
- Cuban (3)
- Another Hispanic, Latino, or Spanish Origin (4)

Q65

What is your race? Mark one or more.

- White (1)
- Black or African American (2)
- American Indian or Alaska Native (3)
- Asian Indian (4)
- Chinese (5)
- Filipino (6)
- Japanese (8)
- Korean (9)
- Vietnamese (10)
- Other Asian (11)
- Native Hawaiian (12)
- Guamanian or Chamorro (13)
- Samoan (14)



Other Pacific Islander (15)

End of Block: Demographics

Start of Block: Recommendations

Display This Question:

If What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy... != Yes

And Are you willing to switch the configuration on the browser that you are using? != Yes

And What types of ads does your extension block? , All ads in general Is Displayed

Or If

What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy... != Yes

And Are you willing to switch the configuration on the browser that you are using? != Yes

And What type of extensions do you run? = Privacy protection

Or If

What is your default browser? != Firefox

And What is your default browser? != Brave

And Are you willing to switch your default browser in order for you to obtain better internet privacy... != Yes

And Are you willing to switch the configuration on the browser that you are using? != Yes

And Which of these topics would you care more about and would be willing to use an extension that pro... = Privacy protection

Q117

We have the following recommendations!

In our research, we have ran extensive testing on all types of available browsers, extensions, and browser configurations to see which ones provided the best protections against website tracking. Below are the suggestions that we have, based on your responses, that we feel would

benefits you the most while online. All of the tools being recommended have proven to be great ways to increase your browsing protections.

Privacy Extension

If you are looking for a privacy-focused extension, we strongly recommend *Ghostery*. Ghostery has proven to block a lot of sites from gathering certain types of information on you.

Ad-Blocking Extension

If ad blocking is something that you are interested in, we recommend using either the *AdBlock* extension or the *uBlock Origin* extension. AdBlock contains a list of sites that they deem 'non-intrusive' and only subjects the users to ads they believe are safe. uBlock Origin, on the other hand, block all ads for the user.

Extension Stores

The following links are to the Firefox and Chrome (respectively) stores where you can search for the extensions just mentioned above.

For Firefox - <https://addons.mozilla.org/en-US/firefox/>

For Chrome - <https://chrome.google.com/webstore/category/extensions>

Alternative Browsers

Most internet users prefer Google Chrome as their default browser. Chrome, however, performs poorly at providing their users with protections from sites that are trying to track their information or show them invasive ads. Firefox is similar, but has a few improvements that make it the better option.

Alternative Chrome Settings

There are simple things that you can do within Chrome itself to increase your privacy-protections slightly. You can follow these instructions in order to enable the *blocking of 3rd party cookies*.

[Check out how to enable the setting here](#)

Alternative Firefox Settings

You can follow these instructions in order to enable the *Strict mode* in Firefox to add a strict list of rules to protect you while browsing. The strict setting on Firefox has proven to be extremely effective in blocking most sites from using their techniques to track you.

[Check it out here](#)

Brave

However, if you did not want to change any configurations of your browser, you can switch to the *Brave Browser*. Brave proved itself to be the best, out-of-the-box browser that provides its users with privacy protections. We recommend switching to Brave if you do not want to add any additional tools onto a browser to help add protections.

[Brave can be found here](#)

Display This Question:

If What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy... =
No

And Are you willing to switch your default browser in order for you to obtain better internet privacy... =
Not sure

And Are you willing to switch the configuration on the browser that you are using? != Yes

And What is something that you wish your extension did better? = Block more intrusive ads

Or If

Are you willing to switch the configuration on the browser that you are using? = No

And What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy... =
No

And Are you willing to switch your default browser in order for you to obtain better internet privacy... =
Not sure

And Which types of ads would you be more inclined to block? = Only intrusive ads

And Would you be willing to switch to the type of extension that you indicated caring about? = Yes

Q116

We have the following recommendations!

In our research, we have ran extensive testing on all types of available browsers, extensions, and browser configurations to see which ones provided the best protections against website tracking. Below are the suggestions that we have, based on your responses, that we feel would benefit you the most while online. All of the tools being recommended have proven to be great ways to increase your browsing protections.

Privacy Extension

If you are looking for a privacy-focused extension, we strongly recommend *Ghostery*. Ghostery has proven to block a lot of sites from gathering certain types of information on you.

Ad-Blocking Extension

If ad blocking is something that you are interested in, we recommend using either the *AdBlock* extension. AdBlock contains a list of sites that they deem 'non-intrusive' and only subjects the users to ads they believe are safe.

Extension Stores

The following links are to the Firefox and Chrome (respectively) stores where you can search for the extensions just mentioned above.

For Firefox - <https://addons.mozilla.org/en-US/firefox/>

For Chrome - <https://chrome.google.com/webstore/category/extensions>

Alternative Browsers

Most internet users prefer Google Chrome as their default browser. Chrome, however, performs poorly at providing their users with protections from sites that are trying to track their information or show them invasive ads. Firefox is similar, but has a few improvements that make it the better option.

Alternative Chrome Settings

There are simple things that you can do within Chrome itself to increase your privacy-protections slightly. You can follow these instructions in order to enable the *blocking of 3rd party cookies*.

[Check out how to enable the setting here](#)

Alternative Firefox Settings

You can follow these instructions in order to enable the *Strict mode* in Firefox to add a strict list of rules to protect you while browsing. The strict setting on Firefox has proven to be extremely effective in blocking most sites from using their techniques to track you.

[Check it out here](#)

Brave

However, if you did not want to change any configurations of your browser, you can switch to the *Brave Browser*. Brave proved itself to be the best, out-of-the-box browser that provides its users with privacy protections. We recommend switching to Brave if you do not want to add any additional tools onto a browser to help add protections.

[Brave can be found here](#)

Display This Question:

If What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy...

!= Yes

And Are you willing to switch the configuration on the browser that you are using? != Yes

And What is something that you wish your extension did better? = Block more ads in general

Or If

What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy... = Not sure

And Are you willing to switch the configuration on the browser that you are using? != Yes

And Which types of ads would you be more inclined to block? = All ads in general

And Would you be willing to switch to the type of extension that you indicated caring about? = Yes

Q117

We have the following recommendations!

Privacy Extension

If you are looking for a privacy-focused extension, we strongly recommend *Ghostery*. Ghostery has proven to block a lot of sites from gathering certain types of information on you.

In our research, we have ran extensive testing on all types of available browsers, extensions, and browser configurations to see which ones provided the best protections against website tracking. Below are the suggestions that we have, based on your responses, that we feel would benefits you the most while online. All of the tools being recommended have proven to be great ways to increase your browsing protections.

Ad-Blocking Extension

If ad blocking is something that you are interested in, we recommend using either the *UBlock Origin* extension. UBlock Origin block all ads for the user, no matter how intrusive or non-intrusive the ad may appear.

Extension Stores

The following links are to the Firefox and Chrome (respectively) stores where you can search for the extensions just mentioned above.

For Firefox - <https://addons.mozilla.org/en-US/firefox/>

For Chrome - <https://chrome.google.com/webstore/category/extensions>

Alternative Browsers

Most internet users prefer Google Chrome as their default browser. Chrome, however, preforms poorly at providing their users with protections from sites that are trying to track their information or show them invasive ads. Firefox is similar, but has a few improvements that make it the better option.

Alternative Chrome Settings

There are simple things that you can do within Chrome itself to increase your privacy-protections slightly. You can follow these instructions in order to enable the *blocking of 3rd party cookies*.

[Check out how to enable the setting here](#)

Alternative Firefox Settings

You can follow these instructions in order to enable the *Strict mode* in Firefox to add a strict list of rules to protect you while browsing. The strict setting on Firefox has proven to be extremely effective in blocking most sites from using their techniques to track you.

[Check it out here](#)

Brave

However, if you did not want to change any configurations of your browser, you can switch to the *Brave Browser*. Brave proved itself to be the best, out-of-the-box browser that provides its users with privacy protections. We recommend switching to Brave if you do not want to add any additional tools onto a browser to help add protections.

[Brave can be found here](#)

Display This Question:

If Are you willing to switch your default browser in order for you to obtain better internet privacy... = Yes

And What types of ads does your extension block? , All ads in general Is Displayed

Or If

Are you willing to switch your default browser in order for you to obtain better internet privacy... = Yes

And What type of extensions do you run? = Privacy protection

Or If

Are you willing to switch your default browser in order for you to obtain better internet privacy... = Yes

And Which of these topics would you care more about and would be willing to use an extension that pro... = Privacy protection

Or If

What is your default browser? = Firefox

And What types of ads does your extension block? , All ads in general Is Displayed

Or If

What is your default browser? = Firefox

And What type of extensions do you run? = Privacy protection

Or If

What is your default browser? = Firefox

And Which of these topics would you care more about and would be willing to use an extension that pro... = Privacy protection

Q118

We have the following recommendations!

In our research, we have ran extensive testing on all types of available browsers, extensions, and browser configurations to see which ones provided the best protections against website tracking. Below are the suggestions that we have, based on your responses, that we feel would benefit you the most while online. All of the tools being recommended have proven to be great ways to increase your browsing protections.

Privacy Extension

If you are looking for a privacy-focused extension, we strongly recommend *Ghostery*. Ghostery has proven to block a lot of sites from gathering certain types of information on you.

Ad-Blocking Extension

If ad blocking is something that you are interested in, we recommend using either the *AdBlock* extension or the *uBlock Origin* extension. AdBlock contains a list of sites that they deem 'non-intrusive' and only subjects the users to ads they believe are safe. uBlock Origin, on the other hand, block all ads for the user.

Extension Stores

The following links are to the Firefox and Chrome (respectively) stores where you can search for the extensions just mentioned above.

For Firefox - <https://addons.mozilla.org/en-US/firefox/>

For Chrome - <https://chrome.google.com/webstore/category/extensions>

Alternative Browser settings

Most internet users prefer Google Chrome as their default browser. Chrome, however, performs poorly at providing their users with protections from sites that are trying to track their information or show them invasive ads. Firefox is similar, but has a few improvements that make it the better option.

You can follow these instructions in order to enable the *Strict mode* in Firefox to add a strict list of rules to protect you while browsing. The strict setting on Firefox has proven to be extremely effective in blocking most sites from using their techniques to track you.

[Check it out here](#)

Display This Question:

If Are you willing to switch your default browser in order for you to obtain better internet privacy... = Yes

And What is something that you wish your extension did better? = Block more intrusive ads

Or If

Are you willing to switch your default browser in order for you to obtain better internet privacy... = Yes

And Which types of ads would you be more inclined to block? = Only intrusive ads

And Would you be willing to switch to the type of extension that you indicated caring about? = Yes

Or If

What is your default browser? = Firefox

And What is something that you wish your extension did better? = Block more intrusive ads

Or If

What is your default browser? = Firefox

And Which types of ads would you be more inclined to block? = Only intrusive ads

And Would you be willing to switch to the type of extension that you indicated caring about? = Yes

Q119

We have the following recommendations!

In our research, we have ran extensive testing on all types of available browsers, extensions, and browser configurations to see which ones provided the best protections against website tracking. Below are the suggestions that we have, based on your responses, that we feel would benefit you the most while online. All of the tools being recommended have proven to be great ways to increase your browsing protections.

Privacy Extension

If you are looking for a privacy-focused extension, we strongly recommend *Ghostery*. Ghostery has proven to block a lot of sites from gathering certain types of information on you.

Ad-Blocking Extension

If ad blocking is something that you are interested in, we recommend using either the *AdBlock* extension. AdBlock contains a list of sites that they deem 'non-intrusive' and only subjects the users to ads they believe are safe.

Extension Stores

The following links are to the Firefox and Chrome (respectively) stores where you can search for the extensions just mentioned above.

For Firefox - <https://addons.mozilla.org/en-US/firefox/>

For Chrome - <https://chrome.google.com/webstore/category/extensions>

Alternative Browser Settings

Most internet users prefer Google Chrome as their default browser. Chrome, however, performs poorly at providing their users with protections from sites that are trying to track their information or show them invasive ads. Firefox is similar, but has a few improvements that make it the better option.

You can follow these instructions in order to enable the *Strict mode* in Firefox to add a strict list of rules to protect you while browsing. The strict setting on Firefox has proven to be extremely effective in blocking most sites from using their techniques to track you.

[Check it out here](#)

Display This Question:

If Are you willing to switch your default browser in order for you to obtain better internet privacy... = Yes

And What is something that you wish your extension did better? = Block more ads in general

Or If

Are you willing to switch your default browser in order for you to obtain better internet privacy... = Yes

And Which types of ads would you be more inclined to block? = All ads in general

And Would you be willing to switch to the type of extension that you indicated caring about? = Yes

Or If

What is your default browser? = Firefox

And What is something that you wish your extension did better? = Block more ads in general

Or If

What is your default browser? = Firefox

And Which types of ads would you be more inclined to block? = All ads in general

And Would you be willing to switch to the type of extension that you indicated caring about? = Yes

Q120

We have the following recommendations!

In our research, we have ran extensive testing on all types of available browsers, extensions, and browser configurations to see which ones provided the best protections against website tracking. Below are the suggestions that we have, based on your responses, that we feel would

benefits you the most while online. All of the tools being recommended have proven to be great ways to increase your browsing protections.

Privacy Extension

If you are looking for a privacy-focused extension, we strongly recommend *Ghostery*. Ghostery has proven to block a lot of sites from gathering certain types of information on you.

Ad-Blocking Extension

If ad blocking is something that you are interested in, we recommend using either the *UBlock Origin* extension. UBlock Origin block all ads for the user, no matter how intrusive or non-intrusive the ad may appear.

Extension Stores

The following links are to the Firefox and Chrome (respectively) stores where you can search for the extensions just mentioned above.

For Firefox - <https://addons.mozilla.org/en-US/firefox/>

For Chrome - <https://chrome.google.com/webstore/category/extensions>

Alternative Browser Settings

Most internet users prefer Google Chrome as their default browser. Chrome, however, performs poorly at providing their users with protections from sites that are trying to track their information or show them invasive ads. Firefox is similar, but has a few improvements that make it the better option.

You can follow these instructions in order to enable the *Strict mode* in Firefox to add a strict list of rules to protect you while browsing. The strict setting on Firefox has proven to be extremely effective in blocking most sites from using their techniques to track you.

[Check it out here](#)

Display This Question:

If What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy...

!= Yes

And Are you willing to switch the configuration on the browser that you are using? = Yes

And What types of ads does your extension block? , All ads in general Is Displayed

Or If

What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy... != Yes

And Are you willing to switch the configuration on the browser that you are using? = Yes

And What type of extensions do you run? = Privacy protection

Or If

What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy... != Yes

And Are you willing to switch the configuration on the browser that you are using? = Yes

And Which of these topics would you care more about and would be willing to use an extension that pro... = Privacy protection

Q121

We have the following recommendations!

In our research, we have ran extensive testing on all types of available browsers, extensions, and browser configurations to see which ones provided the best protections against website tracking. Below are the suggestions that we have, based on your responses, that we feel would benefits you the most while online. All of the tools being recommended have proven to be great ways to increase your browsing protections.

Privacy Extension

If you are looking for a privacy-focused extension, we strongly recommend *Ghostery*. Ghostery has proven to block a lot of sites from gathering certain types of information on you.

Ad-Blocking Extension

If ad blocking is something that you are interested in, we recommend using either the *AdBlock* extension or the *UBlock Origin* extension. AdBlock contains a list of sites that they deem 'non-intrusive' and only subjects the users to ads they believe are safe. UBlock Origin, on the other hand, block all ads for the user.

Extension Stores

The following links are to the Firefox and Chrome (respectively) stores where you can search for the extensions just mentioned above.

For Firefox - <https://addons.mozilla.org/en-US/firefox/>

For Chrome - <https://chrome.google.com/webstore/category/extensions>

Alternative Browser Settings

Most internet users prefer Google Chrome as their default browser. Chrome, however, is not very good at providing their users with protections from sites that are trying to track their information or show them invasive ads.

There are simple things that you can do within Chrome itself to increase your privacy-protections slightly. You can follow these instructions in order to enable the *blocking of 3rd party cookies*.

[Check out how to enable the setting here](#)

Display This Question:

If What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy... != Yes

And Are you willing to switch the configuration on the browser that you are using? = Yes

And What is something that you wish your extension did better? = Block more intrusive ads

Or If

Are you willing to switch the configuration on the browser that you are using? = Yes

And What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy... != Yes

And Which types of ads would you be more inclined to block? = Only intrusive ads

And Would you be willing to switch to the type of extension that you indicated caring about? = Yes

Q122

We have the following recommendations!

In our research, we have ran extensive testing on all types of available browsers, extensions, and browser configurations to see which ones provided the best protections against website tracking. Below are the suggestions that we have, based on your responses, that we feel would benefit you the most while online. All of the tools being recommended have proven to be great ways to increase your browsing protections.

Privacy Extension

If you are looking for a privacy-focused extension, we strongly recommend *Ghostery*. Ghostery has proven to block a lot of sites from gathering certain types of information on you.

Ad-Blocking Extension

If ad blocking is something that you are interested in, we recommend using either the *AdBlock* extension. AdBlock contains a list of sites that they deem 'non-intrusive' and only subjects the users to ads they believe are safe.

Extension Stores

The following links are to the Firefox and Chrome (respectively) stores where you can search for the extensions just mentioned above.

For Firefox - <https://addons.mozilla.org/en-US/firefox/>

For Chrome - <https://chrome.google.com/webstore/category/extensions>

Alternative Browser Settings

Most internet users prefer Google Chrome as their default browser. Chrome, however, is not very good at providing their users with protections from sites that are trying to track their information or show them invasive ads.

There are simple things that you can do within Chrome itself to increase your privacy-protections slightly. You can follow these instructions in order to enable the *blocking of 3rd party cookies*.

[Check out how to enable the setting here](#)

Display This Question:

If What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy... != Yes

And Are you willing to switch the configuration on the browser that you are using? = Yes

And What is something that you wish your extension did better? = Block more ads in general

Or If

Are you willing to switch the configuration on the browser that you are using? = Yes

And What is your default browser? != Brave

And What is your default browser? != Firefox

And Are you willing to switch your default browser in order for you to obtain better internet privacy... != Yes

And Which types of ads would you be more inclined to block? = All ads in general

And Would you be willing to switch to the type of extension that you indicated caring about? = Yes

Q123

We have the following recommendations!

In our research, we have ran extensive testing on all types of available browsers, extensions, and browser configurations to see which ones provided the best protections against website tracking. Below are the suggestions that we have, based on your responses, that we feel would benefit you the most while online. All of the tools being recommended have proven to be great ways to increase your browsing protections.

Privacy Extension

If you are looking for a privacy-focused extension, we strongly recommend *Ghostery*. Ghostery has proven to block a lot of sites from gathering certain types of information on you.

Ad-Blocking Extension

If ad blocking is something that you are interested in, we recommend using either the *UBlock Origin* extension. UBlock Origin block all ads for the user, no matter how intrusive or non-intrusive the ad may appear.

Extension Stores

The following links are to the Firefox and Chrome (respectively) stores where you can search for the extensions just mentioned above.

For Firefox - <https://addons.mozilla.org/en-US/firefox/>

For Chrome - <https://chrome.google.com/webstore/category/extensions>

Alternative Browser Settings

Most internet users prefer Google Chrome as their default browser. Chrome, however, is not very good at providing their users with protections from sites that are trying to track their information or show them invasive ads.

There are simple things that you can do within Chrome itself to increase your privacy-protections slightly. You can follow these instructions in order to enable the *blocking of 3rd party cookies*.

[Check out how to enable the setting here](#)

End of Block: Recommendations

Start of Block: No change

Display This Question:

If We have the following recommendations! In our research, we have ran extensive testing on all type... Is Displayed

And And We have the following recommendations! In our research, we have ran extensive testing on all type... Is Displayed

And And We have the following recommendations! Privacy Extension If you are looking for a privacy-focus... Is Displayed

And And We have the following recommendations! In our research, we have ran extensive testing on all type... Is Displayed

And And We have the following recommendations! In our research, we have ran extensive testing on all type... Is Displayed

And And We have the following recommendations! In our research, we have ran extensive testing on all type... Is Displayed

And And We have the following recommendations! In our research, we have ran extensive testing on all type... Is Displayed

And And We have the following recommendations! In our research, we have ran extensive testing on all type... Is Displayed

And And We have the following recommendations! In our research, we have ran extensive testing on all type... Is Displayed

Q42 In our research, we have ran extensive testing on all types of available browsers, extensions, and browser configurations to see which ones provided the best protections against website tracking. Below are the suggestions that we have, that we feel would benefit you the most while online. All of the tools being recommended have proven to be great ways to increase your browsing protections.

It seems as though the answers that you provided indicates that your current set up is the most desirable for you.

However, if you are looking for additional information for tools that we recommend, check out the following tools:

Privacy extension

If you are looking for a privacy-focused extension, we strongly recommend *Ghostery*. Ghostery has proven to block a lot of sites from gathering certain types of information on you.

Ad-Blocking Extension

If ad blocking is something that you are interested in, we recommend using either the *AdBlock* extension or the *uBlock Origin* extension. AdBlock contains a list of sites that they deem 'non-intrusive' and only subjects the users to ads they believe are safe. uBlock Origin, on the other hand, block all ads for the user.

Extension stores

The following links are to the Firefox and Chrome (respectively) stores where you can search for the extensions just mentioned above.

For Firefox - <https://addons.mozilla.org/en-US/firefox/>

For Chrome - <https://chrome.google.com/webstore/category/extensions>

Alternative Browsers

Most internet users prefer Google Chrome as their default browser. Chrome, however, performs poorly at providing their users with protections from sites that are trying to track their information or show them invasive ads. Firefox is similar, but has a few improvements that make it the better option.

Alternative Chrome Settings

There are simple things that you can do within Chrome itself to increase your privacy-protections slightly. You can follow these instructions in order to enable the *blocking of 3rd party cookies*.

[Check out how to enable the setting here](#)

Alternative Firefox settings

You can follow these instructions in order to enable the *Strict mode* in Firefox to add a strict list of rules to protect you while browsing. The strict setting on Firefox has proven to be extremely effective in blocking most sites from using their techniques to track you.

[Check it out here](#)

Brave

However, if you did not want to change any configurations of your browser, you can switch to the *Brave Browser*. Brave proved itself to be the best, out-of-the-box browser that provides its users with privacy protections. We recommend switching to Brave if you do not want to add any additional tools onto a browser to help add protections.

[Brave can be found here](#)

End of Block: No change

Start of Block: Email

Display This Question:

*If In our research, we have ran extensive testing on all types of available browsers, extensions, an...
Is Displayed*

Q52 If you wish to have your results emailed to you, please leave your email below.

Display This Question:

*If We recommend the following tools for you! Alternative Browser: For a different browser that is...
Is Displayed*

Q53 If you wish to have your results emailed to you, please leave your email below.

Display This Question:

*If We recommend the following tools for you! Alternative Browser: For a different browser that is...
Is Displayed*

Q54 If you wish to have your results emailed to you, please leave your email below.

Display This Question:

*If We recommend the following tools for you! Alternative Browser: We recommend that you try our
Br... Is Displayed*

Q55 If you wish to have your results emailed to you, please leave your email below.

Display This Question:

*If We recommend the following tools for you! Alternative Browser: We recommend that you try our
Br... Is Displayed*

Q56 If you wish to have your results emailed to you, please leave your email below.

Display This Question:

*If We recommend the following tools for you! Alternative Browser: If you are willing to switch the...
Is Displayed*

Q57 If you wish to have your results emailed to you, please leave your email below.

Display This Question:

*If We recommend the following tools for you! Alternative Browser: If you are willing to switch the...
Is Displayed*

Q58 If you wish to have your results emailed to you, please leave your email below.

End of Block: Email

Start of Block: Debrief

Q125 Thank you for participating.

There are lots of incentives for companies that create tools for the internet that track user activity while online. One major factor being profit. For this reason, many sites will utilize technologies such as cookies, and a variety of other hidden techniques that gather information unbeknownst to users. They will use the information to target ads to their customers, make predictions and human profiles, and in some cases, sell your data to scammers or other companies.

In our research, we specifically looked at which tools currently exist that provide users protections from this invasive tracking. Many pieces of technology exists that maintain a business model of providing users with a safer method of browsing the internet and not being tracked. We provided you with recommendations of the tools that we believed accomplished this the best by outperforming its competitors in certain tests.

Today, we wanted to not only give you this recommendation based on your interests and intentions of switching, but we wanted to see if viewing your personal data would enhance these intentions. You were either asked to go to a publicly available site owned by Google to view information based on your own browsing or shown compiled data of what a general WPI student has as hobbies based on a compilation of a few different student's traffic. We will be measuring

if viewing your own personal data has any effect on your indication to switch to these privacy-protection tools.

Once again, none of the information you provided will be recorded and linked to you, all of your responses will remain anonymous. We ask that you please do not share any information about this study with anyone as to maintain the integrity of the study.

Thank you!

End of Block: Debrief

Appendix C: IRB Approval & Training Certificate



Completion Date 17-Jan-2021
Expiration Date 17-Jan-2023
Record ID 38201054

This is to certify that:

Jeffrey Harnois

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

Human Subjects Research

(Curriculum Group)

Human Subjects in Undergraduate Student Projects

(Course Learner Group)

1 - Basic Course

(Stage)

Under requirements set by:

Worcester Polytechnic Institute



Verify at www.citiprogram.org/verify/?w2ccb9ac0-bb15-4e1d-9638-999ae5da188d-38201054

WORCESTER POLYTECHNIC INSTITUTE

100 INSTITUTE ROAD, WORCESTER MA 01609 USA

Institutional Review Board

FWA #00015024 - HHS #00007374

Notification of IRB Approval

Date: 10-Feb-2021
PI: Doyle, James K
Protocol Number: IRB-21-0322
Protocol Title: Exploring Internet Privacy

Approved Study Personnel: Harnois, Jeffrey~Doyle, James K~

Effective Date: 10-Feb-2021

Exemption Category: 2

Sponsor*:

The WPI Institutional Review Board (IRB) has reviewed the materials submitted with regard to the above-mentioned protocol. We have determined that this research is exempt from further IRB review under 45 CFR § 46.104 (d). For a detailed description of the categories of exempt research, please refer to the [IRB website](#).

The study is approved indefinitely unless terminated sooner (in writing) by yourself or the WPI IRB. Amendments or changes to the research that might alter this specific approval must be submitted to the WPI IRB for review and may require a full IRB application in order for the research to continue. You are also required to report any adverse events with regard to your study subjects or their data.

Changes to the research which might affect its exempt status must be submitted to the WPI IRB for review and approval before such changes are put into practice. A full IRB application may be required in order for the research to continue.

Please contact the IRB at irb@wpi.edu if you have any questions.

*if blank, the IRB has not reviewed any funding proposal for this protocol