Characterizing and Measuring “Bad Ads” on the Web

Eric Zeng

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Commitee:

Franziska Roesner, Chair

Tadayoshi Kohno

Jamie Morgenstern

Program Authorized to Offer Degree:
Paul G. Allen School of Computer Science & Engineering

University of Washington

Abstract

Characterizing and Measuring “Bad Ads” on the Web

Eric Zeng

Chair of the Supervisory Committee:
Associate Professor Franziska Roesner
Paul G. Allen School of Computer Science & Engineering

Online advertising is a core part of the modern web; ads sustain websites that provide free content and services to consumers, and inform people about products that they may be interested in. However, the ubiquity of the online advertising ecosystem makes it a potent vector for abuse; malicious actors can use the infrastructure of ad networks to serve scams, malware, and other misleading or detrimental content to millions of users across millions of websites. Though the ad networks that provide this infrastructure make efforts to prevent inappropriate and harmful content from appearing on their platforms through content moderation, many kinds of deceptive and unpleasant ads regularly appear on people’s screens. Due to the opacity of the online advertising ecosystem, it is challenging for external observers to assess the harms and scale of problematic online ads.

This dissertation presents a systematic investigation of the nature and prevalence of problematic content in online advertising, or “bad ads”, on the modern web, through four studies. First, this work investigates users’ perceptions of online advertising, characterizing the reasons why people dislike (and like) ads, and identifying types of ad content which engender negative reactions. Second, this work quantitatively measures the phenomenon of “clickbait” advertising on news and media websites. Using data crawled from over 7000 news and media websites, this work finds that native advertising networks are strong drivers of problematic content such as content farms and advertorials, and are extremely common across a variety of news websites. Third, this work examines problematic content in online political advertising during the 2020 U.S. elections. In a longitudinal measurement study, this work finds evidence of multiple categories of deceptive political content in online ads, including misleading polls and petitions, political clickbait, and misleading political-themed product ads, and found that these ads were targeted at partisan news sources. Lastly, this work empirically measures the targeting of online ads more broadly, through a unique field study using data collected from 286 real users. This dataset provides measurements of the prevalence of different categories of ad content, how such categories are targeted across websites and demographic groups, the monetary value placed on users by advertisers. Together, these works provide a foundation for future regulation, policy, and research aiming to curb problematic content in online advertising, and improve the overall experience for users on the web.

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation covers work done in the last three years of my PhD. The six years I have been in the PhD program have been a period of immense personal and professional growth, and it would not be possible without the many people who have supported me along the way.

Above all, I have been tremendously fortunate to have the mentorship of my advisor Franzi Roesner. Franzi twice took a chance on me, when I out of the blue asked to work with her as a Master's student, and again as a PhD student, for which I am incredibly grateful. And ever since, Franzi has been endlessly supportive, providing me with the freedom to grow and explore, patience and steady encouragement when times were tough, and insightful guidance when I had big decisions to make. Most of all, during the most the difficult moments of my PhD, I never once doubted that Franzi had confidence in me. Thank you Franzi, for everything you have taught me, and all the opportunities that you have given me.

I also extremely grateful to have worked with Yoshi Kohno throughout all of the work in this dissertation. Yoshi and I started working together at a critical time, when I needed to decide what I wanted to make out of my time in grad school. Yoshi challenged me to do the best work that I could, and to become the best version of myself as a researcher. I would not have been able to accomplish so much in these years without Yoshi's belief that I could succeed, and for that I am truly thankful.

Another early mentor that I owe so much to is Hélène Martin, my computer science teacher at Garfield High School. Hélène was an amazing teacher, who opened my eyes to the joy and possibilities of computer science, and she opened many doors for me throughout my career. I could not have asked for a better role model for the beginning of my journey.

I am also grateful to have worked with amazing mentors outside of UW at internships. Thanks to Emily Stark, who was a fantastic mentor, working with her and the team at Chrome Security was one of my greatest professional pleasures, and made me feel like part of the security community for the first time. Thanks as well to Kim Laine, my mentor at Microsoft Research, who was one of the most open minded and thoughtful people I have had the chance to work with. Learning from how Kim navigated challenges broadened my perspectives and had a strong positive impact on how I approach research.

I also greatly appreciate Jamie Morgenstern and Mako Hill for their service on this committee. Their feedback on my thesis proposal was extremely helpful in setting me in the right direction, as it was my first real attempt at independently proposing and executing on research projects.

My graduate school experience would not have been nearly as exciting, enriching, productive, and fun without the past and present members of the Security and Privacy Lab. First off, I am extremely thankful to have been mentored by Ada Lerner on my very first project. Ada's enthusiasm for research was contagious, and her excitement encouraged me to push beyond what I thought was possible of myself. I am also extremely grateful for the friendship of Lucy Simko, Anna Simpson, Camille Cobb and Kiron Lebeck, who immediately made me feel welcome and at home in the lab; for without their warmth and kindness, I may not have stayed on after completing my Master's. I am also extremely fortunate to have collaborated, commiserated, and conspired with Miranda Wei, my favorite co-author, a thoughtful and understanding friend, and now technically the senior student in the lab (big responsibility!). And thank you to all of the other members of the lab: Alex Takakuwa, Chris Geeng, Christine Chen, David Kohlbrenner, Earlence Fernandes, Ivan Evtimov, Kaiming Cheng, Karl Koscher, Kentrell Owens, Michael Flanders, Paul

Vines, Peter Ney, Rachel McAmis, Shrirang Mare, Tina Yeung, Theo Gregersen, and Umar Iqbal for the years of deep conversations, goofing off, late deadlines, happy hours, lunch scavenging, and losing softball games. Knowing you all has made my time here truly special.

I am also grateful for all of my friends inside and outside of graduate school, who kept me going through the tough times, and made life a joy to live. Thanks to Ellis Michael, Maureen Daum, Chandrakana Nandi, Nick Walker, Zach Tatlock, Aishwarya Mandyam, and the rest of Race Condition Running for being steady friends, pacers, and brunch companions. Thanks to my ski partners at school for all of the weekdays we skipped to rip some pow, including Matt Johnson, Galen Weld, Jason Hoffman, Gus Smith, Daniel Revier, Leah Perlmutter, and Chien-Yu Lin. And thanks to all of my old friends, Seth Vanderwilt, Eli McClintock-Shapiro, Brant McPherson, Eamon Gaffney, Caitlin Rochlin, King Xia, Kian Vesteinsson, Lane Aasen, Izzy Majcher, Rama Gokhale, Graeme Britz, Isaac Jaeger, Cindy Fan, Daniel Lin, Kyle Steiner, and anyone else I may have accidentally forgotten to mention, for all of the alpine adventures, board game nights, band practices, house parties, dungeons and dragons, long walks and talks, and late night gaming; I couldn't have done it without your company and support.

Last, but not least, I could not have done any of this without the unconditional support and love of my parents, Weiping Zeng and Qinghong Zhou, and my sister, Aileen Zeng. Thank you, for encouraging me to take a leap of faith, find my own path, and do this PhD; and for supporting me throughout all of the ups and downs of this whole thing.

My time in graduate school, and the projects described in this dissertation were supported in part by the National Science Foundation under Awards CNS-1463968, CNS-1513575, CNS-1513584, CNS-1565252, CNS-1651230, and CNS-2041894; by a Hacherl Endowed Fellowship; the UW Tech Policy Lab, which receives support from: the William and Flora Hewlett Foundation, the John D. and Catherine T. MacArthur Foundation, Microsoft, the Pierre and Pamela Omidyar Fund at the Silicon Valley Community Foundation; the UW Center for an Informed Public; and the John S.

and James L. Knight Foundation.

Chapter 1

Introduction

Like it or not, online advertising is a fundamental part of the World Wide Web as we know it. Many businesses on the web, ranging from behemoths like Facebook and Google, to local news websites and independent creators, can make their content and services mostly free to use, through revenue generated by online ads on their websites. Additionally, many businesses depend on running ads online to promote their products and services, as the internet has become the primary way that people consume information. In 2022, advertisers are projected to spend over $150 billion on online ads in the U.S. alone, and over $500 billion worldwide [197].

Over the past few decades, the online advertising industry has developed an extremely powerful infrastructure for delivering online advertising at an Internet-wide scale. For example, Google claims that they can help advertisers serve their ads on over 2 million websites, and "reach over 90% of internet users across the globe" through the Google Display Network [4]. Advertisers can also use this infrastructure to target ads at people at an extremely granular level, such as people's demographics, interests, previous browsing history, and email addresses.

The infrastructure of online advertising can be misused by dishonest advertisers to deliver misleading, harmful, and unpleasant ads — or what will be referred to as "bad ads" or "problematic ads" in the remainder of this dissertation — to millions of users on the web. These ads commonly use misleading or deceptive designs and language to trick people into clicking on them, such as clickbait headlines and fake user interfaces, and promote things like dubious dietary supplements, potentially unwanted software, and clickbait content farms. And these ads can lead to real harms: the US Federal Trade Commission reported that last year, consumers lost $96 million dollars to scams via online ads [74]. Online ads are also frequently used to spread malware and unwanted software, like fake apps with hidden subscription costs and ads, and ransomware, which encrypts a computer's files and demands payments. Furthermore, these ads simply harm users' experience on the web, and can hurt the reputation of the websites that run them [218].

Due to the opacity and sheer scale of online advertising, we have limited empirical insight into the practices of advertisers and the infrastructure enabling them. Though prior work has studied the privacy impacts of online advertising, revealing extensive ecosystems of web tracking [176, 128, 63, 3, 2, 105, 17], and the scale and impacts of targeting [37, 131, 150, 103, 9, 102], the scope of deceptive and misleading ad content has received less attention. In this dissertation, I present work studying the qualitative and quantitative extent of problematic content online advertising, with the goal of highlighting systemic issues in the practices of online advertisers, and providing data to support the creation of policies, regulations, or technical interventions to protect users from harmful ads and improve the experience of using the web.

Characterizing "Bad" Ad Content

There exist ads whose content are generally considered unfair, harmful, and/or detrimental to user experience on the web. In the 1990s and 2000s, there was much concern about distracting and intrusive banner ads, often coming in animated or popup formats. Deceptive and false advertising techniques predate online advertising, but they have persistently been an issue. Evergreen examples of deceptive online advertising include cure-all supplement ads, and dubious investment advice and opportunities. Other examples are specific to particular time periods and technologies; for example, in the 2000s there were many ads for screensavers or ringtones that included spyware or hidden charges [58]. In the mid 2010s, a major issue was the use of online ads as a vector for malware ("malvertising") [85, 10].

Over time, as technology, regulations, and policies have changed, so have the types of ads and deceptions used by malicious advertisers, to evade scrutiny and content moderation. So what are the common types of "bad ads" today? What kinds of content and deceptive techniques are used? And for ads that exist in the gray area of acceptability, how do we determine what makes an ad "bad" in the first place?

This dissertation investigates the qualitative nature of problematic content in online display advertising, in three studies. Chapter 3 systematically explores users' perceptions of modern display advertising, providing a user-centric perspective on what makes certain ads "bad". Chapter 4 surfaces types of problematic ad content on news and media websites; particularly the phenomenon of "clickbait" native advertising. Chapter 5 examines misleading online political ads observed during the 2020 U.S. elections.

Measuring the Prevalence and Targeting of Ad Content

Though establishing definitions and taxonomies for "bad ads" is useful, it is also important to understand the prevalence and impact of these ads, to assess what kinds of policies, regulations, and other interventions should be prioritized. However, there is little publicly available data on the reach and effectiveness of problematic ad content, because advertisers and ad networks rarely disclose the performance of their ad campaigns and content moderation efforts. But there are still some questions that as researchers, we can answer through external auditing. For example: what proportion of ads shown to people contain problematic content? Are "bad" ads more common on certain websites? Are specific ad networks more involved in delivering bad ads than others?

This dissertation presents several measurement studies that provide an empirical perspective on problematic online advertising. Chapter 4 measures the prevalence of "bad" ads on news and misinformation sites, and across native and traditional ad networks. Chapter 5 measure political ads, how they are contextually targeted to partisan websites, and how problematic political ads are targeted. Chapter 6 examines targeting of ads at large, and the relationship between targeting and the amount advertisers bid to run their ads.

Contributions

User-Centric Taxonomy of Problematic Ad Content Chapter 3 explores user perceptions of online advertising, and systematically investigates the kinds of ads that are detrimental to user experience. Unlike current regulations and content policies, which are largely determined as a matter of law and corporate policy, we establish a taxonomy of what people themselves like and don't like in ads. This taxonomy provides a systematic description of the "gray area" of content moderation, identifying things that people dislike but may go against the incentives of advertising platforms or the limits of regulatory authority. The taxonomy also provides a basis for future research on problematic ad content, and directions for improvement for ad networks and regulators.

Quantifying the Role of Native Advertising in Bad Ads Anecdotal reports in the press suggest that native advertisements on news websites contain a variety of misleading content, such as clickbait, content farms, advertorials, and scams. Chapter 4 presents a quantitative measurement of ads on news websites and misinformation websites. Among the findings, native ads are common on news sites, comprising a high percentage of the total ads shown; native ads contain a higher proportion of problematic content such as advertorials, content farms, and investment pitches; and native ads widely used on both legitimate and misinformation sites. These results provide concrete evidence that native ad networks are primarily responsible for ad content on news sites that may mislead or harm users, and that news websites often tolerate them.

Highlighting Problematic Political Advertising Chapter 5 investigates political advertising during the 2020 U.S. Elections, examining both examples of misleading tactics and misinformation, such as political clickbait news ads, fake political poll ads, and products ads. This study also examines how political ads are targeted at partisan websites, finding that more partisan sites receive more political ads, including problematic political ads. These results highlight the need for greater scrutiny of political ads by ad networks and policymakers.

Empirical Measurements of Targeting and Ad Auctions Chapter 6 presents a field study of online advertising, investigating how ads are targeted at different websites and demographic groups, and how much advertisers pay to show ads to different types of users. Though this study does not address bad ads directly, but investigates the infrastructure and economics of online ads that are leveraged by bad ads. The study finds that targeting and differences in winning bids from advertisers is stark across websites, and vary among individuals broadly, but do not differ substantially across demographic groups. The study also provides empirical measurements on the quantified value of users, and the amount of targeting and personalization broadly.

Methodologies for Auditing the Online Advertising Ecosystem Finally, this dissertation presents novel methods and approaches for auditing the practices of the online advertising ecosystem. The work in this dissertation uses a mixed methods approach, blending qualitative approaches to characterizing ad content, with rigorous quantitative analyses and systematic methods for data collection. Additionally, this work leverages and builds on new tools to enable large-scale analysis of ad content, such as the use of large language models like BERT for content classification and clustering, and tools like the Puppeteer browser automation library and Docker for scalable crawls for data collection. The methods described here (and tools released) provide methodological foundations for researchers in related fields involving content analysis, including security, privacy, trust and safety, social media, and internet measurement.

Together, Chapters 3-6 paint a picture of what "bad ads" look like today, how they are distributed across the web, and provide a foundation for future research, policy, and regulation. Chapter 7 summarizes these contributions, provides directions for future work, and commentary on the state of the online advertising and the web.

Chapter 2

Background and Related Work

Since the first banner ad appeared on hotwired.com in 1994 [145], many have raised concerns about the negative impacts of online advertising on people's security, privacy, and user experience; ranging from concerns about disruptive popup ads in the late 1990s, to current concerns about the privacy impacts of web tracking and behaviorally targeted advertising. In response to the various problematic practices of the online advertising industry, a substantial body of research in the computer security and privacy community, as well as research from the human-computer interaction and marketing research communities, has investigated the misuse and harms of online advertising.

This chapter provides an overview of prior research investigating problematic aspects of online advertising: the privacy impacts of targeted advertising and tracking, computer security issues created or spread via online ads, deceptive and misleading content in online ads, and poor user experiences created by online ads. This dissertation builds on this body of work, through systematic studies that shed light on the qualitative nature and quantitative extent of misleading and low quality advertising on the web.

2.1 Deceptive and Misleading Online Advertising

Some forms of problematic online advertising use deception to achieve their goals, including engaging in false advertising and misleading claims, and visually deceptive techniques such as fake user interfaces and native advertising. These ads can cause material harms to people, by costing them money through fraudulent products or misleading purchase terms, wasting their time and attention, or installing unwanted software or malware. Prior work has catalogued a variety of forms of deceptive advertising, and studied the effectiveness of the deceptions used.

Early Forms of Deceptive Advertising Deceptive advertising has been a problem since the early days of the web. In the 2000s, common types included software and ringtone ads that claimed to be free, but had hidden charges, impersonation ads, fake user interfaces, and spyware. Edelman created a taxonomy the types of deceptive content and techniques found in display ads on the Yahoo ad marketplace, finding many examples of ads that appear to violate the U.S. Federal Trade Commission's rules on unfair and deceptive advertising practices [59, 58].

Native Advertising More recently, in the 2010s, there has been significant concern about deceptively formatted "native" advertisements, which are designed to imitate content on the of the host website, for example, sponsored search results, sponsored posts, and ads that look like articles on news websites. Significant prior work across disciplines suggests that most users do poorly at identifying such ads (e.g., [21, 121, 12, 218]), though people may do better after more experience [115], or with different disclosure designs (e.g., [219, 101]). However, native ads may affect user behavior even when identified [181]. Prior work suggests that native ads can reduce users' perception of the credibility of the hosting site, even if the ads are rated as high quality in isolation [47]. The Federal Trade Commission has attempted to regulate some aspects of native ads, by creating and enforcing rules requiring that native ads include disclosures that indicate that they are paid content from advertisers [69, 126, 68, 72].

Native ads have been observed to contain a substantial amount of low-quality content. Bashir et al. conducted a measurement study of the major native ad networks, finding that a large percentage of native ads have are from low quality advertisers that promote celebrity gossip and dubious financial services and use poor disclosure practices [22]. Anecdotal reports in the media often describe native ads as "clickbait"; i.e. ads that bait users into clicking on them using sensationalist headlines and by hiding information on the landing page [133, 152]. Often times, native ads make implied or explicitly misleading or deceptive claims, particularly ads promoting dietary supplements [207]. Native ads have also been observed to promote political mis/disinformation, or financially support websites that spread mis/disinformation [113, 206]

With the exception of Bashir et al. [22], little work has systematically measured the prevalence and content of native ads on the web. Chapter 4 presents a measurement study of problematic ads on news websites and misinformation websites. Among the findings, I observe that native ads play a leading role in spreading problematic ad content compared to standard display ads, and that native ads are commonly used on both mainstream news and misinformation websites.

Psychology of Deceptive Advertising Prior work studying deceptive advertising predates web ads (i.e., print and TV ads), showing, for instance, that false information in ads can be effectively refuted later only under certain conditions (e.g. [29, 111, 112]), that people infer false claims not directly stated in ads and misattribute claims to incorrect sources (e.g., [97, 177, 164, 109]), and that people's awareness of specific deceptive ads can harm their attitudes towards those brands [110] as well as towards advertising *in general* [49, 35].

Dark Patterns In addition to studies of deception in native advertising and anecdotal evidence of problematic ad content discussed in Section 4.1, the work in this dissertation is thematically related to broader discussions of "dark patterns" [30] on the web and in mobile apps (e.g., [92, 154, 26]). Most closely related is recent work systematically studying affiliate marketing on YouTube and Pinterest [139, 201] and dark patterns on shopping websites [138], though neither considered web ads.

2.2 Computer Security and Online Advertising

Online ads are known to be used as a vector for spreading malware and potentially unwanted programs, a technique colloquially known as "malvertising".

Social Engineering Attacks in Ads Online ads can be used to deceive or socially engineer users into installing or running harmful programs. Nelms et al. conducted a field study of the sources of malicious software downloads on a university network, finding that over 80% were downloaded via an online ad, using deceptive and persuasive techniques such as impersonation, fear, and enticements [151].

Chapters 3 and 4 briefly touch on examples of ads that attempt to deceive users into downloading potentially unwanted programs; examining user reactions to the content of such ads, and their prevalence on news websites.

Drive-by-Download Attacks in Ads Online ads can also be used to directly deploy malicious JavaScript payloads that exploit vulnerabilities in the browser, a technique which is sometimes known as a "drive-by-download" [129, 224].

Adware For some forms of malware, the goal of the malware is to run online ads on the victim's device, and collect the revenue from those ads. This type of malware is commonly called "adware". For example, adware browser extensions inject additional ads into the websites that the user visits, and work by Xing et al. found that the injected ads often contained ads that spread malware as well [220].

Click Fraud Farther afield, online advertisers themselves can be victimized in clickfraud schemes. Website owners can earn fraudulent revenue from advertisers by generating fraudulent impressions or clicks on ads, through techniques like clickjacking, bots, and hidden ads [43].

2.3 Privacy and Online Advertising

Online advertisers have the capability to target individuals at high granularity on millions of websites and apps, using attributes such as a person's location, their previous browsing activity, and identifiers like their email address. These targeting capabilities are enabled by a panopticon of web trackers and data brokers which collect, aggregate, and share data on people's online histories, location, and personally identifiable information. Though this infrastructure allows advertisers to deliver ads that are more relevant to people's interests, it has a variety of negative effects on users and their privacy. A significant amount of prior work has studied various aspects of this infrastructure, to provide transparency on the privacy impacts of online advertising.

Web Tracking and Fingerprinting Researchers have measured the prevalence of various methods for identifying users and tracking their history on the web. Studies have examined the spread of third-party web trackers, such as tracking pixels and analytics scripts, which allow tracking firms to collect a users' browsing history across domains [176, 63, 128]. Other studies have examined the spread of fingerprinting scripts, which evade protections against web trackers that block the use of cookies by third party scripts [153, 105, 17]. And other research has examined the weaponization of web trackers for targeted surveillance [213].

Targeting Researchers have also developed methods for detecting and measuring behavioral targeting — the targeting of ads based on characteristics and interests inferred from past online activity. These methods generally involve crawling for ads on the web using multiple browsers, each browser with a different synthetically constructed browsing profile, and looking for differences in the topics of ads seen by each profile. These studies show clear differences in ads shown to different profiles [37, 131, 150], including problematic differences such as gender discrimination in career ads [50]. Other methods for targeting detection include fine-grained, statistical approaches [124, 125], and a crowdsourced approach based on distributed counting [103].

Though prior work confirmed the widespread usage of behavioral targeting in online advertising, because these studies primarily used crawler-based methods, the findings may not have been representative of what end users actually experience on the web. Chapter 6 fills this gap through a field study of targeting in online advertising, using data gathered from real users to provide novel measurements on how ads differ between individuals and demographic groups in realistic conditions.

User Perceptions of Tracking and Targeting Other work has studied people's perception of the privacy practices of the online advertising industry, and the impacts those practices have on their behavior. People find online behavioral advertising to be "creepy" [211] and disapprove of data collection and sharing practices [210, 217, 171, 191]. People are also concerned that targeted ads could reveal embarrassing or private information about them [5]. The complexity and lack of transparency lead to fears and folk theories [221], e.g. that one's phone can listen to conversations and use the data to target ads [160].

2.4 Discrimination in Ad Delivery

Researchers have also surfaced concerns about how ads may be targeted at users in potentially discriminatory ways. Sweeney conducted some of the first work in this field, finding that Google Search Ads for people's names would turn up different results based on the perceived race of the name, with names more commonly given to Black babies more likely to show ads for arrests and criminal record searches than names commonly given to White babies [202]. Work done by Ali et al., Kinglsey et al., and Imana et al. suggest that even with the absence of targeting parameters set by advertisers, the ad delivery optimization algorithm on platforms like Facebook may serve housing and career opportunity ads in a discriminatory manner across demographic groups [9, 118, 102].

It is still unknown the extent to which discrimination exists in ad targeting and delivery, due to the lack of transparency from ad platforms. Chapter 6 attempts to detect whether discrimination can be empirically detected in display advertising, investigating whether there are disparities in the amount advertisers bid to place ads across demographic groups, and whether ads are targeted across demographic groups.

2.5 Poor User Experiences with Online Advertising

Even when ads are not explicitly malicious or do not cause material harms, many users still dislike seeing ads. A common reason that people dislike web ads is that they are annoying and disruptive, either due to a general aversion to ads, or due to the specific design of the ad. Prior work has studied and summarized design features of ads that lead to perceived or measured reductions in the user experience, including ads that are animated, too large, or pop up [178, 83, 62, 27]. The impacts of these issues include increased cognitive load, feelings of irritation among users, and reduced trust in the hosting websites and in advertising or advertisers [27, 229, 32].

These attitudes motivated the development and widespread adoption of ad blockers — today, around 18% of users in the U.S. and 32% of users in Germany are estimated to use ad blockers [14, 170, 135]. Advertising platforms have taken some steps to restrict some types of poor user experiences. Google and other ad platforms created guidelines against certain intrusive attributes, such as autoplaying videos and ads that heavily impact website performance [77], and Google enforces some of these guidelines in the Chrome browser by blocking ads that violate their standards [23, 180].

Though prior work has investigated how formats and placements of ads, like animations in ads and popup ads, can impact user experience, little work has investigated how the content of ads can impact user experience. Chapter 3 investigates users' perceptions of ad content, surfacing other qualities beyond visual intrusiveness, such as distasteful imagery, pushy language, ugly design, and general untrustworthiness.

Chapter 3

What Makes a “Bad Ad”?

This chapter investigates how to define “bad ads”. Though anecdotal evidence has shown that there is a large variety of potentially problematic content in online ads, there has been little systematic study of which types of ad content are detrimental to user experience, and the reasons why people find them problematic. Towards systematizing the types of problematic ad content, this chapter presents an empirical study of people’s perceptions of problematic ad content, through a pair of online surveys. First, I propose a taxonomy of 15 positive and negative user reactions to online advertising based on a survey of 60 participants, including reactions like “clickbait”, “untrustworthy”, and “distasteful”. Next, I present several classes of online ad content that users dislike or find problematic, using a dataset of 500 ads crawled from popular websites, labeled by 1000 participants using our taxonomy, such as ads for software downloads, listicles, and dietary supplements.

This chapter originally appeared as the paper “What Makes a ‘Bad’ Ad? User Perceptions of Problematic Online Advertising” at the CHI Conference on Human Factors in Computing Systems in 2021 [227].

3.1 Introduction

Many web users dislike online ads, finding them to be annoying, intrusive, and detrimental to their security or privacy. In an attempt to filter such "bad" ads, many users turn to ad blockers [14] — for instance, a 2016 study estimated that 18% of U.S. internet users and 37% of German internet users used an ad blocker [135], a large percentage considering that it takes some initiative and technical knowledge to seek out and install an ad blocker.

There are many drivers of negative attitudes towards online ads. Some users find the mere presence of ads to be problematic, often associated with their (perceived) increasingly disruptive, intrusive, and/or annoying qualities [14] or their impact on the load times of websites [190]. .Users are also concerned about the privacy impacts of ads: Users also find the capabilities of online behavioral advertising to be creepy and privacy-invasive (e.g., [211, 64, 217, 216]) The specific *content* of ads can also cause direct or indirect harms to consumers, ranging from material harms in the extreme (e.g., scams [73, 1, 143], malware [129, 224, 220, 151], and discriminatory advertising [9, 118]) to simply annoying techniques that disrupt the user experience (e.g., animated banner ads [83, 32, 99]).

In this chapter, we focus specifically on this last category of concerns, studying people's perceptions of problematic or "bad" user-visible *content* in modern web-based ads. Driving this exploration is the observation that problematic content in modern web ads can be more subtle than flashing banner ads and outright scams. Recent anecdotes and studies suggest high volumes and a wide range of potentially problematic content, including "clickbait", advertorials or endorsements with poor disclosure practices, low-quality content farms, and deceptively formatted "native" ads designed to imitate the style of the hosting page [126, 139, 152, 113, 133, 134, 226, 84, 16, 219, 12, 47, 200]. While researchers and the popular press have drawn attention to these types of ad content, we lack a systematic understanding of how web users perceive these types of ads on the modern web in general. What makes an ad "bad", in the eyes of today's web users? What are people's perceptions and mental models of ads with arguably problematic content like "clickbait", which falls in a grey area between scams and poorly designed annoying ads? What exactly is it that causes people to dislike (or like) an ad or class of ads? For future regulation and research attempting to classify, measure, and/or improve the quality of the ads ecosystem, where exactly should the line be drawn?

We argue that such a systematic understanding of what makes an ad "bad" — grounded in the perceptions of a range of web users, not expert regulators, advertisers, or researchers — is crucial for two reasons. First, while some ads can clearly be considered "bad", like outright scams, and others can be considered "benign", like honest ads for legitimate products, there is a gray area where it is more nuanced and difficult to cleanly classify. For example, "clickbait" ads for tabloid-style celebrity news articles may not cross the line for causing material harms to consumers, but may annoy many users and use misleading techniques. While the U.S. Federal Trade Commission currently concerns itself with explicitly harmful ads like scams and deceptive disclosures [72, 126, 34], whether and how to address "clickbait" and other distasteful content is more nuanced. As part of our work, we seek to identify ads that do not violate current regulations and policies, but do harm user experiences, in order to inform improvements such as policy changes or the development of automated solutions. Second, research interested in measuring, classifying, and experimenting on "bad" online ads will benefit from having detailed definitions and labeled examples of "bad" ads, grounded in real users' perceptions and opinions. For example, our prior work measuring the prevalence of "problematic" ads on the web used a researcher-created codebook of potentially problematic ad content; that codebook was not directly grounded in broader user experiences and perceptions [226].

Research Questions In this chapter, our goal is thus to systematically elicit and study what kinds of online ads people dislike, and the reasons why they dislike them, focusing specifically on the user-visible content of those ads (rather than the underlying technical mechanisms for ad targeting and delivery). We have two primary research questions:

1. **RQ1 — Defining "bad" in ads**: What are the different types of negative (and positive) reactions that people have to online ads that they see? In other words, *why* do people dislike (or like) online ads?
2. **RQ2 — Identifying and characterizing "bad" ads**: What specific kinds of content and tactics in online ads cause people to have negative reactions? In other words, *which* ads do people dislike (or like)?

While ads appear in many places online — including in social media feeds and mobile apps — we focus specifically on third party programmatic advertising on the web [7], commonly found on news, media, and other content websites. Unlike more vertically integrated social media platforms, the programmatic ad ecosystem is complex and diverse, with many different stakeholders and potential points of policy (non-)enforcement, including advertisers, supply-side and demand-side platforms, and the websites hosting the ads themselves. A benefit of our focus on web ads is that the public nature of the web allows us to crawl and collect ads across a wide range of websites, without needing to rely on explicit ad transparency platforms (which may be limited or incomplete [60, 179]) or mobile app data collection (which is more technically challenging). We expect that many of our findings will translate to ads in other contexts (e.g., social media, mobile), though these different contexts also raise additional research questions about the interactions between the affordances of those platforms and the types of ads that people like or dislike.

Contributions Figure 3.1 shows an overview of the different components of our work and our resulting outputs and contributions. Specifically, our contributions include:

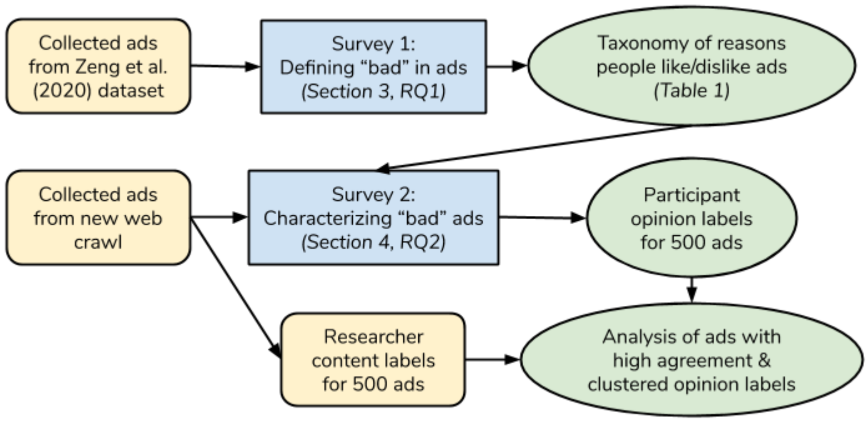


Figure 3.1: An overview of our work and contributions.

1. Based on a qualitative survey characterizing 60 participants' attitudes towards the content and techniques found in modern online web ads, we distill a taxonomy of 15 reasons why people dislike (and like) ads on the web, such as "untrustworthy", "clickbait", "ugly / bad style", and "boring" (Section 3.3, answering RQ1).
2. Using this taxonomy, we generate a dataset of 500 ads sampled randomly from a crawl of popular websites, labeled with 12,972 opinion labels from 1025 people (Section 3.4, towards answering RQ2). This dataset is available in the paper's supplemental materials[1].
3. Combining participant opinion labels with researcher content labels of these 500 ads, and using unsupervised learning techniques, we identify and characterize classes of ad content and techniques that users react negatively to, such as clickbait native ads, distasteful content, deceptive and "scammy" content, and politicized ads (Section 3.4, answering RQ2).

Our findings serve as a foundation for policy and research on problematic online advertising: for regulators, advertisers, and ad platforms, we provide evidence on which types of ads are most detrimental to user experience and consumer welfare, and for researchers, we provide a user-centric framework for defining problematic ad content, enabling future research on the online

[1]Dataset also available at https://github.com/eric-zeng/chi-bad-ads-data

advertising ecosystem.

3.2 Motivation

We identify several key gaps in prior work that we aim to address. First, studies of user perceptions of problematic ad content in the HCI community have focused largely on more traditional design issues (e.g., animated or explicitly deceptive ads), rather than the broader and less well-defined range of "clickbait" and other techniques prevalent on the modern web. Second, research on the potential harms of online advertising in the computer security and privacy community primarily focuses on ad targeting, distribution, and malware, rather than the user-facing content of the ads. Finally, many anecdotes or measurement studies of potentially problematic content in ads rely on researcher-created definitions of what is problematic, rather than being grounded in user perceptions. Are there types of problematic ad content that bother and harm users, but have not been addressed in prior measurement studies or in the policies of regulators and ad companies? And what exactly makes a "bad" ad bad? In this work, we aim to bridge these gaps through a user-centric analysis of ad content, eliciting user perceptions of a wide range of ads collected from the modern web and characterizing which attributes of an ad's content contribute to negative user reactions.

3.3 Survey 1: Why Do People Dislike Ads?

Towards answering our first research question, we conducted a qualitative survey to elicit a detailed set of reasons for what people like or dislike about the content of modern online ads. The resulting taxonomy enables future studies that classify, measure, and experiment on "bad" online ads, including the second part of this paper (Section 3.4).

Though our primary research questions are around reasons that people *dislike* ads, we also collect data about reasons they may *like* ads. This is for two reasons: first, we expect that there are ads that users genuinely like, and that a user may both like and dislike parts of an ad, so we aim to surface the full spectrum of users' opinions. Second, online ads are fundamental to supporting content and services on the modern web, and we aim for our work to ultimately improve the user experience of ads, not necessarily to banish ads entirely.

3.3.1 Survey 1 Methodology

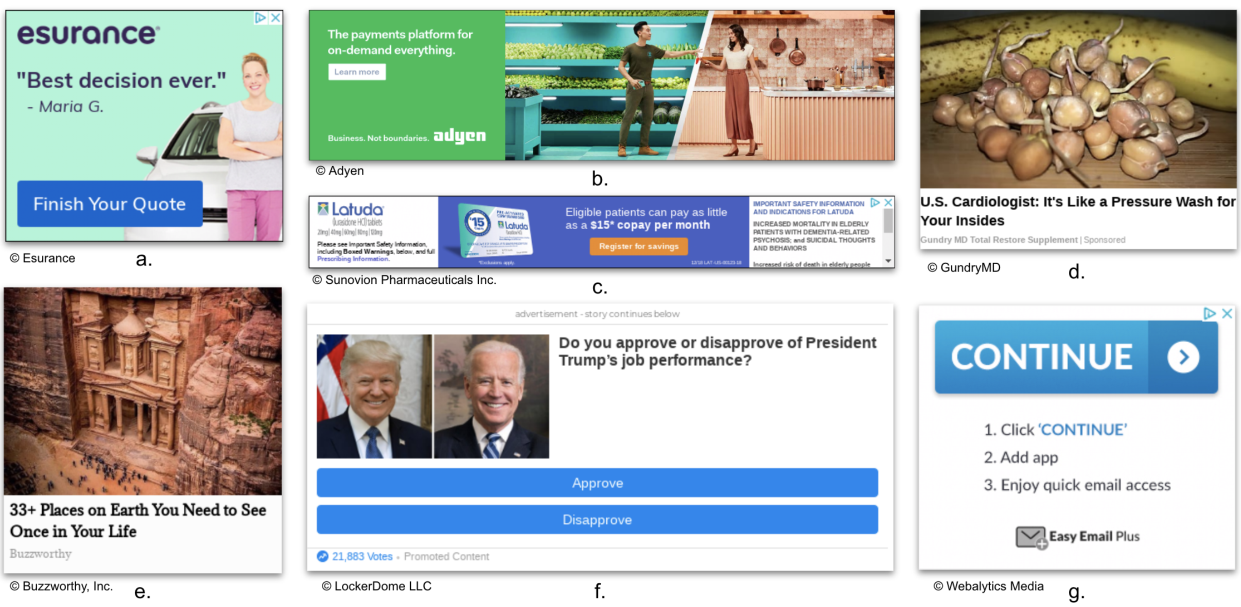


Figure 3.2: A sample of ads shown to participants in Survey 1, selected from the dataset of our prior work [226]. In that study, ads a-c were categorized as "benign", and were each coded as "Product". Ads d-f were categorized as "problematic", using the following codes: d) Supplement, e) Content Farm, f) Political Poll, g) Potentially Unwanted Software.

Survey Protocol

We curated a set of 30 ads found on the web (described below in Section 3.3.1). We showed each participant 4 randomly selected ads, and collected:

- Their overall opinion of the ad (5-point Likert scale).
- What they liked and disliked about it (free response).
- What they liked and disliked about similar ads if they remember them (free response).
- Alternate keywords and phrases they would use to describe the ad.

For each participant, we also asked (a) what they like and dislike about online ads in general (free response), both at the beginning and end of the survey in case doing the survey jogged their memory, and (b) whether they use an ad blocker, and why. See Appendix A.1 for the full survey protocol.

Ads Dataset

To seed a diverse set of both positive and negative reactions from participants, we asked participants to provide their opinions on both "good" and "bad" ads. We selected a set of 30 "problematic" and "benign" ads from a large, manually-labeled dataset[2] of ads that we created in our prior work [226].

We created our previous dataset using a web crawler to scrape ads from the top 100 most popular news and misinformation websites. The ads collected were primarily third-party programmatic ads, such as banner ads, sponsored content ads, and native ads. The dataset did not include social media ads, video ads, search result ads, and retargeted ads. The ads were collected in January 2020. We manually labeled 5414 ads, using a researcher-generated codebook of problematic practices. Ads were considered "problematic" if they employed a known misleading practice, and were labeled with codes such as "Content Farm", "Potentially Unwanted Software", and "Suppplements";

[2] Prior dataset available at https://github.com/eric-zeng/conpro-bad-ads-data.

Demographic Categories	Survey 1 n=60	Survey 2 n=1025	Ad Blocker Usage Survey 1	Survey 2
Gender				
Female	55.0%	45.1%	51.5%	49.1%
Male	45.0%	51.9%	59.3%	63.9%
Prefer not to say	—	0.2%	—	100.0%
No Data	—	2.8%	—	41.4%
Age				
18-24	38.3%	28.1%	69.6%	69.5%
25-34	26.7%	33.2%	56.3%	61.5%
35-44	16.7%	20.1%	20.0%	48.1%
45-54	10.0%	9.0%	33.3%	39.1%
55+	8.3%	6.8%	40.0%	32.9%
No Data	—	2.8%	—	48.3%
Employment Status				
Full-Time	43.3%	43.0%	53.8%	53.8%
Part Time	16.7%	16.3%	40.0%	59.9%
Unemployed	21.7%	17.1%	61.5%	67.4%
Not in Paid Work *(e.g. retired, disabled)*	6.7%	9.1%	25.0%	49.5%
Other	10.0%	8.5%	83.3%	63.2%
No Data	1.7%	6.0%	100.0%	45.2%
Student Status				
Yes	40.0%	29.9%	66.7%	53.7%
No	58.3%	66.0%	45.7%	65.0%
No Data	1.7%	4.1%	100.0%	44.2%

Table 3.1: Participant demographics for Surveys 1 and 2. The "Ad Blocker Usage" columns show the percentage of participants within each demographic group that use ad blockers, in each survey. Our sample skewed young, and used ad blockers more than the overall U.S. population.

otherwise ads were considered "benign", and labeled with codes like "Product".

For this survey we picked 8 "benign" ads, and 22 "problematic" ads from our previous dataset. We show a sample of these ads in Figure 3.2.

We selected ads from this dataset with the goal of representing a wide breadth of qualitative characteristics in a manageable number of ads for the purposes of our survey. However, since ads differ on many different features, and we did not know which features would be salient for participants ahead of time, we used the following set of heuristics to guide the selection of ads: First, we chose at least one ad labeled with each problematic code in our previous dataset. We selected additional ads for a specific problematic code if there was diversity in the code in one of the following characteristics: product type, prominence of advertising disclosure, native vs. display formats, and the use of inappropriate content (distasteful, disgusting, or unpleasant images, sexually suggestive images, political content in non-campaign ads, sensationalist claims, hateful or violent content, and deceptive visual elements). We generated this list of characteristics based on our own preliminary qualitative analysis of the ads in the dataset, and based on the content policies of advertising companies like Google [87].

Analysis

We analyzed the data from our survey using a grounded theory approach. We started with an initial round of open coding, creating codes to describe reasons why participants disliked or liked the ads, using words directly taken from the responses, or words that closely summarized them, such as "clickbait", "fearmongering", and "virus". Then, we iteratively generated a set of hierarchical codes that grouped low level codes, such as "Untrustworthy", and "Politicized". Two coders performed both the open coding and hierarchical coding, after which they discussed and synthesized their codebooks to capture differences how they grouped their codes. Table 3.2 summarizes the resulting categories. The first ten rows are the negative categories distilled from reasons participants disliked ads, and the bottom five rows are the positive categories distilled from reasons participants liked ads.

Participants and Ethics

We recruited 60 participants in the United States to take the survey through Prolific[3], an online research panel. We recruited the participants iteratively until we reached theoretical saturation: recruiting 10-25 participants at a time, coding the results, and repeating until new themes appeared infrequently. The demographics of the participants are shown in Table 6.2. Our participant sample

[3] https://prolific.co

skewed younger, compared to the overall U.S. population, and contained more ad blocker users than some estimates [135].

We ran our survey between June 24th and July 14th, 2020. Participants were paid $3.00 to complete the survey (a rate of $13.85/hr). Our survey did not ask participants for sensitive or identifiable information, and was reviewed and deemed exempt from human subjects regulation by our Institutional Review Board (IRB).

3.3.2 Survey 1 Results

Table 3.2 summarizes the reasons participants liked or disliked ads, based on the codes developed in our qualitative analysis.

Negative Reactions and Feelings Towards Ads

Clickbait The term “clickbait” was used by participants to describe ads with three distinct characteristics: the ad is attention grabbing, the ad does not tell the viewer exactly what is being promoted to “bait” the viewer into clicking it, and the landing page of the ad often does not live up to people’s expectations based on the ad.

Participants described the attention grabbing aspects of clickbait ads with adjectives such as “sensationalist”, “eye-catching”, “scandalous”, “shocking”, and “tabloid”. One participant felt that these attention grabbing techniques were “condescending”. This style is familiar enough that participants often cited common examples:

> I hate any of the ads that say things like "You won’t believe what..." or "They’re trying to ban this video..." or nonsensical click-bait hyperbole.

Many participants observed how clickbait ads tend to omit or conceal information in the ad, to bait them into clicking it, and expressed frustration towards this tactic:

Label	Definition
Boring, Irrelevant	• The ad doesn't contain anything you're interested in • The ad is bland, dry, and/or doesn't catch your attention at all
Cheap, Ugly, Badly Designed	• You don't like the style, colors, font, or layout of the ad • The ad seems low quality and poorly designed
Clickbait	• The ad is designed to attract your attention and entice you to click on it • The ad contains a sensationalist headline, a shocking picture, or a cheap gimmick • The ad makes you click on it ad to find out what it's about • If you click the ad, it will probably be less interesting or informative than you expected
Deceptive, Untrustworthy	• The ad is engaging in false advertising, or appears to be lying/fake • The ad is trying to blend in with the rest of the website • The ad looks like it is a scam, or that clicking it will give your computer a virus/malware
Don't Like the Product or Topic	• You don't like the type of product or article being advertised • You don't like the advertiser • You don't like the politician or issue being promoted
Offensive, Uncomfortable, Distasteful	• Ads with disgusting, repulsive, scary, or gross content • Ads with provocative, immoral, or overly sexualized content
Politicized	• The ad is trying to push a political point of view onto you • The ad uses political themes to sell something • The ad is trying to call out to and use your political beliefs
Pushy, Manipulative	• The ad feels like it's "too much" • The ad demands that you do something • The ad tries to make you feel fear, anxiety, or panic
Unclear	• The ad is hard to understand • Not sure what the product is in the ad • Not sure what the advertiser is trying to sell or promote
Entertaining, Engaging	• The ad is funny, clever, thrilling, or otherwise engaging and enjoyable • The ad is thoughtful, meaningful, or personalizes the thing being sold • The ad gives you positive feelings about the product or advertiser
Good Style and/or Design	• The ad uses eye catching colors, fonts, logos, or layouts • The ad is well put together and high quality
Interested in the Product or Topic	• You are interested in the type of product or article being advertised • You like the advertiser • You like the politician or issue being promoted
Simple, Straightforward	• It is clear what product the ad is selling • The message of the ad is easy to understand • The important information is presented to you up front
Trustworthy, Genuine	• You know and/or trust the advertiser • The product or service in the ad looks authentic and genuine • The ad clearly identifies itself as an ad • Reviews or endorsements of the product in the ad are honest
Useful, Interesting, Informative	• The ad provided information that is useful or interesting to you • The ad introduced you to new things that you are interested in • The ad offered good deals, rewards, or coupons

Table 3.2: The categories of reasons that participants gave for liking or disliking ads, in response to our qualitative Survey 1 (Section 3.3). The top part of the table shows negative categories and the bottom part (below the double-line) shows positive categories. We used these categories as labels for Survey 2 participants, who were also provided with the corresponding definitions (Section 3.4).

> I dislike when an ad doesn't state its actual product...it feels clickbaity, desperate, and lacking confidence in its product.
>
> What is the product? Why do I have to click to find out?

Participants also described the tendency for clickbait ads to fail to meet their expectations, and past experiences where they regretted clicking on such ads. Examples of this include ads for "listicles" or content farms (e.g. Figure 3.2e).

> I know that any of the "##+ things" sites will end up being a slideshow (or multiple page) site that is covered with advertising and slow loading times. It is also likely that the image in the ad is either not included at all, or is the last one in the series.

Psychologically Manipulative Ads Participants disliked when ads tried to manipulate their emotions and actions, such as ads make them feel unwanted emotions, e.g. anxiety, fear, and shock; or ads that "loudly" demand to be clicked or paid attention to. A common example was a dislike of "fearmongering":

> I can't stand ads like this at all. What I dislike most is the "shocking" photo they use to try to scare people into clicking this ad and being fear mongered. They are most likely trying to sell a pill or treatment for this "health condition" that they made up.

Some participants reacted negatively to strong calls-to-actions, such as a political ad which said "Demand Answers on Clinton Corruption: Sign the Petition Now".

> I don't like the political tone and how it asks to demand answers. I feel like it's my personal choice what I should and I shouldn't do, they don't need to tell me.

More generally, participants commented on how some ads manipulate people's emotions; one participant disliked ads that are "prying on emotions/sickness", another characterized an advertiser in our study as "impulse pushers" that "use too much psychology in a negative way".

Distasteful, Offensive, and Uncomfortable Content Participants reacted to some ads in our survey with disgust, such as the ad showing a dirty laundry machine, and an ad with a poorly lit picture of sprouting chickpeas in front of a banana (Figure 3.2d). Participants reacted to these ads with words like “gross”, “disgusting”, and “repulsed”.

Some participants had similar reactions to content that they found offensive or immoral. For example, in an ad for the Ashley Madison online dating service, premised on enabling infidelity, one participant said:

> I dislike that this ad for many reasons, one of them being the idea that a person should leave their partner for a hotter one. Gross.

Others reacted negatively to ads that they perceived as unnecessarily sexually suggestive, or was “using sex to sell”.

Cheap, Ugly, and Low Quality Ads Participants disliked the aesthetics of some ads, describing them as “cheap”, “trashy”, “unprofessional”, and “low quality”. Some features they cited include poor quality images, the use of clip art images, “bad fonts”, or a feeling that the ad is “rough” or “unpolished”. Some participants felt that the poor quality of the ad reflected poorly on the product, saying that it makes the company look “desperate”, or that it made them think the ad looked like a scam. Participants also disliked specific stylistic choices, like small fonts or “garish”, too-bright colors.

Dislikes of Political Content in Ads Participants disliked politics in their ads, for different reasons. Most obviously, some participants disliked ads when they disagreed with the politician or political issue in the ad. Others disliked political ads because they dislike seeing any kind of political message or tones in an advertisement:

> [I dislike] everything. At least there’s no stupid President in my face, but come on,

> get your politics and agenda away from me. I even agree with this ad but it's still managing to annoy me! Go away!

Some participants observed that ad that looked like a political poll (Figure 3.2f) was intended to activate their political beliefs and lure them into clicking to support their preferred candidate.

> The ad makes me feel fear that the opposite political party will win, and it makes me feel pride towards my own political party. I feel like I need to answer the question on this ad to help promote my preferred candidate.

> It calls to the political side of people in order to lure into their ad. It is probably just a scam.

Untrustworthy and Deceptive Ads Participants disliked ads that felt untrustworthy to them, describing such ads using words like "deceptive", "fake", "misleading", "spam", and "untrustworthy".

Related to "clickbait", participants mentioned disliking "bait and switch" tactics, where something teased or promised in an ad turns out not to exist on the landing page.

> I don't like ads that mislead what the application/ product actually does. For example, there are sometimes ads that show a very different style of gameplay for an app than is actually represented.

Participants were also sensitive to perceived lies, false advertising, and fake endorsements. For the ad headlined "US Cardiologist: It's Like a Pressure Wash for your Insides" (Figure 3.2d), a participant said:

> "U.S. Expert" — who is it? It sounds like a lie.

Participants also disliked visually deceptive ads. Several participants called out an ad that appeared to be a phishing attempt (Figure 3.2g):

> I don't like ads that try to deceive the user, or use buttons like "continue" to try to get

> them to be confused with what is an ad and what is part of the site.

Some participants disliked ads labeled as "sponsored content", seeing through the attempts to disguise the ad as content they would be interested in.

> I dislike everything about this ad, because from my experience, this ad leads to an article that pretends to be an informed article, but is actually paid by one of the phone companies to advertise their brand.

> What I dislike is the paid product placement, disguised as a genuine article.

Scams and Malware Many participants suspected that the ads that they did not trust were scams, or would somehow infect their computer with viruses or malware.

> It just looks like a very generic ad which would give you a virus. It doesn't even state the company, etc.

> I disliked all of this ad. Just by glancing at the headline, it seems like a scam and does not seem like it is from a reputable source. The image doesn't really add much either. I don't like how the company/brand is in a tiny box either. It's like they're trying to hide it somehow?

Some participants suspected ads of spreading scams and malware whether or not the ad had to do with computer software. For example, for a suspicious ad about mortgages, one participant said:

> It seems like a scam. The graphics are badly done and it seems like it would sell my information to someone else or download a virus.

Boring, Irrelevant, and Repetitive Ads Participants generally reported disliking ads which bored them, were not relevant to their interests, or ads that they saw repeatedly (on the web in general, not in our survey).

Unclear Ads Many participants found some of the ads shown to them in the survey to be confusing and unclear. A common complaint was that it was unclear from looking at the ad what exactly the product was; participants said this about ads from both the problematic and benign categories (e.g. Figure 3.2b and c).

Targeted Advertising While perceptions of privacy and targeted advertising were not the main focus of this study, some participants mentioned these as concerns when asked about ads they disliked in general. Three participants mentioned disliking retargeted advertisements, i.e., ads for products which they had looked at previously, as they found these ads repetitive.

Other Disliked Topics and Genres of Ads When asked about what ads they disliked in general, participants called out other specific examples and genres of ads, unprompted by the ads we showed in the survey. 10 participants independently said they disliked ads for video games, particularly mobile game ads that use dishonest bait-and-switch tactics. Some participants mentioned disliking certain kinds of ads on social media, like ads for “drop-shipping” schemes, ads with endorsements perceived to be inauthentic, and ads that “blend in” to the feed. Participants also mentioned disliking specific topics such as dating ads, celebrity gossip ads, beauty ads, and diet/supplement ads.

Positive Reactions to Ads

We now turn to participants’ positive reactions to ads. While our primary research questions are around negative reactions, we also wish to characterize the full spectrum of people’s reactions to ads, especially when people might have different opinions about the same ads (e.g., one person might find annoying an ad that another finds entertaining), and to help identify types of ads that do not detract from user experience.

Trustworthy and Genuine Ads Participants responded positively to ads that they described as "honest", "trustworthy", "legitimate", and "authentic". Some signals people cited for these traits include ads that look "refined" and high quality, images that accurately depict the product, and ads that include brands they recognized.

Good Design and Style Participants liked aesthetically pleasing ads, including ads with appealing visuals like pleasing color choices, images that are "eye-catching", interesting, beautiful, or amusing, and a "modern" design style.

Entertaining and Engaging Ads Participants liked ads that they found entertaining, engaging, or otherwise gave them positive feelings. They variously described some of the ads as "humorous", "clever", "fun", "upbeat", "calming", "unique", and "diverse".

Relevant, Interested in the Product Many participants, when asked about what kinds of ads they liked in general, said that they enjoyed ads which were targeted at their specific interests. Various participants mentioned liking ads for their specific hobbies, food, pets, and for products they are currently shopping for, etc.

Simple and Straightforward Participants appreciated ads that were easy to understand, and straightforward about what they were selling.

Some participants mentioned that it was important that ads were clearly identifiable as ads, present information up front, and clearly mention the brand:

> When I am browsing, I enjoy ads that are unique and advertise the brand name clearly, without disrupting the content I am viewing. Specifically, side banners and top banners are fine

And others appreciated direct approaches, as opposed to "clever" tactics or other appeals:

> Simple–not too pushy. If I'm looking for insurance it's there. Not trying to be too clever or emotional. Nice palette–few and easy-to-focus-on visuals

Useful, Interesting, and Informative Participants liked ads that provided them with useful information. Some participants genuinely liked seeing ads to discover new products:

> I like seeing ads of events happening nearby me and products concerning sports and electronics because i feel they are in a way an outlet for me to know whats out there.

Others appreciated when the ads were informative about the product being sold:

> [Explaining why they like a clothing ad] The picture of the guy. It gives me a good idea of what it would look like on me.

Stepping back, we organize and summarize the taxonomy of both positive and negative reactions that participants had ad content in Table 3.2. We note that participants did not always agree on their assessment of specific ads — some of the positive and negative reactions we reported above referred to the same ads, suggesting that a range of user perceptions and attitudes complicates any assessment of a given ad as strictly "good" or "bad". We explore this phenomena quantitatively, and in greater detail, in the next section, and we return to a general discussion combining the findings from both of our surveys in Section 3.5.

3.4 Survey 2: Which Ads Do People Dislike?

Equipped with our taxonomy of reasons that people dislike ads from Survey 1, we now turn to our second research question: specifically *which* ads do people dislike, and for which reasons? What are the specific characteristics of ads that evoke these reactions? Can we characterize ad content on a spectrum, ranging from ads that people nearly universally agree are "bad" or "good" to the gray area in between where subjective opinions are mixed?

# of Ads	Site Type	Example Domains
412	News, Media, and Blogs	nytimes.com, food52.com
27	Non-Article Content	marvel.com, photobucket.com
22	Reference	merriam-webster.com, javatpoint.com
17	Software, Web Apps, and Games	speedtest.net, armorgames.com
13	Social Media and Forums	slashdot.org, serverfault.com
9	E-Commerce	amazon.com, samsclub.com

Table 3.3: Categories of websites that ads in Survey 2 appeared on. Ads primarily appeared on news, media, and blog websites.

To answer these questions, we collected a large (new) dataset of ads from the web and surveyed a large number of participants. At a high level, we (1) collected a dataset of 500 ads that we randomly sampled from ads appearing on the top 3000 U.S.-registered websites, (2) asked 1000 participants to rate and annotate 5 ads each with one or more *opinion labels*, derived from our taxonomy from Survey 1, (3) manually labeled each ad ourselves with on *content labels* to describe objective properties of the ads (e.g., topic, format), and (4) analyzed the resulting labeled ad dataset.

3.4.1 Survey 2 Methodology

Ads Dataset

We wanted to collect participant ratings on a large, diverse dataset of actual ads from the web. Thus we created a new dataset by crawling the top 3000 most popular U.S.-registered websites, based on the Tranco top ranked sites list [123], matching the crawling methodology used to collect the ads in Survey 1 [226].

We crawled these sites using a custom-built web crawler based on Puppeteer, a browser automation library for the Chromium browser [88]. When the crawler visits a site, it identifies ads using the Easylist [57], a popular list of definitions used by ad blockers, and takes screenshots of each ad on the page. Our crawler visited the home page of each domain in the top 3000 list, scraped any ads found on the page, and then attempted to find a subpage on the domain that contained ads (to account for cases where the home page did not have ads but a subpage did) by clicking links on the home page, scraping ads if a page with ads was found. Each crawler ran in a separate Docker container, which was removed after crawling each domain to remove all tracking cookies and other identifiers.

Most of the ads in the dataset came from online news sites, blogs, and articles. We categorize the type of sites the ads appeared on in Table 3.3. Matching the types of ads in Survey 1, the ads we collected consisted primarily of third party programmatic ads on news and content sites, such as banner ads, sponsored content, and native ads, and excluded social media, video, and retargeted ads (as our crawler did not explore social media feeds, and deleted its browsing profile between sites).

We ran our crawl on July 30th, 2020. We crawled 7987 ads from 854 domains (2146 domains did not contain ads on the home page or the first 20 links visited). We filtered out 2700 ads that were blank or unreadable, due to false positives, uninitialized ads, or ads occluded by interstitials such as sign up pages and cookie banners, and 3359 ads that were duplicates of others in the dataset, leaving 1838 valid, unique ads in our dataset. We randomly sampled 500 ads from this remaining subset for use in our survey.

Survey Protocol

We designed a survey asking each participant to evaluate five ads from our dataset. For each participant, we first collected (a) their overall feelings towards ads (7-point Likert scale, from extremely dislike to extremely like seiing ads), to provide context on their baseline feelings towards ads, and (b) whether they use an ad blocker. Then, for each of the five ads a participant labeled, we collected:

- Their overall opinion of the ad (7-point Likert scale, from extremely negative to extremely positive).

- One or more *opinion labels* describing their reaction to the ad. Participants were asked to select all that applied from the list of 15 categories derived from the previous study (Table 3.2). Participants were given the definitions of those labels and could view these definitions throughout the course of the survey.
- For each opinion label they selected, their level of agreement with that label (5-point Likert scale).
- Optionally, participants could write in a free response box if the given opinion labels were not sufficient.

See Appendix A.2 for the full survey protocol.

Expert Labels of Ad Content

To understand what features and content may have influenced participants' opinions of ads, we performed a separate content analysis of the ads and generated *content labels* for each of our 500 ads. Two researchers coded the ads: the first researcher generated a codebook while coding the first pass over the dataset, the second researcher used and modified the codebook in a second iteration, then both researchers discussed and revised the codebook, and resolved disagreements between their labels. The final codes are organized into three broad categories:

- *Ad Format*, which describe the visual form factor of the ad (e.g., image, native, sponsored content);
- *Topic*, which are topical categories for the products or information promoted by the ad (e.g., consumer tech); and
- *Misleading Techniques*, such as "decoys", where an advertiser puts what appears to be a clickable button in the ad, intended to mislead users into thinking it is part of the parent page's UI [151].

A full listing of content codes with their definitions are available in Appendix A.3.

Analyzing User Opinions of Ads as Label Distributions

We expected that different participants would label the same ad with different sets of opinion labels, because of different personal preferences and experiences regarding online ads. Thus, no single opinion label (or set of labels) can represent the "ground truth" of how users perceived the ad. Instead, we assigned 10 participants to evaluate each ad in our dataset to capture the spread of possible opinions and reactions to ads, meaning that each ad had a *distribution* of opinion labels. We recruited 1025 participants (each evaluating 5 ads) to collect 10+ evaluations for each of the 500 ads in our dataset.

We analyzed the opinion labels on each ad as a *label distribution*. We count all of the opinion labels used by participants to produce a categorical distribution of labels, with each opinion label as a category. For example, a given ad might have 20% of participants label it as "Simple", 10% label it as "Trustworthy", 40% label it as "Boring/Irrelevant", and 30% label it as "Unclear".

Participants and Ethics

We recruited 1025 participants to take our survey through Prolific, and ran the study between August 20 and September 14, 2020.[4] Participants were paid $1.25 to complete the survey (a rate of $11.12/hr). Our survey, which did not ask participants for any sensitive or identifiable information, was reviewed and determined to be exempt from human subjects regulation by our Institutional Review Board (IRB). The demographics of our participant sample are shown in Table 6.2. Our sample was younger than the overall U.S. population, and contained more ad blocker users than some estimates [135].

[4] Due to a bug in the survey, we added 25 participants to our original target of 1000 to ensure each ad was labeled by at least 10 people.

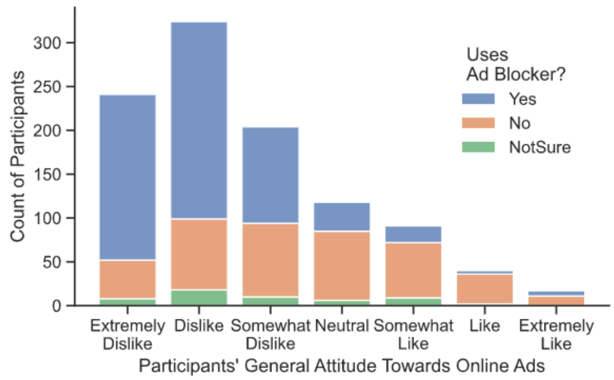


Figure 3.3: Histogram of how much participants reported liking/disliking seeing online ads in general. Overall, most participants disliked seeing ads; ad blocker users disliked ads more.

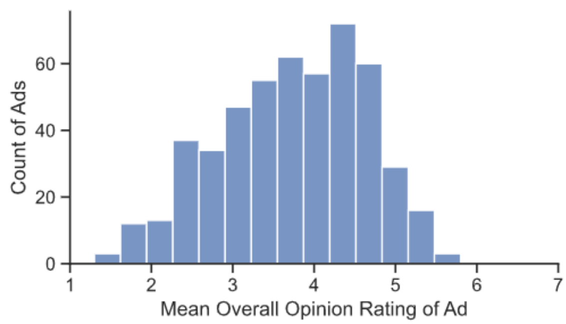


Figure 3.4: Histogram of average overall opinion rating for ads in our dataset, where the values 1-7 map to a Likert scale ranging from extremely negative to extremely positive. Ratings for ads skewed negative, with a median score of 3.8.

3.4.2 Survey 2 Results

General Attitudes Towards Ads

Our participants generally skewed towards disliking ads to begin with. Figure 3.3 shows participants' general attitude towards online ads; most participants disliked seeing ads in general, and the majority of those who dislike seeing ads use an ad blocker. 57% reported using an ad blocker, 38% did not use an ad blocker, and 5% were not sure if they used one.

	Number (%) of ads with >n% agreement		
Opinion Label	>25%	>50%	>75%
Simple	327 (65.4%)	156 (31.2%)	12 (2.4%)
Clickbait	189 (37.8%)	103 (20.6%)	23 (4.6%)
Good Design	231 (46.2%)	93 (18.6%)	10 (2.0%)
Ugly/Bad Design	212 (42.4%)	68 (13.6%)	6 (1.2%)
Boring/Irrelevant	243 (48.6%)	62 (12.4%)	4 (0.8%)
Deceptive	140 (28.0%)	56 (11.2%)	9 (1.8%)
Unclear	137 (27.4%)	38 (7.6%)	6 (1.2%)
Interested in Product	142 (28.4%)	24 (4.8%)	0 (0.0%)
Useful/Informative	103 (20.6%)	21 (4.2%)	0 (0.0%)
Dislike Product	111 (22.2%)	20 (4.0%)	1 (0.2%)
Politicized	22 (4.4%)	13 (2.6%)	3 (0.6%)
Distasteful	27 (5.4%)	8 (1.6%)	0 (0.0%)
Entertaining	56 (11.2%)	8 (1.6%)	0 (0.0%)
Pushy/Manipulative	55 (11.0%)	7 (1.4%)	0 (0.0%)
Trustworthy	62 (12.4%)	7 (1.4%)	0 (0.0%)
Any Negative Label	414 (82.8%)	226 (45.2%)	51 (10.2%)
Any Positive Label	380 (76%)	207 (41%)	21 (4.2%)

Table 3.4: The number and proportion of ads in the dataset where >25%, >50%, or >75% of participants annotated the ad with the same label. Negative labels are italicized. Note that each ad can have multiple labels with higher agreement than the threshold, so the number of ads where 50% of participants agreed on *any* negative or positive label is not simply the sum of the relevant counts.

Prevalence of "Bad" Ads

How prevalent were "bad" ads in our sample of 500 unique ads crawled from the most popular 3000 U.S.-registered websites? In this section, we analyze the quantity of ads that participants rated negatively in our dataset. While we cannot directly generalize from our sample to the web at large (due to the fact that our dataset only captures a small slice of all ad campaigns running at one point in time, and that ads may have been targeted at our crawler and/or geographic location), our results provide an approximation of how many "bad" ads web users see when visiting popular websites.

Overall Opinion of Ads in the Dataset Most ads in the dataset had negative overall opinion ratings participants. Figure 3.4 shows a histogram of the average opinion rating for each ad (on

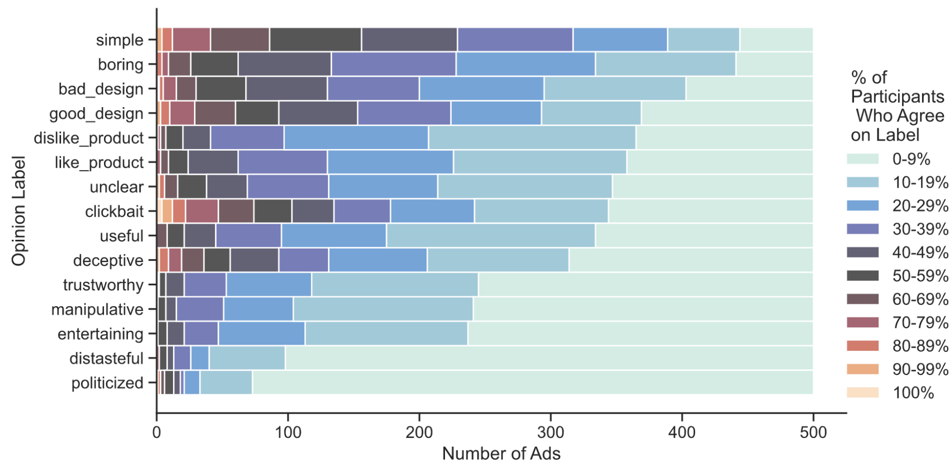


Figure 3.5: A stacked histogram representing the distribution of agreement values for each opinion labels. Each bar represents the number of ads annotated with an opinion label, subdivided into bars representing ranges of agreement values. For example, the width of the black sub-bar for the "Simple" code represents the number of ads where 50-59% of annotators labeled the ad as "Simple". The number of ads with high agreement on any label was fairly low, but specific labels like "clickbait" and "simple" had more ads with high agreement — indicating that certain ads acutely embodied this label.

a 7-point Likert scale from extremely negative (1) to extremely positive (7)). The median of the average opinion ratings across all ads was 3.8, less than the value for the "Neutral" response (4). The Fisher-Pearson coefficient of skewness of the distribution was -0.281 (a normal distribution would have a coefficient of 0), and a test of skewness indicates the skew is different from a normal distribution (z=-2.558, p=0.011), indicating that participants' perceptions of ads skew negative. Additionally, no ads had an average rating over 6, while some had ratings under 2, indicating that there were *no* ads that most people were extremely positive about.

Opinion Label Frequencies Next, we analyzed the number of ads labeled by participants with each opinion label, for example, the number of ads labeled "clickbait". Since opinion labels do not have a ground truth value, but are instead a distribution of 10 participants' opinions, we cannot simply count the number of ads labeled with each opinion label. Instead we calculated *agreement*: the percentage of participants who annotated the ad with a specific opinion label, out of all participants who rated the ad. Because agreement is a continuous value (rather than binary), we analyze the distribution of agreement values when counting the number of ads labeled a specific opinion label.

Table 3.4 shows the quantity and percentage of ads in the dataset where more than 25%, 50%, or 75% of participants agreed on an opinion label. Figure 3.5, visualizes the distribution of agreement values: for each opinion label, it shows the number of ads at different levels of agreement, in bins of width 10% (e.g. the number of ads with 40-49% agreement on the "clickbait" label).

Nearly half of ads were perceived negatively by a majority of participants — 226, or 45% — were labeled with any negative label by over 50% of its annotators. 20.6% of ads were labeled as "clickbait by a majority of annotators. Of the other negative labels, 13.6% of ads were seen as "cheap, ugly, badly designed", and 11.2% were seen as "deceptive" by a majority of their annotators. Few ads were seen as "manipulative", "distasteful", and "politicized", with fewer than 3% of ads reaching 50% agreement on those labels.

Overall, there were few ads with high agreement on opinion labels: for example, only 23 ads had over 75% agreement on the "clickbait" label. There are two likely contributing factors: first, participants may have differing, inconsistent understandings of each opinion label (we discuss this possible limitation further in Section 3.5.4). Second, participants have diverse personal preferences for advertising, and are unlikely to unanimously agree on the usage of subjective opinion labels, except in a small number of extremely good or extremely bad ads. In the next section, we leverage the subjectivity and disagreement in participants' opinion labels to identify clusters of ads with a similar "spread" of labels.

Characterizing the Content of Ads with Similar Opinion Label Distributions

Towards answering our primary research question of this survey — which ads do people dislike? — we performed a clustering analysis to identify groups of ads with similar opinion label distributions (i.e., ads participants felt similarly about), and we characterize the content of those groups of ads using our researcher-coded content labels.

Clustering Methodology We used an unsupervised learning algorithm to cluster our opinion label distributions, partially borrowing the method described by Liu et al. for population label distribution learning (PLDL), which was designed to model scenarios precisely like ours, where a small sample size of human annotators label each item using subjective criteria [132].

We use one of the unsupervised learning algorithms proposed for use in PLDL, specifically the *finite multinomial mixture model* (FMM) with a Dirichlet prior $\pi\ Dir(p, \gamma = 75)$, learned using the Variational Bayes algorithm.[5] This algorithm was found by Liu et al. [132] and Weerasooriya et al. [215] to have the best clustering performance on similar benchmark datasets, measured using Kullback–Leibler (KL) divergence, a measure of the difference between probability distributions [119]. The highest performing FMM model on our dataset was trained with 40 fixed clusters, and achieved a KL-divergence of 0.227, similar to the performance measured by Liu et al. and Weerasooriya et al. for other PLDL datasets [132, 215].

Clustering Results Overview Our model produced 16 thematically distinct clusters, which we summarize in Table 3.5 (ordered by decreasing average participant rating of the ads overall in each cluster). We removed 5 additional clusters which contained three or fewer ads and/or had dissimilar opinion and content labels (these account for the missing alphabetical cluster names). Next, we describe findings based on a qualitative analysis of these clusters, with examples and

[5] We used an implementation of FMM-vB in the bnpy library (https://github.com/bnpy/bnpy) [100].

free-responses from participants.

Clickbait Ads and Native Ads We found 4 clusters (R, S, T, and U) where a majority of participants labeled ads as "clickbait" (61-68% of labelers). Participants disliked the ads in these clusters: they represent the four lowest-ranked clusters in terms of participants' overall opinion, with average ratings ranging from 2.21 to 2.8 (on a 1-7 scale).

These clusters contain a diverse set of ad content including listicles, potentially fraudulent supplements, sexualized images, and tabloid news. The common thread among them is that many are *native ads* (43%-72% of ads in these clusters), also known as content recommendation ads, or colloquially as "chumboxes" [134]. These ads imitate the design of links to news articles on the site, and have been considered borderline deceptive by the FTC and researchers [68, 121, 101, 218].

Numerous participants suspected that "clickbait" native ads, such as the ad in Figure 3.6a, are content farms:

> It seems like this ad would lead to an actual article but I think the website would be loaded with other advertisements.

They also commented on the tendency for listicle-style native ads to do a bait-and-switch.

> (Figure 3.6a) It tries to fool into clicking something that may or may not have anything to do with the add by giving me misleading or tangential information in the headline.

Participants also found the lack of clear disclosure of the advertiser or brand in native ads confusing:

> It's difficult to tell that this is an ad rather than a legitimate recommended article.

Clickbait and Distasteful Content Cluster R contains a high number of "clickbait" ads containing sexualized or gross images, mostly in the native ad format. We counted 12 ads featuring sexually suggestive pictures of women and 2 of men, mainly for human interest "listicles". We also counted 5 pictures participants described as "gross" and disgusting, like dogs eating an unknown

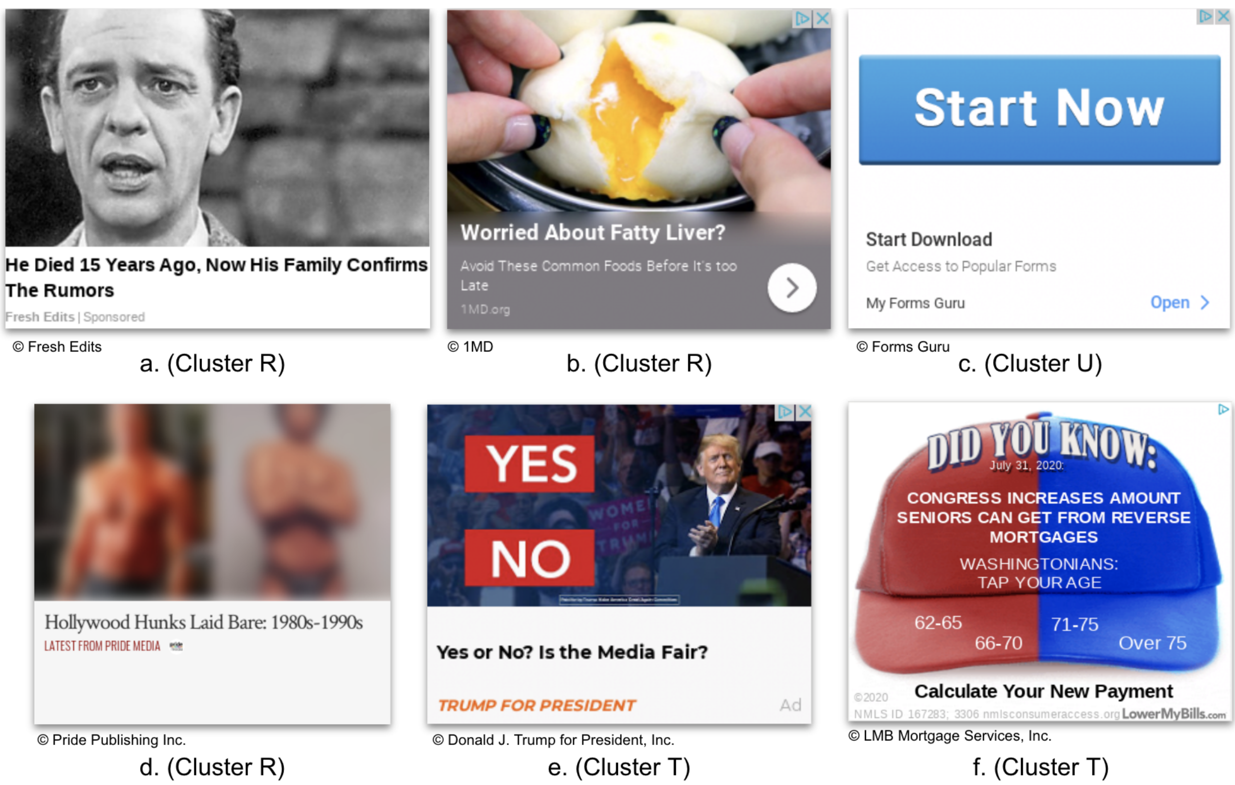


Figure 3.6: Examples of "bad" ads from the Survey 2 dataset, which appeared in clusters where participants frequently used labels such as "clickbait", "deceptive/untrustworthy" "distasteful", and "politicized". (a) is a native ad, for a celebrity news "listicle", sometimes called a "content farm", an article designed to maximize ad revenue per viewer. (b) is a Google responsive display ad, for a dietary supplement, viewed by some participants as disgusting. (c) is a "decoy" software download ad, designed to look UI on the parent page, seen by participants as deceptive. (d) is a native ad for a listicle featuring sexualized imagery (blurred by us), which participants found sexist and distasteful. (e) is a political campaign ad, designed to look like a poll, seen by participants as politicized. (f) is an ad for reverse mortgages which uses political imagery to attract attention, also seen as politicized.

Cluster	# Ads	User Rating	Description	Avg. Opinion Label Dist.	Ad Formats	Top Topics	Misleading Techniques
A	13	$\mu = 5.23$ $\sigma^2 = 1.23$	Very glossy and well-designed consumer ads	Good Design (62%) Simple (45%) Like Product (40%)	Image (92%) Spon. Content (8%)	Entertainment (23%) Consumer Tech (15%) Travel (15%)	Advertorial (8%)
C	15	$\mu = 4.86$ $\sigma^2 = 1.34$	High quality simple, and well designed ads	Simple (59%) Like Product (40%) Good Design (38%)	Image (67%) Spon. Content (20%) Native (7%)	Apparel (13%) Political Content (7%) B2B Products (7%)	Political Poll (7%)
D	115	$\mu = 4.53$ $\sigma^2 = 1.29$	General pool of quality consumer ads	Simple (54%) Good Design (45%) Like Product (25%)	Image (83%) Google Resp. (10%) Spon. Content (3%)	B2B Products (16%) Household Prod. (13%) Entertainment (10%)	
E	3	$\mu = 4.52$ $\sigma^2 = 1.18$	Unclear, but well designed ads	Unclear (68%) Good Design (67%) Simple (32%)	Image (100%)	Apparel (33%) B2B Products (33%) Sports (33%)	
G	59	$\mu = 4.07$ $\sigma^2 = 1.64$	Average quality consumer ads, incl. native ads	Simple (35%) Good Design (32%) Clickbait (29%)	Image (46%) Native (31%) Google Resp. (14%)	COVID Products (14%) Consumer Tech (10%) Food and Drink (8%)	Advertorial (10%) Spon. Search (7%) Listicle (5%)
I	101	$\mu = 3.81$ $\sigma^2 = 1.35$	Average quality niche interest or B2B ads	Simple (38%) Boring/Irrelevant (37%) Unclear (27%)	Image (70%) Google Resp. (24%) Spon. Content (5%)	B2B Products (39%) Journalism (10%) Apparel (9%)	Spon. Search (3%)
J	4	$\mu = 3.67$ $\sigma^2 = 1.72$	Boring, mildly politicized ads	Simple (40%) Politicized (35%) Boring/Irrelevant (33%)	Image (50%) Spon. Content (25%) Google Resp. (25%)	Weapons (25%) Journalism (25%) Political Campaign (25%)	
L	46	$\mu = 3.29$ $\sigma^2 = 1.48$	Average quality B2B ads and native Ads	Ugly/Bad Design (38%) Boring/Irrelevant (34%) Deceptive (34%)	Google Resp. (46%) Image (33%) Native (15%)	B2B Products (30%) Software Download (13%) Health/Supplements (9%)	Advertorial (11%) Spon. Search (9%)
M	10	$\mu = 3.13$ $\sigma^2 = 1.62$	Generally political content; TV shows, political T-shirts	Politicized (39%) Dislike Product (30%) Ugly/Bad Design (28%)	Image (90%) Poll (10%)	Apparel (20%) Political Content (20%) Journalism (20%)	Political Poll (10%)
N	3	$\mu = 3.12$ $\sigma^2 = 1.39$	Strongly disliked products; e.g. vape pens	Dislike Product (71%) Good Design (31%) Pushy/Manipulative (31%)	Image (100%)	Journalism (33%) Recreational Drugs (33%) Education (33%)	
P	15	$\mu = 3.04$ $\sigma^2 = 1.28$	Vague/unclear ads; no visible brand names	Unclear (54%) Boring/Irrelevant (53%) Ugly/Bad Design (46%)	Google Resp. (47%) Image (40%) Poll (7%)	B2B Products (73%) Humanitarian (7%) Sports (7%)	Spon. Search (7%) Advertorial (7%)
Q	29	$\mu = 2.95$ $\sigma^2 = 1.56$	Ugly ads and confusing clickbait ads	Ugly/Bad Design (48%) Clickbait (47%) Boring/Irrelevant (35%)	Native (38%) Google Resp. (31%) Image (28%)	Household Prod. (14%) B2B Products (10%) Investment Pitch (10%)	Advertorial (21%) Listicle (7%) Spon. Search (3%)
R	39	$\mu = 2.8$ $\sigma^2 = 1.57$	Clickbait; sexualized and distasteful content	Clickbait (63%) Deceptive (37%) Ugly/Bad Design (28%)	Native (72%) Google Resp. (15%) Image (10%)	Human Interest (23%) Health/Supplements (23%) Celebrity News (10%)	Listicle (41%) Advertorial (28%)
S	2	$\mu = 2.52$ $\sigma^2 = 1.47$	Politicized native ads	Politicized (72%) Clickbait (53%) Boring/Irrelevant (52%)	Native (50%) Google Resp. (50%)	Senior Living (50%) Political Campaign (50%)	Political Poll (50%) Listicle (50%)
T	7	$\mu = 2.31$ $\sigma^2 = 1.44$	Deceptive and politicized ads; using politics as clickbait	Clickbait (61%) Deceptive (55%) Politicized (42%)	Native (43%) Image (29%) Google Resp. (29%)	Mortgages (29%) Human Interest (14%) Political Memorabilia (14%)	Listicle (29%) Advertorial (29%) Political Poll (14%)
U	31	$\mu = 2.21$ $\sigma^2 = 1.3$	Scams; supplements and software downloads	Clickbait (68%) Deceptive (61%) Ugly/Bad Design (45%)	Native (45%) Google Resp. (32%) Image (23%)	Health/Supplements (32%) Software Download (29%) Computer Security-related (10%)	Advertorial (26%) Decoy (23%) Listicle (13%)

Table 3.5: Ads in out dataset clustered by user opinion label distribution. "User Rating" shows the average overall rating of ads in the cluster (1-7 scale). "Description" qualitatively summarizes the ads in the cluster. "Average opinion label distribution" shows the mean percentage of participants who labeled an ad using the listed opinion labels. "Ad Formats", "Top Topics", and "Misleading Techniques" show the percentage of ads in the cluster labeled with the listed content label.

purple substance, and a dirty toilet. On average, 27% of participants labeled ads in this cluster as "distasteful", the highest percentage for that label in any cluster. Participants reacted negatively to these ads (the average opinion rating was 2.8) and described their visceral dislike of the ads in the free response:

> (Figure 3.6b) The picture of the egg yolk oozing out looks disgusting. The ad also uses threatening language such as "before it's too late".

In response to a particularly sexually suggestive ad:

> Blatant soft-porn sexism. Completely disgusting.

Clickbait and Deceptive Content Cluster U contains "scammy" clickbait ads — on average 61% of participants labeled ads from this cluster as deceptive. This cluster also has the lowest average rating from participants of all of the clusters ($\mu = 2.21$), indicating a wide dislike for deception in advertising. Software download ads that used decoys and phishing techniques were common (29% of ads in the cluster), such as ads for driver downloads, PDF readers, and browser extensions (Figure 3.6c).

> Looks like an advertisement a scammer would use to get you to download bad software on to your computer.

We also observed numerous ads for supplements (32% of ads in the cluster) which claimed to help with conditions such as weight loss, liver health (Figure 3.6b), and toenail fungus, but we did not find ads for legitimate prescription drugs or medical services here, suggesting that people consider supplements to be particularly "scammy" or deceptive.

Clickbait and Politicized Ads Clusters S and T encompass ads that participants frequently rated as both "politicized" and "clickbait". Of the 9 ads in these two clusters, two were ads from a political campaign, both from U.S. President Donald Trump's re-election campaign. Both of these ads present themselves as a political poll, asking “Approve of Trump? Yes or no?” (Figure 3.6e), and “Yes or No? Is the Media Fair?”, likely a tactic to bait users to click, exploiting a desire to make their political opinions heard.

The remainder of the politicized ads were not for political campaigns, but used political themes to attract attention. For example, we found ads that prominently used symbols associated with Donald Trump: an ad for mortgage refinancing that uses imagery reminiscent of the “MAGA” hat (Figure 3.6f), and a native ad for a commemorative coin, with the headline “Weird Gift from Trump Angers Democrats!”.

Participants broadly disliked these politicized ads; the average opinion rating was 2.31 and 2.52 for clusters T and S respectively. The low ratings may in part be due to the political beliefs of our participants: 5 of 9 ads support or use pro-Trump imagery, and our participant pool skewed Democratic: 51% identified as Democrats, 16% as Republicans, and 26% as Independent.

Other Negatively Perceived Ad Clusters

- *Cheap, Ugly, and Badly Designed Ads:* Participants appear to dislike visually unattractive ads in general. Cluster Q contains ads that do not have much in common in terms of content, but on average, 48% of participants labeled ads in this cluster as poorly designed, with an average opinion rating of 2.95.
- *Unclear or Irrelevant Business-to-Business (B2B) Product Ads:* Participants rated ads in cluster P as unclear and boring/ irrelevant, on average 54% and 53% of participants per ad respectively, and the overall rating was 3.04. 73% of these ads were aimed at businesses and commercial customers, indicating that these ads were likely confusing and not relevant to participants. Many ads also used Google’s Responsive Display ad format (47%), which sometimes lacked images, potentially adding to participants’ confusion.
- *Strongly Disliked Products:* Cluster N contained only three ads, but 71% of participants on

average said they disliked the product (overall rating was 3.12). These ads contain socially undesirable products: vape pens, a medical school advertising that it does not require an MCAT exam score, and subscriptions to a tabloid magazine.

- *Non-Clickbait Politicized Ads:* Cluster M contains ads that on average 39% of participants labeled as politicized, 30% as disliking the product, and 28% as boring or irrelevant. Compared to the political ads in clusters S and T, most of these ads do not employ clickbait or deceptive tactics. They include ads like for political TV programming (e.g., Fox News) and political T-shirts. In general, participants appear to dislike these ads (the overall average rating was 3.13) more because they disagree with the politics than concerns about the ad's design or tactics.

"Good" or Neutral Ads The remainder of the clusters contain ads participants rated only slightly below average, or above average (with average overall opinions 3.29-5.23). Factors characterizing these clusters included:

- *Attractive Ads:* Participants' favorite cluster of ads (A) contained glossy, visually appealing image ads (Figure 3.7a), for popular products, like TV shows (Figure 3.7b), travel destinations, and dog food. For the average ad in cluster A, participants labeled it as "good design" (62%), "simple", (45%), and "interested in product" (40%).
- *High Relevance — Consumer Products:* Clusters C, D, and G contain a large number of ads for many different types of consumer products, ranging from mobile apps to face masks (Figure 3.7c) to lotions (Figure 3.7d). The format of most of these are image ads, rather than native ads. Participants viewed these clusters positively, with overall opinion ratings of 4.07-4.86, and as simple, well-designed, and relevant to their interests.
- *Low Relevance — B2B Products and Niche Products:* Clusters I and L contain many ads for commercial and business customers, e.g., ads for cloud software (Figure 3.7e). They also

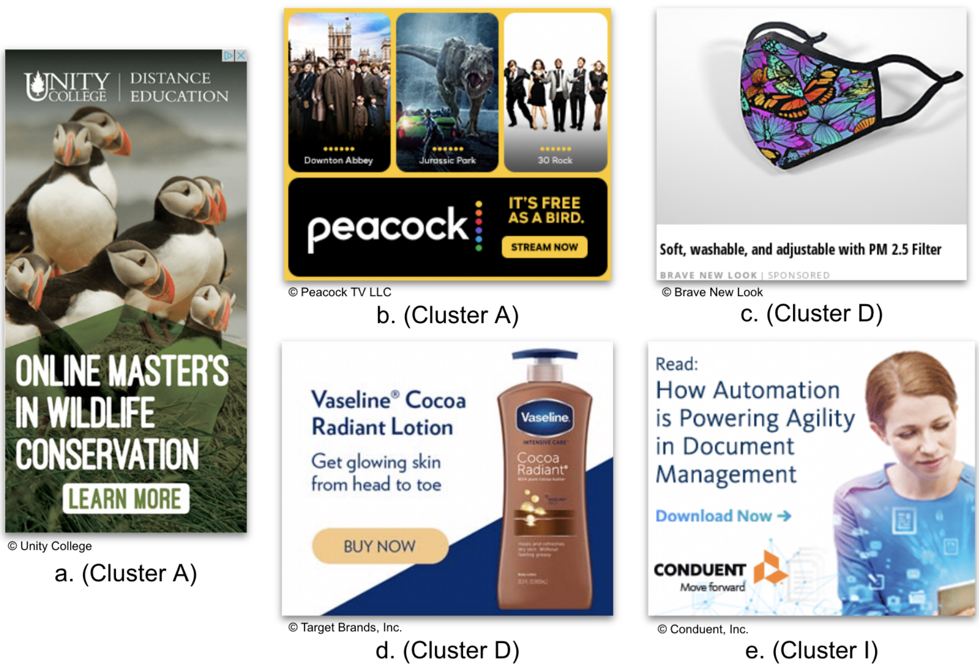


Figure 3.7: A sample of "good" ads from our Survey 2 dataset, that participants labeled positively, with labels like "good design", "simple", and "interested in product". (a) and (b) are ads for consumer services, and were in the highest rated cluster because of attractive visuals and appealing products. (c) and (d) are also consumer products from clusters with above average ratings. (e) is an example of an ad for B2B or enterprise products, which participants didn't find problematic, but rated as boring or not relevant.

Effect	Estimate	Std. Error	p-value
(Intercept)	3.325	0.060	<0.001***
Interested in Product	0.164	0.009	<0.001***
Entertaining	0.163	0.012	<0.001***
Attitudes towards Ads	0.139	0.013	<0.001***
Good Design	0.138	0.008	<0.001***
Simple	0.134	0.007	<0.001***
Useful	0.127	0.010	<0.001***
Trustworthy	0.114	0.011	<0.001***
Unsure If Uses Ad Blocker	0.086	0.082	0.292
Unclear	-0.057	0.008	<0.001***
Politicized	-0.078	0.017	<0.001***
Boring	-0.083	0.007	<0.001***
Uses Ad Blocker	-0.085	0.040	0.034*
Pushy/Manipulative	-0.087	0.011	<0.001***
Ugly/Bad Design	-0.106	0.008	<0.001***
Clickbait	-0.131	0.008	<0.001***
Deceptive	-0.133	0.009	<0.001***
Distasteful	-0.169	0.016	<0.001***
Random Effects	Variance	Std. Dev.	
Participant (Intercept)	0.1661	0.4075	
Ad (Intercept)	0.313	0.1769	
Residual	0.6674	0.8169	

Table 3.6: Summary of linear mixed model of participants' overall ratings for individual ads, with opinion labels and attitudes towards ads as fixed effects. Negative labels (italicized) such as "distasteful" and "clickbait" have a negative impact on ad ratings, while positive labels have a positive impact. Prior ad attitudes are positively correlated with ratings for ads, while ad blocker usage has a negative effect.

contain consumer products, but ones with narrower appeal, like specific articles of clothing or specific residential real estate developments. These ads, likely less relevant to the average person, scored slightly lower than the consumer products clusters above, with scores of 3.81 and 3.29, and more frequent use of labels like boring or irrelevant (34-37%).

Impact of Individual Opinion Labels on the Overall Perceptions of Ads

Lastly, we investigate which of the reasons people dislike ads impact their overall opinion of an ad most adversely. We fit a linear mixed effects model, with participants' overall opinion rating as the outcome variable, and the opinion labels as fixed effects. We also modeled other context participants provided in the survey as fixed effects: their general feelings towards online ads (1-7 Likert scale), and whether they used an ad blocker. We modeled the participant and the ads as random effects.

```
Opinion of Ad = Σ(Opinion Labels) + Use Ad Blocker?  + General
Opinion of Ads + (1|Participant) + (1|Ad)
```

We report the results of the maximal random effects structure in Table 3.6, in order of coefficient estimates (the effect size for each variable). We found that, as expected, positive opinion labels are correlated with higher ratings, and negative opinion labels are correlated with lower ratings. The negative opinion labels that had the largest effect on opinion ratings were "Distasteful, Offensive, Uncomfortable", "Deceptive", and "Clickbait", which had nearly twice the effect of labels like "Boring, Irrelevant", suggesting that these opinion categories qualitatively describe "worse" traits in ads.

Participants opinion ratings were also affected by their overall attitudes towards ads; participants who self-reported liking ads more in general rated specific ads more positively, and participants who used ad blockers tended to rate ads more negatively, though these factors had less effect than the opinion labels (i.e. their substantive perception of the ad).

3.5 Discussion

Finally, we discuss the implications of our findings for policy and future research on online advertising.

3.5.1 Broader Lessons on the Online Advertising Ecosystem

Ad Content Policy Is Lagging Behind User Perceptions

Our results show that users are unhappy with a significant proportion of ads. On average, participants gave 57.4% of ads in our sample a lower user rating than "neutral", and 20% of ads had a lower average rating than "somewhat negative". In the clusters of ads with the lowest ratings, users labeled the content of these ads as deceptive, clickbait, ugly, and politicized. These ads often contained supplements and software downloads, ads with sexually suggestive and distasteful pictures, and product ads leveraging political themes. Clusters with low user ratings also had a much higher proportion of native ads than clusters with high user ratings.

Though most advertising platforms have policies against inappropriate ad content (e.g. Google Ads prohibits malware, harassment, hacked political materials, and misrepresentation [87]), it appears that these policies are insufficient. Though we did not observe acutely malicious ads in our study, it appears that a significant proportion of ads do not meet users' expectations for acceptability.

Misaligned Incentives and Distributed Responsibility for Content Moderation

Our results indicate that bad ad content is a problem specifically in (but not limited to) the programmatic ads ecosystem. Our sample of ads came mostly from news and content websites, who generally use third parties to supply programmatic ads (unlike social media platforms like Facebook, which deliver ads end-to-end).

So who in the programmatic advertising ecosystem is currently responsible for moderating ad content? It is unclear, because ads are delivered to users via a complex supply chain of ad tech companies, and any one of them could play a role. Advertisers work with ad agencies to create their ads. Agencies run the ads via demand side platforms (DSPs), who algorithmically bid on ad exchanges to place the ads on websites. Available ad slots are submitted to ad exchanges by supply side platforms (SSPs), who publishers work with to sell ad space on their websites [7]. Some publishers also use ad quality services to help monitor and block bad ads on their websites.

The distributed nature of this system creates a disincentive for any individual ad tech company to act on "bad" ads. In related work on how programmatic ads support online disinformation, Braun et al. [28] found that the large number of companies in the marketplace creates a prisoner's dilemma "wherein each individual firm has an incentive (or an excuse) to do business with 'bad actors.' " For example, if one SSP decided to stop allowing ads from a DSP that has been providing too many low quality ads, then they would lose the sales volume, and another SSP would replace them. And because intermediate actors do not deal directly with advertisers, publishers, or users, they can dodge responsibility by "pointing the finger at other firms in the supply chain". Indeed, industry reports suggest that both SSPs and DSPs alike are seen as "not doing enough" to "stop bad ads" [130, 148, 188].

Moreover, even publishers themselves are incentivized to run "bad ads" at times, despite potential reputational risk [47], because "bad ads" can help generate revenue, especially when legitimate advertisers pull back on spending [24].

Our findings on the gap between users and current policy, combined with others' findings on the incentives of the programmatic advertising marketplace, suggest that challenging, structural reforms are needed in the online ads ecosystem to limit "bad ads" that harm user experience.

3.5.2 Recommendations

We propose policy recommendations for ad tech companies, regulators, and browsers to address structural challenges in the online ads ecosystem that have enabled the proliferation of "bad ads".

Immediate Policy Changes

In the short term, we suggest that SSPs, DSPs, and publishers implement policy changes to ban some of the characteristics that our participants found to be problematic, such as "clickbait" content farms, distasteful food pictures, political ads designed like polls (Figure 3.6a, b, and e), and to invest in their content moderation efforts to screen these types of ads more effectively.

We also suggest that the U.S. Federal Trade Commission (FTC) explore expanding their existing guidance on deceptively formatted advertisements to cover characteristics of ads in our study that users viewed as deceptive. Though the current guidance [71] and enforcement policy [70] focuses on disclosure practices of native ads, further scrutiny could be applied to other forms of deceptive formatting. For example, this might apply to software ads whose primary visual element is a large action button labeled "Download" or "Continue", and contains little information about the product or advertiser (e.g. Figures 3.2g and 3.6c).

Incorporating User Voices in the Moderation of Ad Content

In the long term, we recommend that the online advertising industry incorporate users' voices in the process of determining the acceptability of ad content. As we discussed above, there is a gap between the types of ads that users find acceptable and the policies of the online ad ecosystem, and the current system does not sufficiently incentivize ad tech to prioritize quality user experience.

We propose that the advertising industry implement a standardized reporting and feedback system for ads, similar to those found on social media posts. Users could provide reasons for why they want to hide or report the ad, based on our proposed taxonomy of user perceptions of ads (Table 3.2). User reports could be propagated back up each layer of the programmatic supply chain, so that all parties involved with serving the ad are notified. Ad tech companies could temporarily take down and review ads that exceed a user report threshold, and adjust their content policies if necessary. Eventually, user reports could be used to train models to detect and flag potentially problematic ads pre-emptively.

User feedback mechanisms do exist on display ads from Google Ads, which include an “X” icon near the AdChoices badge in the top right. However, this mechanism has not been adopted widely in the ecosystem, since it is purely voluntary. Additionally, users are likely unaware of the existence of this feature; a previous usability study found serious discoverability issues with AdChoices UIs [79].

We suggest two policy approaches that could encourage greater adoption of effective user feedback systems in online advertising. First, browser vendors could require that third-party ad frames implement feedback mechanisms, or else block the ad from rendering, similar to Google Chrome’s policy of blocking poor ad experiences [180]. Second, through regulation or legislation, online ads could be required to include a mechanism for user feedback, and ad tech companies could be required to provide transparency about the number of reports they receive.

3.5.3 Future Research Directions

Measuring “Regret”: Time and Attention Wasting Ads

Which kinds of ads do people “regret” clicking on, and how often do people do so? Participants in our study anecdotally reported that they “regretted” clicking on ads for clickbait content farms and slideshows, because the quality of the content was than lower than expected, or the page did not contain the content promised. Measuring feelings of regret and of being misled could be used as a metric for identifying ads that waste people’s time and attention, which could provide a basis for new legislation on online advertising, or could provide evidence for violations of existing FTC regulations against “misleading door openers” [70].

Targeting of "Bad Ads"

What is the role of ad targeting and delivery in the distribution of "bad" ads? For example, do certain demographic groups receive disproportionately many misleading health and supplement ads? Understanding whether the ad targeting and delivery infrastructure is being used to target vulnerable populations could contribute to ongoing discussions of regulations and algorithmic fairness and privacy in the advertising ecosystem [9, 165, 124, 125].

Automated Classification of "Bad Ads"

Our methods and data provide a basis for potential automated approaches to detecting "bad ads". Using the population label distribution learning approach [132], our dataset[6] from Survey 2 could be used to train a classifier that predicts user opinion distributions based on the image and/or text content of the ad. Such a classifier could be used for future web measurement studies, or user-facing tools, like extensions to block only bad ads, or browser features to visually flag bad ads.

3.5.4 Limitations

Our study only examines third-party, programmatic advertising common on news and content websites, and may not generalize to other types of online ads. Due to our crawling methodology, we did not cover ads on social media, video ads, and ads targeted at specific behaviors or locations. Additionally, our study is U.S. centric: we obtained ads using a U.S.-based crawler, from U.S. registered sites, we surveyed U.S.-based people, and make U.S.-based policy recommendations.

We did not show participants the full web page that the ads appeared on, which could affect their perception of the ads. For example, certain ads might be "acceptable" on an adult website

[6]Dataset is available in the Supplemental Materials of this paper, or at https://github.com/eric-zeng/chi-bad-ads-data

but not on a news website. (Regarding that specific example, we excluded adult sites from our dataset.) The screenshots we showed included a margin of 150 pixels of surrounding context on each side of the ad (see Figure A.1 in Appendix A.2).

Our participant samples skewed towards younger people and ad-blocker users. This reflects the overall userbase of Prolific[7] and the tendency for younger people to use ad blockers [135]. As a result, our data may somewhat overestimate the level of negativity towards ads. However, our regression analysis (Section 4.2.4) indicates that though ad blocker users are likely to rate ads more negatively, how users perceived the specifics of the ad were generally more important. Despite this bias, our results still are useful for understanding the phenomenon of "bad" ads, by systematizing qualitative reasons for disliking ads, and surfacing the concerns of users who actively choose to block ads.

Though we chose the sample of 30 ads in Survey 1 to cover a broad range of ad characteristics, it is nevertheless a small sample and our resulting taxonomy describing user perceptions of ads is unlikely to be comprehensive. We note that no methodology can cover all possible ads, since any crawl-based approach of obtaining display ads is inherently a snapshot of a subset of the ad campaigns running at that time. Though different ads could result in different user reactions, we believe that our approach of selecting a qualitatively diverse set of ads from our previous study's labeled dataset [226] surfaced many common reactions to ads from participants, and provides a useful basis for future work.

In Survey 2, it is possible that participants interpreted the taxonomy inconsistently, and assigned different meanings to the categories than us (or other participants). Therefore it is possible that differences in participants' understanding of the taxonomy decreased agreement in the opinion labels for some ads. We tried to mitigate potential confusion by making definitions of the categories easily accessible throughout the survey. Despite this limitation, our results still

[7] Prolific panel demographics: https://prolific.co/demographics/

provide useful insight into clusters of ads that participants had unambiguously negative or positive views of.

3.6 Conclusion

Though online advertisements are crucial to the modern web's economic model, they often elicit negative reactions from web users. Beyond disliking the presence of ads or their potential privacy implications in general, web users may be negatively impacted (financially, psychologically, or in terms of time and attention resources) by the content of specific ads. In this work, we studied people's reactions to a range of ads crawled from the web, investigating *why* people dislike different types of ads and characterizing specifically *which* properties of an ad's content contribute to these negative reactions. Based on both a qualitative and a large-scale quantitative survey, we find that large fractions of ads in our random sample elicit concrete negative reactions from participants, and that these negative reactions can be used to generate and characterize meaningful clusters of "bad" ads. Our findings, taxonomy, and labeled ad dataset provide a user-centric foundation for future policy and research aiming to curb problematic content in online ads and improve the overall quality of content that people encounter on the web.

Chapter 4

Clickbait Native Ads on News and Misinformation Websites

In recent years, journalists and researchers have raised concerns about problematic content appearing on news websites, such as clickbait, misinformation, scams, and malware. In particular, there was concern about the proliferation of native advertising, ads which imitate the look and feel of first party content; anecdotally, some of the lowest quality ads were found in native ads on news websites, in what are called “chumboxes”. This chapter presents a systematic measurement study of ad content on mainstream news sites and known misinformation sites. Using ads crawled from over 7000 websites, and a mixed of qualitative and quantitative analysis, this study provides evidence that there are a significant number of problematic ads on popular news and misinformation sites, primarily served through native ad networks.

This chapter originally appeared as the paper “Bad News: Clickbait and Deceptive Ads on News and Misinformation Websites” at the Workshop on Technology and Consumer Protection in 2020 [226].

4.1 Introduction

Online advertisements are an unavoidable fact of the modern web — they are embedded in and financially support the majority of content websites. Significant prior work in the computer security and privacy community has studied the ecosystem of online advertising, particularly in terms of its privacy implications (e.g., [176, 213, 128, 63, 124, 125, 153, 212, 67, 21]) or the use of ads to spread malware (e.g., [129, 173, 224, 220]). What has not been substantively considered in the security community, however, is the *visible, user-facing content* of these advertisements (except to the extent it relates to privacy, e.g., people finding highly personalized ad content or ad targeting explanations "creepy" [211, 64]).

Meanwhile, there is significant anecdotal evidence that the content of online advertisements can be deeply problematic [133, 113, 134, 206, 152, 207, 151] — consider the examples in Figure 4.1, a row of low-quality ads colloquially called a "chumbox". These concerns have been voiced particularly about *native advertising*, that is, ads that appear to be first-party content on the hosting website (such as inline search results or recommended articles) but are actually paid for by an advertiser. Concerns about native ads include the fact that they are deceptive: users are not reliably able to identify them as ads [219, 12, 101, 218, 72] and may click on them thinking that they are reading another story on a news site. Anecdotally, native ads also commonly use "clickbait" techniques or other "dark patterns", e.g., curiosity-provoking headlines or shocking imagery to attract attention and entice users to click. Further, these ads seem to often lead to low-quality content, misinformation, or even outright scams (e.g., cure-all supplements) and malware.

Despite these issues — or perhaps because of them — native ads are appealing to ad networks and hosting websites, as they have the potential to generate a significant amount of revenue. Prior work has shown, and native ad platforms themselves claim that they generate significantly higher clickthrough rates (0.2% vs. 0.05%) than traditional "display ads" [16, 189, 120]. As a result, online

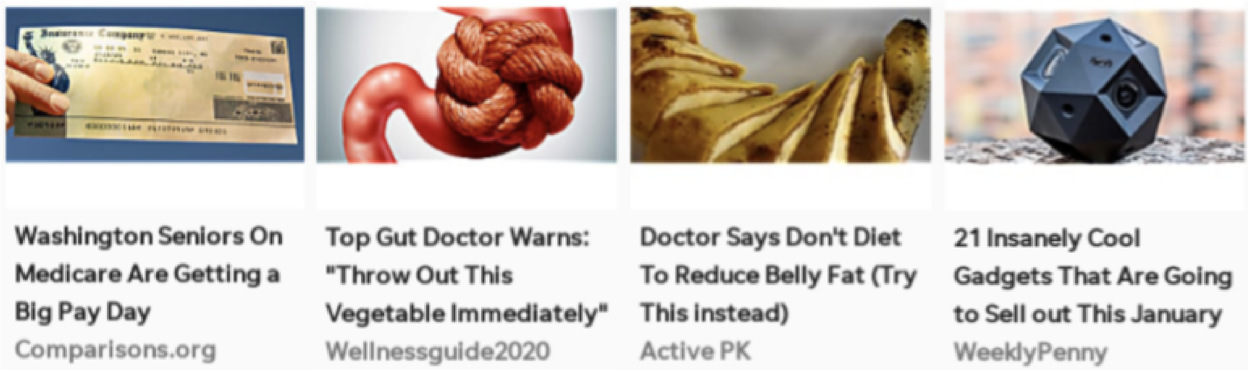


Figure 4.1: Example of native ads from a news website (aka a "chumbox"), showing four ads that use "clickbait" techniques to entice clicks (such as distasteful imagery, sensationalism, provoking curiosity, and urgency). Such ads often lead to low-quality sites, misinformation, or outright scams.

news and media publications, which have recently struggled to raise revenue [93, 82], frequently host native advertising on their properties.

Despite these many growing concerns about problematic content and dark patterns in online advertising, there has been limited systematic, scientific study of this phenomenon. We argue that these issues should be a concern of the computer security and privacy community, alongside the now well-understood privacy concerns regarding how those ads are targeted. First, these ads use misleading, deceptive, and in some cases illegal practices — impacting users financially, wasting their time and attention, and spreading scams, misinformation, and malware. At the same time, not all problematic ads are equally harmful: we must understand the spectrum of problematic ad content practices, their prevalence, and their impacts. Second, the locations where these ads appear can compound their harms: for example, on mainstream news and media websites, deceptive native ads may benefit from the trust that users have in the hosting website. Moreover, there is growing evidence that online ads are used to financially support news and media websites that spread *disinformation* (e.g., [80, 81, 51, 155, 113, 193, 48]). To fully understand the potentially harmful impacts, we must understand where these ads appear on the web and how they are targeted at individual consumers. The security and privacy community has the right tools (e.g., web crawlers, ad and tracker detectors), experience, and mindset to conduct a systematic study of this ecosystem.

In this work, we lay the foundation for such a systematic study of problematic ad content. We present the results from an initial measurement study of ad content on news, media, and known misinformation websites, and we surface hypotheses and directions for future work in the security and privacy community. Specifically, we focus on the following research questions:

1. How prevalent are different types of problematic ad content on the modern web?
2. How does the prevalence of problematic ad content *differ* across different types of ads (native vs. display), different ad platforms, and different types of websites (news/media vs. known misinformation)?

We performed a mixed-methods measurement study, using quantitative and qualitative techniques to explore ad content on popular news/media and known misinformation[1] sites in January 2020. Among other findings, we present empirical evidence that native ads use problematic techniques significantly more often than traditional display ads. We also find that both popular news sites and misinformation sites both run a significant amount of problematic ads, but that this phenomenon is not evenly distributed — that is, some sites choose to run problematic ads while others do not. Comparing ad platforms, we find that Taboola is responsible for serving the majority of problematic ads in our dataset, that Google also serves a significant number of problematic ads (though these represent a small percentage of their ads overall), and that there are certain (smaller) native ad platforms that appear more frequently on misinformation sites.

Our results and systematic measurement methodology lay a foundation for future work to further understand this ecosystem — e.g., studying the concrete impacts of problematic ads on

[1]Information that is deliberately false is often called "disinformation", while unintentionally incorrect information is called "misinformation" [107]. For simplicity, we default to the term *misinformation*, as we do not always know — and do not aim to clarify — the underlying intent of the creator.

users, or the ways that these ads may be targeted at more susceptible populations — in order to ultimately inform technical and/or regulatory solutions.

4.2 Methodology

We designed a rigorous methodology to allow us to study the prevalence of different types of problematic ad content. At a high level, our methodology involved crawling websites of interest, scraping the ads from these sites, and performing a systematic manual qualitative analysis of ad and landing page content for selected samples of ads.

4.2.1 Input Datasets

Mainstream News and Media Sites We collected a dataset of 6714 news and media sites from the Alexa Web Information Service API [13], which categorizes websites in the Alexa top 1 million by topic. We scraped all domains in the "News" category, and all domains in subcategories in other top-level categories that ended in "News and Media" or "Magazines and E-Zines". We excluded known misinformation sites.

Misinformation Sites We compiled a dataset of 1158 known misinformation websites (spreading political disinformation, hoaxes, conspiracies, and other misleading and false content) based on a combination of existing sources [66, 166, 117, 142, 157, 168, 98, 6]. These lists are surely incomplete, but allow us to study ads on *known* misinformation sites.

4.2.2 Crawling Infrastructure

We built a web crawler using Puppeteer [88], a browser automation and instrumentation library for the Chromium browser. Our crawler takes a URL as input, visits the URL, identifies each ad on the page using the EasyList filter list for Adblock Plus [57], a popular list of CSS selectors and domains used by many ad blockers to detect ads and trackers. The crawler screenshots each ad, stores its HTML content, and then clicks on each ad, and screenshots and scrapes the landing page.

Because ads that appear on a site's homepage may differ from those on article pages (e.g., some sites show native ads only at the bottom of articles), we crawled both the homepage and one article page for each site in our dataset. We found the URLs for articles using three heuristics: extracting the RSS feed from the site's HTML metadata, guessing the RSS feed by appending "/feed" or "/rss" to the domain, and randomly clicking links on the homepage and using Firefox's Readability library [146] (which transforms web articles into a simpler format) as an article-detection heuristic.

Clicking on ads raises ethical questions, since advertisers pay per click. We note that prior works have used similar methodologies [173, 220] and that even loading ads can lead to (smaller) costs (per impression). We believe that our measurements were small-scale compared to the overall business of the companies potentially impacted, and that fully studying this ad ecosystem, including landing pages, is crucial to understanding and reducing problematic content online.

Identifying Ad Platforms In addition to studying the content of the ads, we are also interested in the ad platforms responsible for delivering ads. The process for serving an individual ad is complex: often many companies are involved in taking an ad from an advertiser to a publisher, via supply side providers, ad exchanges, demand side providers, and ad servers. For the purposes of this study, we attempt to identify the third-party platform used directly by publishers to allow ads to run on their websites, such as Google Ad Manager. These platforms usually appear as a Javascript file or iframe embedded in the publisher's website (i.e., the host website).

To identify these publisher-side ad platforms, we use two complementary approaches. First, we detected well-known platforms like Google Ad Manager, Taboola, and Outbrain using CSS selectors that match HTML classes that we determined to be associated with the platform, based on manual inspection. For native ads that contain multiple ads in a single area, we also built selectors to split each individual ad into a separate record in our database. Second, for each DOM subtree we detected as an ad, we recorded each third-party resource in the subtree (iframes, anchors, images, and scripts), as well as any modifications made to the subtree via third-party Javascript elsewhere in the document. Post-crawl, we manually identified the publisher-side ad platform or other entity (e.g., ad exchange or third-party image host) behind the 100 most popular third-party resources — we did this by examining the resources and reading promotional materials or documentation at the domain of the resources. Lastly, we labeled the ad platforms we identified in both approaches as either native ad platforms or display ad platforms, based on how they describe their own product on their websites.

Studying Site-Based, Not Profile-Based, Targeting To enable comparisons between ads that appear on different types of sites, we wanted to maximize the chance that if we see a problematic ad, it was served based on the site we were visiting, not on the fact that our crawler has visited many misinformation sites in the past.

We thus visit each site using a separate browser instance in a new Docker container (i.e., containing no tracking cookies or other persistent browser state), to approximate a new user without a tracking profile. However, we must assume that the ad ecosystem may nevertheless successfully track our crawler, even across Docker instances, using fingerprinting, IP targeting, and other techniques [153, 63]. Embracing this reality, we thus "warm" the profile of our crawler by visiting all the sites in our input datasets twice, in random order, collecting data only on the second run (still using new containers for each site in each run). In other words, we ensure that the crawler's browsing profile looks consistent throughout the measurement to any ad networks able to fingerprint our crawler.

Crawls We created our dataset during January 15-19, 2020, successully crawling 6498 mainstream news sites (plus 5831 articles) and 1055 misinformation sites (plus 863 articles). Across these pages, we detected 81,870 ads, 55,045 of which were visible HTML elements.

4.2.3 Qualitative Analysis

Finally, we qualitatively analyzed and labeled the ads we observed on a subset of the websites we crawled. We generated a codebook to describe different types of problematic ads we observed in a preliminary analysis of the dataset, with each code describing a set of ads with similar advertisers, products, and advertising tactics. Our codes ranged from ads for things that could cause material harm, such as potentially misleading ads for supplements and investment pitches, to ads that people find irritating, such as ads for celebrity news content farms. The codebook was informed by prior academic research, regulations, and journalism on deceptive advertising, clickbait, malvertising, and advertising industry practices [133, 113, 134, 206, 152, 207, 151, 72]. The full codebook with definitions is included in Table 4.1.

Because we crawled 55,045 ads in total, we could only manually analyze a subsample of our dataset. For this preliminary work, we coded three different samples of websites, focusing on sites that users visit most: (1) 100 of the most popular news sites and their articles, (2) 100 of the most popular misinformation sites and their articles, and (3) 100 news sites (and articles) that have a similar popularity to the misinformation set. For the first and second samples, we discarded sites that our crawler could not reach and supplemented them with additional sites from the ranked lists until our sample size was 100 for each. The third sample allows us to control for the effect of site popularity on the types of ads that appear.

For each site in the samples, we coded each ad that appeared on the home page and article page, using a single code per ad. In total, we coded 2058, 1308, and 2048 ads from the top 100 news

Category	Definition
Content Farms	News sites and blogs that contain a high density of ads, often broken up into slideshows to artificially increase ads loaded. The content of the articles are typically about human interest news, celebrity news, or political news.
Insurance Advertorials	Ads appearing to be news articles about people saving money on car or health insurance, to persuade consumers to give personal information to insurance companies for quotes. The landing page does not clearly disclose that it is an ad.
Mortgage Advertorials	Ads for mortgage refinancing, promising large savings, sometimes citing changes to government policies. The goal is to collect consumers' personal information and send it to lenders for quotes. Unclear advertising disclosure.
Investment Pitches	Ads for investment opportunities that make sensationalist claims about their returns, "secret stock picks", or predictions of imminent economic turmoil. The advertisers are not affiliated with established brokerages or financial institutions.
Misleading Political Polls	Ads that appear to be political opinion polls, about politically polarizing candidates or issues, but require users to submit names and email addresses — likely for fundraising or advertising purposes.
Potentially Unwanted Software	Ads for software downloads that primarily consist of misleading UI elements, like large buttons labeled "Download" or "Watch Now", rather than advertising the name of the product or its functionality.
Product Advertorials	Ads for consumer products written in the style of a blog post or news article that do not obviously disclose that they were written by the advertiser, other than in the fine print in the header or footer of the page.
Sponsored Editorial	Articles hosted on news sites paid for and/or authored by an advertiser, to sell products or promote their views.
Sponsored Search	Ads for products or travel packages, but rather than linking to a specific business, links to search results for the product.
Supplements	Ads for supplements which claim about solve various chronic medical conditions, such as tinnitus, dark spots, weight loss, and toe nail fungus, but are not FDA approved.
Charities / PSAs	Charitable causes, public service announcements, class action lawsuit settlements, and other ads in the public interest.
Political Campaigns	Ads for political candidates or advocacy organizations, intended to spur people into taking action, including voting, signing petitions, donating, or other forms of political participation.
Products and Services	Straightforwards ads for various consumer products. No deception about the intent or identity of the ad is used.
Self Links	Ads that link to a page on the parent domain. Some native ad platforms will recommend both sponsored content and 1st party articles from the publisher.

Table 4.1: Problematic Ad Codebook. Labels used to describe ads in our qualitative analysis. The top section includes ad content we consider problematic, based on prior work, while the bottom section includes more neutral ad content.

sites, top 100 misinformation sites, and 100 similar popularity news sites respectively, for a total of 5414 ads.

4.2.4 Limitations

Our dataset contains a significant number ads of that could not be analyzed, because they were *not initialized* and were not rendered, or because of being *occluded* by other site content. Our crawler was unable to take screenshots of approximately one-third of ads detected using Easylist, because they were uninitialized and had zero height and/or width. Of the 5413 ads in our manually labeled sample, 1813 had no screenshot (33.5%), and 1182 (21.8%) were occluded or otherwise did not contain meaningful content. While the percentage of occluded and uninitialized ads were similar across our three samples of coded ads (40-47%), we observed that a substantially larger number of display ads, primarily from Google, were not rendered compared to native ads (56% vs. 31%).

To sanity check the quality of the data we collected via our crawler, we ran the ad detection algorithm described above in a standard desktop browser on 10 randomly sampled news and misinformation sites, and found 55.2% of ads were uninitialized, occluded, or otherwise false positives, compared to 58.3% on the same sites in our crawled dataset, suggesting that our crawled data is similar to what users actually see.

We suggest several reasons why some ads were not loaded or visible: (1) the elements were false positives in the ad blocker's detection algorithm, (2) the Docker environment and virtual frame buffer interfered with the browser's rendering, (3) content on the website, such as sign-up interstitials or cookie banners, occluded the ad content, and (4) the ad platform chose not to fill the ad space, e.g., due to detecting our visits as anomalous, low demand for ads, or high latency during real-time bidding. In drawing our conclusions, we assume that the distribution of problematic content among the ads that did not load because of the crawling environment is similar to that among the ads that did. Future work must validate this assumption and address this measurement challenge.

Additionally, our method for identifying ad platforms was not comprehensive (we did not identify ad platforms for 20.7% of the ads crawled), nor does it perfectly describe the entity "responsible" for working with problematic advertisers. For example, sites might configure Google Ad Manager to allow ads from a third-party ad exchange, where many third-party supply-side providers may bid on the site's ad inventory. Nevertheless, we chose to investigate the ad platforms used directly by publishers, as these platforms often have content policies in place against malicious and harmful content [203, 87, 86].

4.3 Results

4.3.1 Which ad platforms show problematic ads?

We first investigate whether native ad platforms are the primary culprit for problematic content in ads on news and misinformation sites. Table 4.2 shows the count of each ad content code across *all* of our samples, comparing their prevalence across native and traditional display ad platforms.

Based on the subtotals for all native ad platforms and display ad platforms, we highlight several high-level conclusions. First, a significant fraction of all coded ads contain some kind of problematic content: of the 2419 ads we coded, 1078 of (44.6%) them were labeled as problematic. Second, native ads are indeed primarily responsible for these issues: 87% of native ads (that loaded during the crawl) were labeled as problematic, compared to 20% of display ads. Third, however, display ads do also include non-trivial numbers of problematic ads (particularly for supplements) — thus, conversations about ad content should not focus exclusively on native ads.

Next, we analyze the prevalence of problematic ads on specific ad platforms, from two perspec-

		Display Ad Platforms					Native Ad Platforms						
Code	Total	Amazon	Concert	Google	TownNews	Subtotal	Outbrain	PowerInbox	RevContent	Taboola	Zergnet	Subtotal	Unknown
Content Farms	283	0	0	13	0	13	0	0	1	178	87	266	4
Insurance Advertorials	96	0	0	21	0	21	0	0	15	59	0	74	1
Investment Pitches	43	0	0	10	0	10	0	2	6	24	0	32	1
Misleading Political Poll	14	0	0	10	0	10	0	0	0		0	0	4
Mortgage Advertorials	29	0	0	4	0	4	0	0	4	21	0	25	0
Potentially Unwanted Software	8	0	0	7	0	7	0	0	0	1	0	1	0
Product Advertorials	103	0	0	8	0	8	0	0	2	92	0	94	1
Sponsored Editorials	50	0	0	29	0	29	0	0	0	9	0	9	12
Sponsored Search	196	0	0	17	0	17	0	0	0	177	0	177	2
Supplements	256	0	0	106	0	106	0	2	38	98	0	138	12
Problematic Ads Subtotal	1078	0	0	225	0	225	0	4	66	659	87	816	37
Charities and PSAs	17	0	0	17	0	17	0	0	0	0	0	0	0
Political Campaign	28	0	0	12	0	12	0	0	0	0	0	0	16
Products and Services	1214	0	8	1050	1	1059	1	2	0	93	0	96	59
Self-Link	82	0	8	28	1	37	5	0	4	17	0	26	19
Benign Ads Subtotal	1341	0	16	1107	2	1125	6	2	4	110	0	122	94
Total Coded	2419	0	16	1332	2	1350	6	6	70	769	87	938	137
Occluded/Uninitialized Ads	2995	1	16	1666	7	1690	292	0	16	72	45	425	876
Grand Total	5414	1	32	2998	9	3040	298	6	86	841	132	1363	1013

Table 4.2: Raw counts of coded ads across ad platforms, from the samples of misinformation, top news, and similar popularity news sites. Subtotals for native ad platforms and display ad platforms are listed inline. The percentage of a particular ad code contributed by a platform can be calculated by dividing the cell by the row-wise total (e.g., 55% of ads for investment pitches were served by Taboola, or 24/43). The percentage of ads within a platform of a specific code can be calculated by dividing the cell by the column-wise total (e.g., 8% of Google Ads were labeled as "Supplements", or 106/1332).

tives. First, which ad platforms serve the largest absolute number of problematic ads, contributing most to what users see? Second, which ad providers serve disproportionately many problematic ads, as a fraction of all ads they serve?

We observe that Taboola served the largest number of problematic ads in our samples (61.1% of all problematic ads), and that proportionally, most of the ads served by Taboola were problematic (85.7%). Taboola also served a large diversity of problematic ads: we saw examples for all categories in our codebook except for misleading political polls. By contrast, other native ad platforms with significant numbers of ads in our samples served a more concentrated selection of problematic ad types: 100% of the ads on the Zergnet network were for content farm-style articles and slideshows, and 57.6% of all RevContent ads advertised some sort of supplement.

Google was the most popular ad platform in our sample, making up nearly the entirety of the display ads that we coded. While most ads served through Google were benign ads for various products and services, 16.9% of Google-served ads were problematic, accounting for 20.8% of problematic ads in our samples. While Google's platform does not serve as many problematic ads proportionally, due to its large volume of ads in general, we note that the number of problematic ads it serves is substantial, second only to Taboola.

These results suggest that while the advertising ecosystem is large and complex, a large proportion of problematic content flows through large platforms popular with publishers, like Google Ads and Taboola. Efforts to eliminate problematic ads could start by focusing on regulating or improving ad content moderation on these platforms.

4.3.2 Are problematic ads more frequent on misinformation websites?

We next consider whether problematic ads appear disproportionately more often on misinformation sites, compared to legitimate news/media sites. We initially hypothesized that we would see

	Top 100 News				Top 100 Misinfo				100 Popularity Adjusted News			
	Homepage		Article		Homepage		Article		Homepage		Article	
Code	n	%	n	%	n	%	n	%	n	%	n	%
Content Farms	46	12.5%	66	12.1%	3	1.5%	62	15.4%	37	9.4%	69	13.7%
Insurance Advertorials	18	4.9%	21	3.9%	6	2.9%	20	5.0%	5	1.3%	26	5.2%
Investment Pitches	9	2.4%	10	1.8%	5	2.4%	10	2.5%	2	0.5%	7	1.4%
Mortgage Advertorials	0	0.0%	13	2.4%	0	0.0%	5	1.2%	1	0.3%	10	2.0%
Misleading Political Polls	1	0.3%	1	0.2%	5	2.4%	7	1.7%	0	0.0%	0	0.0%
Potentially Unwanted Software	0	0.0%	1	0.2%	3	1.5%	2	0.5%	0	0.0%	2	0.4%
Product Advertorials	12	3.3%	33	6.1%	0	0.0%	16	4.0%	8	2.0%	34	6.7%
Sponsored Editorials	14	3.8%	11	2.0%	1	0.5%	0	0.0%	14	3.6%	10	2.0%
Sponsored Search	39	10.6%	56	10.3%	6	2.9%	41	10.2%	20	5.1%	34	6.7%
Supplements	22	6.0%	72	13.2%	35	17.1%	73	18.2%	11	2.8%	43	8.5%
Problematic Ads Subtotal	161	43.6%	284	52.1%	64	31.2%	236	58.7%	98	24.9%	235	46.6%
Charities and PSAs	7	2.0%	0	0.0%	3	1.5%	1	0.3%	4	1.1%	2	0.4%
Political Campaigns	0	0.0%	3	0.6%	10	5.0%	13	3.3%	1	0.3%	1	0.2%
Products and Services	179	51.6%	240	45.5%	124	61.7%	143	36.4%	273	69.1%	255	51.7%
Self Links	22	6.0%	18	3.3%	4	2.0%	9	2.2%	18	4.6%	11	2.2%
Benign Ads Subtotal	208	56.4%	261	47.9%	141	68.8%	166	41.3%	296	75.1%	269	53.4%
Total # of Ads Coded	369		545		205		402		394		504	
Occluded/Uninitialized Ads	466		678		255		446		594		556	

Table 4.3: Counts of ads we labeled across our samples of news and misinformation sites. Percentages are computed columnwise (with the total number of coded ads as the denominator). We do not see evidence for substantial differences in the prevalence of problematic ad content across these samples.

Avg. # of Ads/Page	Mainstream News Homepage	Mainstream News Article	Misinformation Homepage	Misinformation Article
All Ads	5.80	6.37	2.37	5.16
Display Ads	5.57	5.22	1.78	2.79
Native Ads	0.23	1.15	0.59	2.37
All Coded Ads	8.27	12.35	4.90	8.88
Coded Display Ads	6.89	9.05	4.53	6.58
Codes Native Ads	1.38	3.30	0.37	1.38

Table 4.4: Average number of ads per page. Top: Ads on *all* crawled pages. Bottom: Manually labeled ads. While mainstream news sites tend to have more ads on the homepage, misinformation sites run more native ads. (Note that the native ad fraction is an underestimate, since uncommon, unknown ad providers are considered display ads here.)

such a difference, because news sites might choose to include higher quality ads, and/or because the ad targeting ecosystem might be more likely to serve problematic ads to misinformation sites. Table 4.3 investigates this relationship, breaking down labeled ads between the three samples of websites, considering both homepages and article pages.

We draw several conclusions. First, although we see some differences, the numbers are small — overall, we do *not* see evidence for significant differences between the types of sites. In other words, it does not appear that visitors to popular misinformation sites are significantly more likely to encounter problematic ads. Second, in all samples, problematic ads appear more on articles than homepages. This may be because some sites "hide" problematic ads beyond the homepage.

Without automated classification of problematic ads, we cannot consider the prevalence of these issues below the top-ranked websites that we studied manually. However, recall that our measurement infrastructure automatically identifies a set of popular ad providers associated with ads. Based on this metadata, which is available even for ads that did not load properly, we can estimate the proportion of native ads in our *whole* crawled dataset, i.e., thousands of sites.

Table 4.4 shows the average number of native and display ads per page, for sites in our full dataset. On average, we see that news sites run more ads than misinformation sites on their

	Top 100 News		Top 100 Misinfo		100 Popularity Adjusted News	
	Homepage	Article	Homepage	Article	Homepage	Article
Some Problematic Ads	43	44	24	42	29	40
Ads, None Problematic	54	47	47	34	61	50
No Ads	3	9	29	24	10	10

Table 4.5: Counts (or percents) of sites in our three samples that include no ads, only "clean" ads, and at least one problematic ad. Problematic ads are clustered: a large fraction of sites in each sample include only "clean" ads. We caution that these are underestimates, due to ads that were not loaded.

homepages, but both run similar numbers of ads on their articles. However, news sites appear to use a significantly greater fraction of display ads compared to misinformation sites, *when considering the full dataset.* This result suggests that as we consider lower-ranked news and misinformation sites, the gap between the quality of ads on those sites might be larger than what we observe for the popular subset.

More broadly, Table 4.4 also provides large-scale evidence that misinformation sites heavily leverage the targeted ad ecosystem for monetization — supporting recent reports [81, 48] and underscoring the need for advertisers and ad platforms to consider their role in supporting (or combating) these actors.

4.3.3 Are problematic ads evenly distributed across sites?

The previous section showed that problematic ad content appears roughly equally often, on average, on different samples of sites. However, this result does not imply that all sites include equal numbers or fractions of problematic ads.

Table 4.5 divides sites into three categories: those that contain problematic ads, those that do not, and those that do not have any ads at all. What we find is that sites do indeed differ on this point: the problematic ads we see are clustered in 32%–57% of the ad-supported sites in each sample, though we do not see evidence for large differences between the samples. In other words, certain sites use ad platforms or preferences that allow problematic ads to run, but others run only or primarily “clean” ads.

Due to the challenges with many ads not loading discussed above, and because ads that appear are not consistent across page loads, the number of sites that run problematic ads may be an *underestimate*. Due to this concern, we manually investigated a sample of “clean” sites, which indeed appeared to only include display ads for benign products and services.

Also anecdotally, we observed that sites *with* problematic ads are also not created equal: some sites include a mix of “clean” display ads and one native ad, while others contain 10+ problematic native ads.

4.3.4 Do misinformation sites use a different set of ad providers?

Lastly, we investigate whether misinformation sites use different ad platforms than news sites. Are there specific ad platforms that are more popular among misinformation site operators? We might expect to see such difference because, for example, these site operators tolerate lower quality advertisements, or because certain ad platforms are willing to work with misinformation sites but not others.

Table 4.6 shows the distribution of ad platforms used by misinformation sites compared to news sites across our entire dataset. We see that Google Ads are common in both populations, comprising 52.8% of ads on misinformation homepages, and 68.3% of ads on news site homepages. Taboola is the second most common ad platform and most common native ad platform, especially on article pages, making up 10.2% and 15.8% of ads on misinformation and news article pages. However, we see that certain native ad providers, such as content.ad, RevContent, and Zergnet, are much more popular among misinformation sites. We note that these ad platforms also run high proportions of problematic ads (see Table 4.2).

		Misinformation				Mainstream News				
		Home Page		Article		Home Page		Article		
Platform	Ad Format	n	%	n	%	n	%	n	%	Total
Ad Butler	Display	1	0.0%	1	0.0%	109	0.3%	102	0.3%	213
Amazon	Display	5	0.2%	10	0.2%	23	0.1%	50	0.1%	88
AuctionNudge	Display	0	0.0%	0	0.0%	2	0.0%	1	0.0%	3
Concert	Display	0	0.0%	0	0.0%	77	0.2%	72	0.2%	149
Google	Display	1322	52.8%	1572	35.3%	25753	68.3%	20947	56.3%	49594
TownNews	Display	0	0.0%	0	0.0%	452	1.2%	321	0.9%	773
Connatix	Interactive	1	0.0%	0	0.0%	20	0.1%	35	0.1%	56
Insticator	Interactive	3	0.1%	4	0.1%	168	0.4%	169	0.5%	344
AdBlade	Native	13	0.5%	47	1.1%	0	0.0%	18	0.0%	78
content.ad	Native	163	6.5%	495	11.1%	3	0.0%	165	0.4%	826
FeedNetwork	Native	8	0.3%	0	0.0%	0	0.0%	0	0.0%	8
MGID	Native	58	2.3%	234	5.3%	0	0.0%	75	0.2%	367
Outbrain	Native	15	0.6%	148	3.3%	694	1.8%	1470	4.0%	2327
PowerInbox	Native	20	0.8%	43	1.0%	0	0.0%	1	0.0%	64
RevContent	Native	197	7.9%	567	12.7%	55	0.1%	418	1.1%	1237
Taboola	Native	111	4.4%	452	10.2%	1336	3.5%	5866	15.8%	7765
Zergnet	Native	69	2.8%	277	6.2%	89	0.2%	558	1.5%	993
Unknown		517	20.7%	601	13.5%	8928	23.7%	6939	18.6%	16985
Total		2503		4451		37709		37207		81870

Table 4.6: Counts of ads from each ad platform, across *all* crawled pages. Percentages are computed columnwise (with the total number of ads as the denominator). Google Ads and Taboola are similarly popular across both populations, but many smaller native ad platforms are present on misinformation sites but rare on news sites, such as content.ad and RevContent.

Our results suggest that misinformation sites appear largely to be able to work with the same types of ad platforms as mainstream news sites — i.e., we do not see much evidence that they have been systematically "deplatformed" by any major providers. These results are consistent with prior work from GDI showing that misinformation sites generate revenue from roughly the same ad exchanges as mainstream news sites [81]. Our data also suggests that with the exception of Taboola, mainstream news sites tend to avoid using many native ad platforms that misinformation sites use, perhaps due to the low quality ad content served by those platforms.

4.4 Discussion

We argue that problematic content of online — particularly native — ads should be a subject of systematic study by the computer security and privacy community. In this paper, we provide initial results, which raise many additional questions and lay a foundation for future work. For example:

Larger-Scale Systematic Measurement Our work considers a small set of popular sites. While these are (by definition) the sites users are most likely to visit, our results raise the question of how things look in the longer tail. For example, perhaps lower-ranked sites tolerate more problematic ads, or perhaps (as suggested by Table 4.4) lower-ranked misinformation sites are worse than similarly-ranked news sites. One key challenge to a larger-scale analysis is the need for *automated classification* of problematic ad content; future work might build on our labels in the Appendix and prior work on clickbait or adversarial ad detection (e.g., [39, 167, 187]). The methodology we present can also lay a foundation for future measurements, but we highlight several additional *measurement challenges* that must be addressed: (1) classifying problematic ads often requires considering both the ad itself and the landing page, but automatically clicking on ads should be thought through carefully given that it impacts the ad ecosystem, and (2) many ads were not loaded by our crawler (perhaps due to anomaly detection by ad networks due to our clicking). Prior work on tracking detection either did not have to contend with the challenge of ads not loading due to anomaly detection (because it did not require clicking on ads) or did not notice the limitation (because it did not inspect ads visually).

Role of Ad Targeting The types of ads that appear on a website result from a combination of the ad platform's policies and partners, options chosen by site's owner, and the ad platform's targeting of the end user. We described and used a methodology that isolated ad targeting based on hosting site, not the user. While we studied news and misinformation sites, other types of sites warrant investigation (e.g., sites targeted at children). Additionally, we hypothesize that there is an interplay between problematic ad content and the fine-grained (and privacy-invasive) *user targeting* enabled by today's online ad ecosystem. Who is being targeted with different types of problematic ads? Are there some potentially vulnerable populations (e.g., seniors, or people who frequently visit known misinformation sites) being disproportionately exploited?

Understanding and Differentiating Impacts on Users Beyond studying the ad ecosystem technically via web measurement, it is crucial to also study the actual *human impacts* of these problematic ad practices. Not all of the practices we discuss are equally harmful, and to combat them, particularly through policy and regulation, we must understand their relative harms. For example, false advertising and scams are not only problematic but illegal under existing regulations. But is "clickbait" merely annoying, or actively harmful? Future work should conduct user studies to help clarify these harms. For example, how do people actually perceive and interact with these ads? How much time do people spend on low-quality sites reached via ads, and how do they value that time compared to the time they spend elsewhere on the web? How well do the various "dark patterns" we see work in practice, and on which types of users — are some manipulative techniques disproportionately successful, and are some users particularly vulnerable? While prior work in the marketing literature has considered related issues (e.g., [34, 109, 45, 49]), these works typically focus on deception in legitimate product ads and/or do not include large-scale measurement studies.

Defenses: Policies, Regulations, Tools Many ad platforms, including native ad platforms, have explicit policies against problematic ad content (e.g., [87, 203]). In our analysis, however, we saw many examples of ads that either violate these policies or only technically meet them. Understanding the root cause of this discrepancy requires further investigation: perhaps some violating ads are difficult to detect, some policies are inconsistently enforced, or the policies as written are insufficient to prevent the types of ads we identified as problematic. At the same time, some types of problematic ads may be annoying, but are not sufficiently problematic to ban outright (especially by U.S. regulatory agencies, which are constrained by the First Amendment). Combining systematic web measurements with user studies (proposed above) to understand the concrete impacts on end users may provide clarity on where to draw the line. Beyond policy, technical defenses may play an immediate role in helping end users. For example, future work might explore designing and evaluating a browser extension that detects and warns users of problematic content in ads, or that blocks only problematic ads.

4.5 Conclusion

The potential harms of online ads have become a core interest of the computer security and privacy community in the last decade. In this work, we expand that focus to consider the visible content of advertisements. We aim for our work to lay the foundation to rich future investigations into this aspect of the online ad ecosystem, ultimately reducing the spread of misinformation and other low-quality content online.

Chapter 5

Problematic Political Advertising on News and Media Websites During the 2020 U.S. Elections

Online advertising can be used to mislead, deceive, and manipulate Internet users, and political advertising is no exception. This chapter presents a measurement study of online advertising around the 2020 United States elections, with a focus on identifying dark patterns and other potentially problematic content in political advertising. The study collected ad content from 745 news and media websites from six geographic locations in the U.S. from September 2020 to January 2021, collecting 1.4 million ads. Through a systematic qualitative analysis of political content in these ads, as well as a quantitative analysis of the distribution of political ads on different types of websites, the results reveal the widespread use of problematic tactics in political ads, such as bait-and-switch ads formatted as opinion polls to entice users to click, the use of political controversy by content farms for clickbait, and the more frequent occurrence of political ads on highly partisan news websites. The chapter ends with some policy recommendations for online political advertising, including greater scrutiny of non-official political ads and comprehensive standards across advertising platforms.

This chapter originally appeared as the paper “Polls, Clickbait, and Commemorative $2 Bills: Problematic Political Advertising on News and Media Websites Around the 2020 U.S. Elections” at the Internet Measurement Conference in 2021 [228].

5.1 Introduction

The 2020 United States general elections were one of the most important and contentious elections in recent history. Issues facing the U.S. included the COVID-19 pandemic and ensuing economic crisis, controversy surrounding President Donald Trump’s first term, and renewed movement for racial justice following the murder of George Floyd and other police violence. During this election season, online political advertising was more prominent than ever: campaigns turned to online ads as the pandemic reduced in-person events and canvassing [208], and spent record sums advertising on Google and Facebook [175]. The misuse of online ads in non-political contexts is a well-known problem, ranging from distasteful clickbait ads to outright scams and malware [226, 227, 129, 224, 151]. In this paper, we investigate misleading and manipulative tactics in online political advertising, for purposes such as collecting email addresses and driving traffic to political content websites.

We take a broad view of what constitutes a “political” ad in our work, considering any ad with political content, whether or not the ad was placed by an official political campaign committee. In our investigation, we ask: Who ran political ads during this period? What was the content of these ads, and do they use problematic techniques? Did the number of political ads on different types of websites differ?

To answer these questions, we conducted measurements of online advertising before, during, and after the Nov. 3rd elections. We collected a daily crawler-based sample of ads from 745 online news and media websites from September 2020 to January 2021, providing insight into the ads people saw while reading news during this period. We continued collecting data through several post-election developments: contested vote counting in multiple states, the Georgia U.S. Senate runoff election on January 5, and attack on the U.S. Capitol on January 6. Our crawlers collected data from six locations with varying political contestation: Atlanta, GA; Miami, FL; Raleigh, NC; Phoenix, AZ; Salt Lake City, UT; and Seattle, WA.

Using a combination of qualitative and quantitative techniques, we analyze the political ads in our dataset, including identifying examples of misleading and manipulative techniques, the distribution of political ads across websites of different political biases, and political affiliations and organization types of the advertisers.

Scope Our crawler-based dataset provides a complementary perspective to the political ad archives from Google and Facebook. Though our dataset is not as complete as the political ad archives, and partially overlaps Google's, our dataset encompasses *all* ads on the pages we crawled — including non-political ads, political-themed ads were not officially classified as political and thus do not appear in Google's archive, and ads served via ad networks outside of Google Ads. Additionally, we capture the URL of the website that each ad appeared on, allowing us to measure contextual targeting of political ads on news and media websites.

Contributions First, we characterize the quantity and content of online advertising longitudinally during the 2020 U.S. Presidential Election and shortly thereafter, and at scale.

- We observe differences in the number of political ads in different geographical locations.
- We observe shifts in the quantity of political ads through the election, and the effects of political ad bans.

- We characterize the topics of all online advertisements that we collected during this time period.

Through our qualitative analysis, we observed several problematic types of online political advertising, such as:

- The use of misleading and manipulative patterns in political ads. For example, ads that purport to be political polls, but use inflammatory framing, and appear to be used for gathering email addresses.
- Political topics in clickbait and native advertising. These ads imitate the look of links to news articles, but link to external sites. Headlines often imply controversy about candidates, and may fuel disinformation.

We also find that problematic political ads are more common on partisan and low-quality news sites.

- More partisan websites have more political ads, on both ends of the political spectrum.
- Problematic categories of ads, such as political products and polls, appear more frequently on right-leaning sites.

We discuss the potential harms from the problematic political ads we observed, and we make recommendations for platform policies, government regulation, and future research. We also release our full dataset of ads and metadata.

5.2 Background and Related Work

5.2.1 The 2020-21 U.S. Elections

Between September 2020 and January 2021, the U.S. held a presidential election, congressional elections, and numerous state and local elections. In the presidential election, Joe Biden, a

Democrat, and his running mate, Kamala Harris, ran against Donald Trump, the incumbent Republican president, and his running mate, Mike Pence [20].

Election day was November 3, 2020, but the results of the election were significantly delayed due to the COVID-19 pandemic as states continued to receive mail-in votes and count ballots in subsequent days [42]. During this time, Trump and his campaign maintained that there was widespread voter fraud [195]. Most major news outlets declared the results — that Biden had obtained enough electoral votes to defeat Trump — on November 7 [127].

Sparked by a speech from Donald Trump on January 6, 2021 in which he continued to falsely claim that he had won the election, thousands of his supporters marched to the U.S. Capitol complex, where Congress had assembled to certify the electoral result [205]. The storming of the Capitol resulted in over 140 injuries [185] and 5 deaths [65]. The certification was completed the next day and President Biden's inauguration was held on January 20, 2021.

On November 3, elections were also held for seats in the Senate and House of Representatives. In state and local politics, elections were held for 13 governorships in 11 states and 2 territories, as well as for state legislative chambers, attorneys generals, state supreme court seats, and various referendums and ballot measures. In the state of Georgia, no Senate candidates received a majority of the vote during the first round, leading to a run-off election on January 5, 2021.

5.2.2 Online Political Ads Policy During the 2020-21 U.S. Elections

Before the election, tech companies faced mounting pressure to address concerns about political advertising spreading misinformation and causing other harms. Some companies had already banned political ads (Pinterest in 2018 [91], Twitter in 2019 [53]), at least in part due to revelations that Russian organizations had purchased political ads during the 2016 presidential election [116]. Google and Facebook allowed political ads in 2020, (and had resisted regulation around political ad disclaimers previously [96]), but implemented several short-term bans. Our dataset of display ads was likely impacted by Google's bans from Nov. 4 through Dec. 10 [75, 186], and again after the storming of the Capitol between Jan. 14 and Feb. 24 [76]. The day after the election, Google began an ad ban on ads "referencing candidates, the election, or its outcome" to reduce misinformation, which it maintained until December 10 [75, 186]. Google then implemented another ban between January 14 and February 24, following the storming of the Capitol, to mitigate confusion around "sensitive political events."

Still, political ads around the 2020-21 elections set new records for ad spending, with overall spending in the billions. On Facebook and Google alone, the Trump campaign spent $276 million and the Biden campaign spent $213 million [175].

5.2.3 Online Political and Problematic Ads

Prior work studies the online ad ecosystem from various perspectives. In the computer security and privacy community, researchers have often studied the privacy implications of online ads and the tracking enabling them (e.g., [176, 153, 125, 212, 21, 183]). In this work, we focus on the content of ads and contextual targeting that may cause different ads to appear on different types of sites, rather than on the underlying privacy-invasive mechanisms.

Recent work in computer science identifies types of problematic content in ads (e.g., clickbait, distasteful ads, misleading content, manipulative techniques) [226, 227], and types explicitly malicious ads (e.g., spreading malware) [129, 224, 220, 151, 173]. Online ads play a role in spreading mis/disinformation (e.g., during the 2016 and 2018 U.S. elections) [46, 192, 193, 61] as well as in monetizing mis/disinforma-tion websites [81, 155, 113, 48]. Other work has shown that ads (e.g., on Facebook) may be targeted in discriminatory ways [9, 118]. Studies of misleading and manipulative patterns (often called "dark patterns") beyond ads also inform our work (e.g., [139, 149]), particularly a recent study of such patterns in political campaign emails [140].

Significant work in other fields (e.g., political science and marketing) also studies political ads. Kim et al. identified political ads on Facebook purchased by "suspicious" groups, including Russian groups known for spreading disinformation [116]. Stromer-Galley et al. [199] studied U.S. political ads on Facebook in 2016 and 2020, while Ballard et al. [19] characterized political campaign web display ads during the 2012 U.S. elections. Other work considered deceptive political advertising, (not necessarily online) including deceptively formatted "native" ads (e.g., [144, 56]). Van Steenburg provides a systematic literature review of political advertising research and proposes a research agenda, identifying the study of the impact of technology (i.e., the internet) as one key theme and area for future work (but does not discuss the manipulative patterns or non-official political ads that we see in our dataset) [198].

Our work considers ads appearing on websites rather than social media, and we capture all ads (not only those marked as political ads). Prior work has found that Facebook's ad archives are incomplete and use a limited definition of "political" [61, 60, 194]. Indeed, we found many ads that contained political themes but were not placed by an official campaign.

5.3 Methodology

In this section, we describe our methodology for measuring ads throughout the 2020 U.S. elections. In summary, we selected a group of popular mainstream and alternative news websites and scraped ads from these sites using crawlers in different locations. We collected 1.4 million ads in total from September 2020 to January 2021. We analyzed the content of our ads dataset using a combination of natural language processing, to automate tasks like identifying which ads were political, and manual qualitative analysis techniques, to provide greater context such as the party affiliation of the advertiser. See Figure 5.1 for a summary of our analysis pipeline.

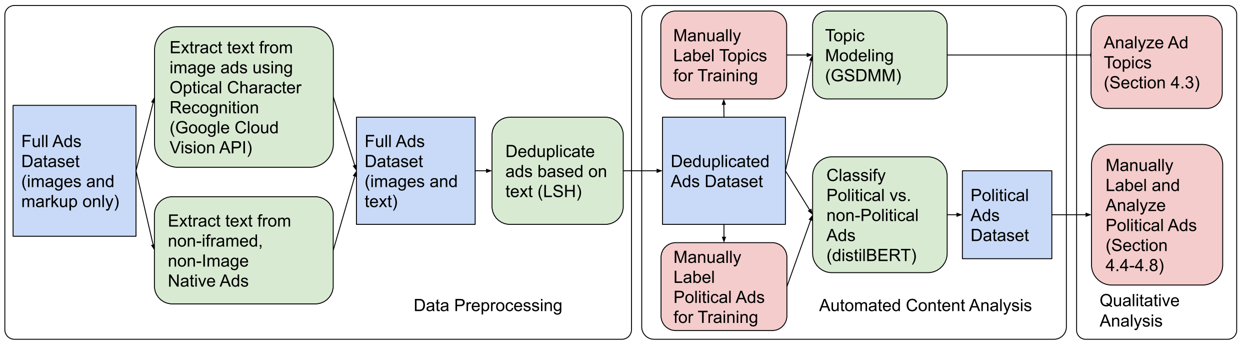


Figure 5.1: Overview of our analysis methodology. We used NLP techniques to preprocess and organize our dataset, and then conducted manual content analyses to explore political ads in greater detail, and to validate automated outputs. Blue boxes represent data, green boxes represent automated processes, and red boxes represent manual and qualitative analyses.

5.3.1 Ad Crawling

Seed Websites

Site Bias	# Sites	Examples
Mainstream News and Media Websites		
Left	63	jezebel.com, salon.com
Lean Left	57	miamiherald.com, theatlantic.com
Center	46	npr.org, realclearpolitics.com
Lean Right	18	foxnews.com, nypost.com
Right	44	dailysurge.com, thefederalist.com
Uncategorized	376	adweek.com, nbc.com
News Websites Labeled as Misinformation		
Left	13	alternet.org, dailykos.com
Lean Left	6	greenpeace.org, iflscience.com
Center	1	rferl.org
Lean right	11	rt.com, newsmax.com
Right	60	breitbart.com, infowars.com
Uncategorized	50	globalresearch.ca, vaxxter.com

Table 5.1: Summary of seed websites.

To collect ads, we crawled news and media websites that spanned the political spectrum and information ecosystem. We identified 6,144 mainstream news websites in the Tranco Top 1 million [123], using categories provided by the Alexa Web Information Service [13]. These mainstream sites included national newspapers, local newspapers, TV stations, and online digital media. We also compiled a list of 1,344 websites which we refer to as "misinformation websites". Websites in this list were identified as "fake news", alternative news, mis/disinformation, highly partisan, propaganda, or conspiracy websites by fact checkers (Politifact [196], Snopes [117], Media Bias/Fact Check [142], and others [157, 98, 66]).

To ensure that our crawlers could complete the crawl list in one day, we truncated the list to 745 sites by picking all sites with a ranking higher than 5,000 (411 sites), and then sampling from the remaining tail (334 sites) by choosing 1 site per bucket of 10,000 site rank, to ensure that lower ranked sites were represented. In Table 5.1, we show the number of sites in our crawl list by misinformation label and political bias. The political bias of websites were aggregated from Media Bias/Fact Check [142] and AllSides [11].

Crawler Implementation

We built a web crawler to scrape ads based on Puppeteer [88], a Chromium-based browser automation library. Each crawler node crawls the seed list once per day, crawling 6 domains in parallel in random order. For each seed domain, the crawler loads the root page and detects ads using CSS selectors from EasyList [57], a filter list used by ad blockers. Elements smaller than 10 pixels in width or height (like tracking pixels) were ignored. The crawler scrolls to each ad, takes a screenshot, and collects the HTML content. Then, the crawler clicks the ad, and collects the URL and content of the landing page. Because ads may differ on site homepage vs. subpages, for each seed domain, the crawler also visits and collects ads from an article on the site.

To minimize behavioral ad targeting, we crawled each seed domain using a clean browser profile (similar to prior work [226]). For each domain we visited, we ran separate browser instances inside a new Docker container, so that no tracking cookies or other state persisted across domains (though fingerprinting may be possible).

Crawler Nodes and Locations

We crawled ads using 4 nodes from geographical locations where we predicted the political landscape could result in different ads.

- *Sep. 25, 2020 – Nov. 12, 2020*: We first crawled from two cities in states predicted to be contested (Miami, FL; Raleigh, NC) and two uncompetitive (Seattle, WA; Salt Lake City, UT).
- *Nov. 13, 2020 – Dec. 8, 2020*: Due to contested election results, we switched two crawlers to Phoenix, AZ and Atlanta, GA. The other two crawlers alternated between the 4 previous locations (Seattle, Salt Lake City, Miami, Raleigh).
- *Dec. 9, 2020 – Jan. 19, 2021*: After the presidential election was resolved, we crawled from Atlanta, GA and Seattle, WA to observe the Georgia special election. Due to the Capitol insurrection, we continued crawling for 2 weeks.

To simulate crawling from these locations, we tunneled our traffic through the Mullvad VPN service. Mullvad's VPN servers ran on rented servers in local data centers (100TB, Tzulo, and M247). We verified that the VPN servers were located in the advertised locations using commercial IP geolocation services.

In sum, we ran 312 daily crawls, on 4 machines, using Chromium 88.0.4298.0, on a Debian 9 Docker image. The hardware was: Intel Core i7-4790 3.6GHz 32GB RAM, Intel Core i7-7740X 4.3 GHz 64GB RAM, and Intel Core i5-6600 3.30GHz, 16GB RAM (2x).

Data Collection Errors

No data was collected globally from 10/23–10/27 (VPN subscription lapsed), nor 12/16–12/29 and 1/15–1/19 in Seattle (VPN server outage). Some individual crawls also sporadically failed. In total, 33 of 312 daily crawl jobs failed.

5.3.2 Preprocessing Ad Content

Extracting Text from Ads

To enable large-scale analysis of the content of our dataset, we extracted the text of each ad. For ads where 100% of the visual content is contained in an image, we used the Google Cloud Vision API to perform optical character recognition (OCR). We extracted text from 877,727 image ads (62.6%) using this method. For native ads (i.e., sponsored content headlines), the text is contained in the HTML markup, so we automatically extracted the text from these ads using JavaScript. We extracted text from 524,518 native ads (37.4%) using this method.

Ad Deduplication

Many ads in our dataset appeared multiple times, some appearing tens of thousands of times. To reduce redundancy during qualitative coding and the runtime of machine learning tasks, we de-duplicated ads using the extracted text. We grouped our dataset by the domain of the landing page of the ad, and for each group, we used MinHash-Locality Sensitive Hashing[1] (LSH) to identify ads with a Jaccard similarity > 0.5. We maintained a mapping of unique ads to their duplicates, which we used later to propagate qualitative labels for unique ads to their duplicates, enabling analysis of the whole dataset. After deduplication, we obtained a subset of 169,751 unique ads.

5.3.3 Analyzing Ad Content with Topic Modeling

To help us broadly understand the content of the ads in our dataset, we used topic modeling to automatically create groups of semantically similar ads, allowing us to qualitatively analyze those groups. We experimented with several topic modeling and text clustering algorithms, and selected

[1]We used the MinHash LSH implementation from the datasketch Python library: http://ekzhu.com/datasketch/lsh.html.

Gibbs-Sampling Dirichlet Mixture Model (GSDMM) [222], which performed best on our dataset. Second, we automatically generated qualitative descriptions of each ad cluster, by using c-tf-idf to extract the most significant words from the text cluster [95]. We applied GSDMM & c-tf-idf to describe the topics in our overall ads dataset (Sec. 5.4.3) and political product ads (Sec. 5.4.7).

5.3.4 Analyzing Political Ads In-Depth

Our main focus is the content of political ads in our dataset. We defined a political ad broadly: any ad with political content, whether or not the advertiser was a political campaign. This includes ads with incidental political content, such as ads for products incorporating election imagery or ads promoting political news articles.

Our analysis of political ads consisted of three phases. First, we used machine learning to automatically identify political ads in our overall ads dataset. Second, we manually labeled the attributes of each political ad, such as the purpose of the ad, and the advertiser's political affiliation. Lastly, we performed quantitative analyses of the labeled political ad data.

Political Ads Classifier

To analyze political ads, we first needed to isolate political ads from the overall ads dataset. We implemented a binary text classifier based on the BERT language model, to classify our ads as political or non-political.

We started by generated a training set of political and non-political ads by labeling a random sample of ads in our dataset, obtaining 646 political ads and 1,937 non-political ads. We supplemented this data by crawling 1,000 political ads from the Google political ad archive [90] to balance the classes. We implemented the classifier by fine-tuning the DistilBERT model [184] for a binary classification task. We trained our model with a 52.5% / 22.5% / 25% Train / Validation /

Test split. Our model achieved an accuracy of 95.5%, and an F_1 score of 0.9. We ran the classifier on our deduplicated dataset (169,751 unique ads) and it classified 8,836 unique ads as political (5.2%).

Qualitative Analysis of Political Ads

Next, we we qualitatively coded the 8,836 unique political ads in our dataset to build a systematic categorization of the ads' content and characteristics [182]. Prior work in computer science and political science has also analyzed ad content using qualitative coding [226, 199]. We describe the development of our qualitative codebook and coding methods below.

We generated a qualitative codebook for political ads using grounded theory [147], an approach for generating themes categories via observation of the ground-level data. First, three researchers conducted a preliminary analysis of around 100 political ads each, creating open codes describing the characteristics of ads. We met to discuss and organized them into axial codes (i.e., multiple choice categories for different concepts) that best addressed our research questions.

Using these codes, three researchers coded the 8,836 ads, meeting multiple times during the process to iteratively refine the codebook based on new data. To assess the consistency of the coding, all coders coded a random subset of 200 ads, and we calculated Fleiss' κ (a statistical measure of intercoder agreement, $\kappa = 0$ indicates zero, $\kappa = 1.0$ indicates perfect) on this subset. We achieved an average $\kappa = 0.771$ across our 10 categories ($\sigma = 0.09$), indicating moderate-strong agreement [141].

Supplementing our qualitative codes, one researcher also labeled each campaign-related ad with the advertisers' name and legal classification (e.g., 501(c)(4) nonprofit), using information such as the "paid for" box in the ad, or the organization's website.

Our codeboook included three mutually exclusive high-level themes: **(1) campaigns and advocacy ads**, **(2) political product ads**, and **(3) political news and media ads**. To account for technical errors in crawling and classification, ads were classified as **Malformed/not political** if the extracted text and/or image content was incomplete or non-political, e.g., if screenshots failed to capture the whole ad, pop-ups or other material covered the ad, multiple ads were captured, incorrect model classification.

Campaigns and Advocacy Ads We define campaign and advocacy ads as those that explicitly addressed or promoted a political candidate, election, policy, or call to action. Within this category, we further define the level of election, the purpose of the ad, and advertiser-related information.

The *level of election* refers to candidate's jurisdiction, e.g., Senate elections were classified as federal. Specific codes of election level are: presidential, federal, state / local, no specific election, none. These codes are mutually exclusive. Note that "state / local" encompasses ballot initiatives and referenda as well as candidates.

The *purpose of ad* is a mutually inclusive code, meaning one campaign and advocacy ad can be assigned multiple purposes, e.g. voter information coupled with promoting a candidate. We coded for five purposes: promote candidate or policy; poll, petition, or survey; voter information; attack opposition; fundraise.

To facilitate insights into the advertisers, we identified the *Advertiser Affiliation and Organization Type* (both mutually exclusive). First, we labeled each advertiser by name, using information from the ad content and/or the landing page (e.g., disclosures that say "Paid for By…").

Then, for each advertiser, we investigated their legal organization status, based on criteria developed by Kim et al. [116]. Organizations listed on the Federal Election Commission website, or state elections boards were labeled as Registered Committees. 501(c)(3), 501(c)(4), and 501(c)(6) tax-exempt nonprofits, and legitimate foreign nonprofits that were visible in the Propublica Nonprofit Explorer or Guidestar were labeled as Nonprofit organizations. Advertisers whose websites' home pages were news front pages were labeled as news organizations (regardless of their legitimacy). Elections boards, state Secretaries of State, or any other state or local government institutions were labeled as Government Agencies. Advertisers who ran poll ads, and were listed FiveThirtyEight's Pollster Ratings were labeled as poll organizations. Ads from corporations and other commercial ventures were listed as businesses. Any ads where the advertiser was not identifiable was listed as unknown.

We also attempted to determine the political affiliation of the advertiser. We coded affiliations as Democratic party, Republican party, or independent if the advertiser was officially associated with those political parties (local or national branches), or a candidate running under that party's ticket. Codes of right/conservative, liberal/progressive, and centrist apply to advertisers not officially associated with a party, but that explicitly indicate their political alignment with words like "conservative" or "progressive", either in the ad itself or on their websites. Nonpartisan affiliation refers to explicitly nonpartisan advertisers or nonpartisan election positions, e.g. some local sheriff offices.

Political Product Ads We define political products ads as those centered on selling a product or service, using political imagery or content. This is further delineated into three mutually exclusive subcategories: political memorabilia, nonpolitical products using political topics, and political services.

Political memorabilia includes all ads marketing products with some form of political design, e.g. 2nd-amendment-themed apparel, keepsakes such as election trading cards, and merchandise such as Trump flags. This encompasses products sold for profit and those marketed as free or giveaways.

We coded ads as *nonpolitical products using political topics* if they used political messaging or context to advertise products ordinarily unrelated to politics. For instance, this covers investment firms marketing their stock reports in the context of election uncertainty.

Political services includes ads promoting services directly involved in political industry such as lobbying or election prediction sites.

Political News and Media Ads We define political news and media ads as those advertising a specific political news article, video, program, or event, regardless of the content style or quality. This categorization encompasses political clickbait and tabloid-style coverage of political figures as well as traditional news and media. We further define two mutually exclusive subcategories: sponsored articles / direct links to stores, and news outlets, programs, and events.

We coded ads as *sponsored articles / direct links to stories* if they advertised a specific news article or media piece, e.g. an authored story or video regarding a current event. We automatically assigned 1,038 ads to this category from Zergnet, a well-known content recommendation company, as we determined via their advertisement methods that all ads from their domain fit this category.

News outlets, programs, and events ads are distinguished from sponsored articles / direct links to stories in specificity, longevity, or reference. This category includes ads for political news outlets (as opposed to individual news pieces), lasting programs such as NBC election shows (in contrast to a single media clip), or future events such as panels or livestreams (rather than already existing news). We also included ads that were related media, such as podcasts, books, and interviews.

5.3.5 Ethics

Our data collection method had two types of impacts on the web. First, our crawler visited web pages and scraped their content. We believe this had a minimal impact: all sites we visited were public-facing content websites, contained no user data, and were visited by our crawlers no more than 4 times per day.

Second, our crawler clicked on ads to scrape the landing page of the ads. By clicking on the ads, we may cause the advertiser to be charged for the clickthrough (unless our click is detected as illegitimate), which is paid to the website and various middlemen.

We determined that clicking on ads was necessary because it was the only way for us to obtain the content and URL of the landing page for each ad. Many ads obscure their landing page through nested iframes and redirect chains. This data was needed for automatically determining the identity of the advertiser and for manually investigating the landing pages during qualitative coding (when the ad itself did not have sufficient context).

It is difficult to estimate the costs incurred to advertisers as a result of our crawls, but we believe the amount was low enough to be inconsequential. We cannot precisely determine the cost because the bid for each ad is not visible, and we do not know if advertisers pay using a cost-per-impression model or cost-per click model. For advertisers who pay based on impressions, we estimate the amount charged to be $3.00 per thousand impressions [204]. If all advertisers paid by impression, we estimate the total cost to *all* advertisers to be approximately $4,200. For the average advertiser, the mean number of ads we crawled was 63, and the median was 3, resulting in a mean cost of $0.19, and median cost of $0.009. If advertisers instead paid per click, we estimate a cost of approximately $0.60 per click [106]: in this case, the the mean advertiser would have been charged $37.80, and the median would have paid $1.80. The outlier advertisers in our dataset who received the most clicks were predominantly intermediary entities, such as Zergnet (36k ads), mysearches.net (26k ads), and comparisons.org (9k ads). These intermediaries place ads on other websites on behalf of advertisers on their platform, meaning that costs incurred for these intermediaries were spread among many individual sub-advertisers.

Stepping back, as we discuss further in Section 5.5, because of the distributed nature of the web ad ecosystem and the complex incentives of different stakeholders, we believe it is critical that external audits investigate the content and practices in this ecosystem, as we do in this study. Towards that end, we believe that the (small) costs of our study were justified. It is only through the process of clicking on ads, and evaluating the resulting landing pages, that can one fully understand the impact to users if they were to click on the ads. This is akin to the observation that malware websites may be linked from ads, potentially requiring search engine companies aiming to develop lists of known malware sites to engineer their crawlers to click on ads [169]. Moreover, similar methodologies have been used in prior works studying ads [173, 220].

5.3.6 Limitations

Our crawling methodology provided an incomplete sample of political advertising on the web. Our crawlers only visited a finite set of news and media websites, excluding other places that political ads appear, e.g., Facebook. Because we only visited each site once, we only saw a fraction of all ad campaigns running at that time. Our crawlers also only see political ad campaigns that were served to them — ongoing political ad campaigns may not have been shown to the crawler e.g. because of targeting parameters. We may have failed to load landing pages for ads because of detection and exclusion of our crawler by ad platforms. Due to VPN outages and crawler bugs, some days are missing from the data (Sec. 5.3.1).

We relied on categorizations from the fact checkers AllSides [11] and Media Bias/Fact Check [142] to identify the political bias of our input websites. 42% of our input sites had a rating: some uncategorized sites were non-political news websites (e.g., espn.com), while others may not have been popular enough to be rated.

Our automated content analyses were based on text extracted with OCR and did not use visual context from images. Some ads contained text artifacts, which negatively impacted downstream analyses. Based on the sample we labeled, we estimate that 18% ads in our dataset were malformed, i.e., impossible to read the ad's content. This was typically caused by modal dialogs (such as newsletter signup prompts) occluding the ad, which are difficult to automatically and consistently dismiss.

For the majority of ads, our data did not allow us to identify the ad networks involved in serving the ads. Though our crawler collected the HTML content of each ad (including iframes), this alone was rarely sufficient to identify ad networks.

Despite the above limitations, our dataset presents a unique and large-scale snapshot of political (and other) web ads surrounding the 2020 U.S. election. These include ads that do not appear in Google's (or others') political ad transparency reports. To support future research and auditing of this ecosystem, we will release our full dataset along with the publication of this paper, including ad and landing page screenshots, OCR data, and our qualitative labels.

5.4 Results

In this section, we present an analysis of the ads in our dataset. We begin by providing an overview of the dataset as a whole, including: How many ads appear overall, and how many of these are political ads of different types (Section 5.4.1)? How did the number of ads (political and non-political) change over time and location (Section 5.4.2)? Overall, what ad topics were common (Section 5.4.3)?

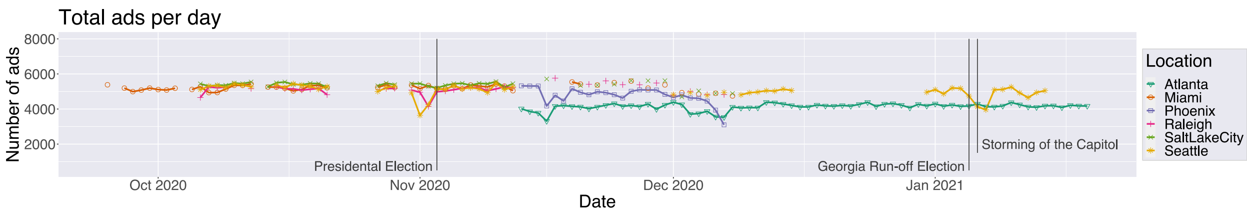
Then, we dive more deeply into our analysis of political ads. We investigate and characterize the sites political advertising appeared on (Section 5.4.4), advertisers running official campaign and advocacy ads (Section 5.4.5), misleading/manipulative campaign ads (Section 5.4.6), and political product ads (Section 5.4.7) and news and media ads (Section 5.4.8).

5.4.1 Dataset Overview

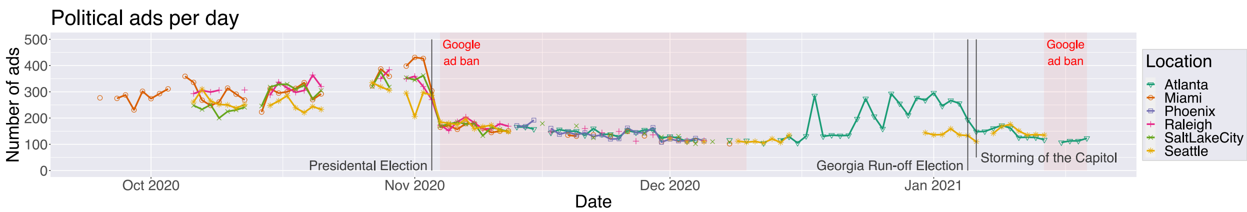
Between September 26, 2020 and January 19, 2021, we collected 1,402,245 ads (169,751 unique ads) from 6 locations: Atlanta, Miami, Phoenix, Raleigh, Salt Lake City, and Seattle. Our political ad classifier and qualitative coding, detected 67,501 ads (8,836 unique) with political content, or 3.9% of the overall dataset. During our qualitative analysis of political ads, we removed 11,558

Ad Categories	**Count**	**%**
Political News and Media	29,409	52%
Sponsored Articles	25,103	45%
News Outlets, Programs, Events	4,306	7%
Campaigns and Advocacy	22,012	39%
Level of Election		
Presidential	5,264	9%
Federal	5,058	9%
State/Local (including initiatives/referenda)	2,320	4%
No Specific Election	2,150	4%
None	7,220	13%
Purpose of Ad (not mutually exclusive)		
Promote Candidate or Policy	10,923	20%
Poll, Petition, or Survey	7,602	14%
Voter Information	4,145	7%
Attack Opposition	3,612	6%
Fundraise	2,513	4%
Advertiser Affiliation		
Democratic Party	5,108	9%
Right/Conservative	5,000	9%
Republican Party	4,626	8%
Nonpartisan	4,628	8%
Liberal/Progressive	1,673	3%
Unknown	781	1%
Independent	172	<1%
Centrist	24	<1%
Advertiser Organization Type		
Registered Political Committee	12,131	22%
News Organization	4,249	8%
Nonprofit	2,736	5%
Business	931	2%
Unregistered Group	913	2%
Unknown	781	1%
Government Agency	241	<1%
Polling Organization	30	<1%
Political Products	4,522	8%
Political Memorabilia	3,186	6%
Nonpolitical Products Using Political Topics	1,258	2%
Political Services	78	<1%
Political Ads Subtotal	55,943	100%
Political Ads - False Positives/Malformed	11,558	
Non-Political Ads Subtotal	1,347,810	
Total	1,402,245	

Table 5.2: Summary of the types of political ads in our dataset.



(a) The number of ads collected in each crawler location. We collected a relatively constant number of ads for each location.



(b) The number of political ads, classified as political by our text classifier, collected in each crawler location. The number of political ads was higher prior to the elections in November and January, were lower in the period after the elections.

Figure 5.2: Longitudinal graphs showing the number of total ads and political ads, collected in six locations from Sept. 2020 to Jan. 2021. Salient U.S. political events, as well as ad bans implemented by Google, are superimposed for context. Gaps from mid-Nov. to mid-Dec. are because we scheduled crawls on nonconsecutive days. Other gaps are due to VPN outages (see Section 5.3.1).

false positives and malformed ads (3,201 unique), resulting in 55,943 political ads. In Table 5.2, we show the number of political ads, across our qualitative categories. About a third of ads were from political campaigns and advocacy groups; over half advertised political news and media, and the remainder political products.

5.4.2 Longitudinal and Location Analysis

Ads Overall

We show the quantity of ads collected by location in Figure 5.2a. The number of ads per day stayed relatively stable in each location: consistently around 5,000 ads per day. The stability in ad counts indicates that changes in demand for ad space before and after the election had little impact on websites' ad inventory.

We collected about 1,000 fewer ads per crawler day in Atlanta than other locations. We do not know if this was due to differences in location-based targeting or an artifact of our crawling (e.g., limitations of the Atlanta VPN provider).

Political Ads

The amount of political ads over time and locations is visualized in Figure 5.2b. Leading up to the presidential election on Nov. 3, 2020, the number of ads per day in each location increases from less than 250 to peaks of 450. After election day, the number of political ads seen by crawlers sharply decreases, to below 200 ads/day. This decrease could be a natural consequence of less political attention following election day; it likely was also due to Google's first ad ban, from Nov. 4 to Dec. 10. We believe Google's ad bans help contextualize our results, given Google's large presence in web ads — but because we did not determine the ad networks used by each ad, we cannot prove a causal connection.

During Google's first ban, we collected 18,079 political ads. 76% of these ads were political news ads and political product ads. In the 4,274 campaign and advocacy ads during this period, 82% were from nonprofits and unregistered groups, such as Daily Kos, UnitedVoice, Judicial Watch, and ACLU. The remaining 18% (783 ads) were from registered committees, some from candidates in special elections (e.g., Luke Letlow, Raphael Warnock), but others from PAC groups specifically referencing the contested Presidential election. For example, an ad from the Democratic-affiliated Progressive Turnout Project PAC reads: "DEMAND TRUMP PEACEFULLY TRANSFER POWER – SIGN NOW".

Google lifted their political ad ban on Dec. 11. At this time, we only collected data from Seattle and Atlanta, and observed a rise in the number of political ads per day in Atlanta until the Georgia run-off election on Jan. 5, 2021, but no corresponding rise in Seattle. The increase in Atlanta came

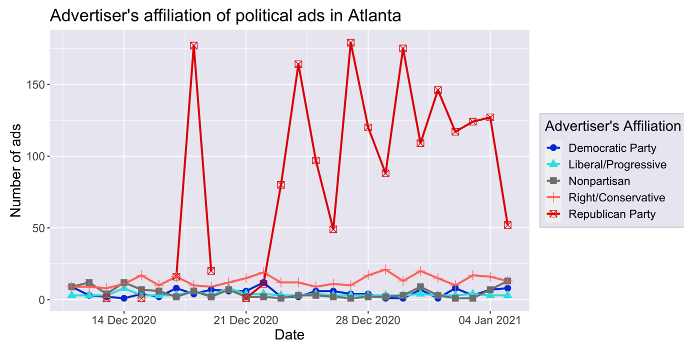


Figure 5.3: Campaign ads observed in Atlanta in Dec 2020–Jan 2021, prior to the Georgia special elections. Almost all ads during this time period were run by Republican groups.

almost entirely from Republican-affiliated committees — Democratic-affiliated advertisers seem to have bought very little online advertising for this election (Figure 5.3).

Following the Georgia election, we again observed a sharp drop in ads per day from the Atlanta crawler, matching the Seattle crawler at less than 200 political ads per day.

Though we observe that the volume of political advertising generally fell after elections, Google's ban on political advertising did not stop all political ads — other platforms in the display ad ecosystem still served political advertising.

5.4.3 Topics of Ads in Overall Dataset

To provide context before diving into political ads (Section 5.4.4-5.4.8), we present results from a topic model of the entire dataset. Table 5.3 displays the 10 largest topics in the data, each with a manually assigned topic description, the top c-TF-IDF terms, and the number of ads assigned to the topic.

The largest topic regarded "enterprise" ads, e.g., a Salesforce ad to "empower your partners to accelerate channel growth with external apps." The second largest topic included "tabloid" ads, e.g., "the untold truth of Arnold Schwarzenegger," as well as many clickbait and native advertisements.

Topic	c-Tf-IDF Terms	Ads	%
enterprise	cloud, data, business, software, marketing	93,475	6.7
tabloid	look, photo, star, upbeat, celebrity, celeb, truth	90,596	6.5
health	fungus, trick, fat, try, cbd, dog, doctor, knee, tinnitus	73,240	5.2
politics	vote, trump, biden, president, election, yes, sure	71,240	5.1
sponsored search	search, senior, yahoo, living, car, might, visa	70,613	5.0
entertainment	stream, original, music, watch, listen, tv, film	50,248	3.6
shopping (goods)	boot, shipping, jewelry, newchic, mattress, rug	49,457	3.5
shopping (deals/sales)	friday, black, deal, sale, cyber, review, monday	45,022	3.2
shopping (cars/tech)	suv, luxury, phone, commonsearch, deal, net, auto	44,179	3.2
loans	loan, mortgage, payment, rate, apr, fix, nml	43,629	3.1

Table 5.3: Top Topics in the Overall Ad Dataset.

The model's fourth largest topic, "politics", contained 71,240 ads: a 64.8% overlap with the 55,943 political ads identified by our classifier and qualitative coding.

These topics give us a sense of the context within which political ads were embedded. Like the web ad content studied in prior work [226, 227], political ads were surrounded by ordinary or legitimate ads for products and services, as well as low-quality and potentially problematic ads.

5.4.4 Distribution of Political Ads On Sites

Next, we examine how political ads were distributed across sites by political bias, misinformation label, and popularity.

Political Bias of Site Overall, we find that political ads appeared more frequently on sites with stronger partisan bias. Figure 5.4 shows the fraction of ads that were political across websites' political biases for mainstream and misinformation sites.

The percentages we calculate are the number of ads normalized by the total number of ads collected from sites for each level of bias. The number of ads collected from sites in each bias level varies, but no group of sites had overwhelmingly more ads. From Left to Right, the number of ads collected per site in each group were: 1,888, 1,950, 2,618, 2,092, and 2,172, and 1,676 had unknown bias.

Two-sample Pearson Chi-squared tests indicate a significant association between the political bias of the site and the percentage of ads that were political, for both mainstream news sites ($\chi^2(5, N = 1150676) = 25393.62, p < .0001$) and misinformation sites ($\chi^2(5, N = 206559) = 8041.43, p < .0001$). Pairwise comparisons using Pearson Chi-squared tests, corrected with Holm's sequential Bonferroni procedure, indicate that all pairs of website biases were significantly different ($p < .0001$).

On mainstream news sites, conservative sites had more political ads than others; 9% and 10.3% of ads on right-leaning and right sites were political, but only 6.9% and 4.4% of ads on left and left-leaning sites. On misinformation sites, 26% of ads on left sites were political, substantially more than right leaning sites. In 4 of the 7 left misinformation sites (AlterNet, Daily Kos, Occupy Democrats, Raw Story) over 19% of ads were political.

We also find that political advertisers tend to target sites matching their political affiliation: Democratic and liberal groups ran the majority of their ads on left-of-center sites, and likewise for Republican and conservative groups on right-of-center sites (Figure 5.5). In particular, ads for Democratic political candidates and progressive nonprofits and causes ran substantially more on 2 of 7 Left misinformation sites (Daily Kos and Occupy Democrats).

Two-sample Pearson Chi-squared tests indicate a significant association between the political bias of the site and the number of ads based on the advertiser's political affiliation, for both mainstream news sites ($\chi^2(25, N = 1,150,676) = 22575.49, p < .0001$) and misinformation sites ($\chi^2(20, N = 206,559) = 22168.50, p < .0001$). Pairwise comparisons using Pearson Chi-squared tests, corrected the Holm-Bonferroni method, indicate that all pairs of website biases were significantly different ($p < .0001$) except for the (Lean Left, Uncategorized) Misinformation Sites.

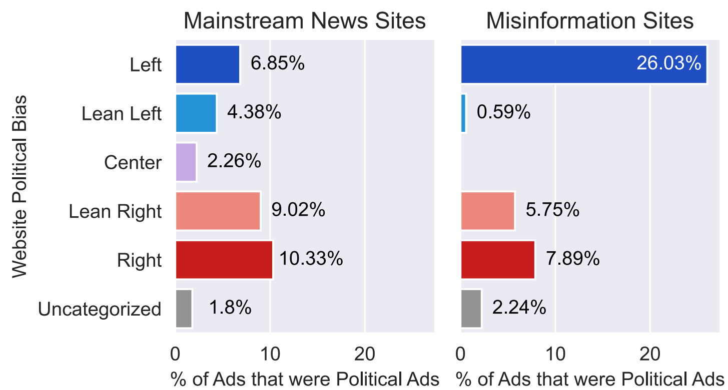


Figure 5.4: The percentage of ads, out of all ads on those sites, that were political, by sites' political bias and misinformation label. Higher percentages of ads on partisan sites were political, compared to centrist/uncategorized sites.

Site Popularity We found little relationship between site popularity and the number of political ads on it (Figure 5.6). While sites hosting many political ads tended to be popular politics sites (e.g., dailykos.com, mediaite.com), some popular sites (e.g., nytimes.com, cnn.com) ran <100 political ads. A linear mixed model analysis of variance indicates no statistically significant effect of site rank on the number of political ads ($F(1, 744) = 0.805$, *n.s.*).

At a high level, we find that political ads are seen more on websites that are political and partisan in nature. We hypothesize that this is either due to contextual targeting (political groups advertising to co-partisans), and/or because neutral news websites choose to block political advertising on their sites to appear of impartiality.

5.4.5 Advertisers of Campaign Ads

Next, we analyze the advertisers who ran campaign and advocacy ads: their organization type, their affiliations, and how many they ran. Figure 5.7 shows these ads by organization type and affiliation.

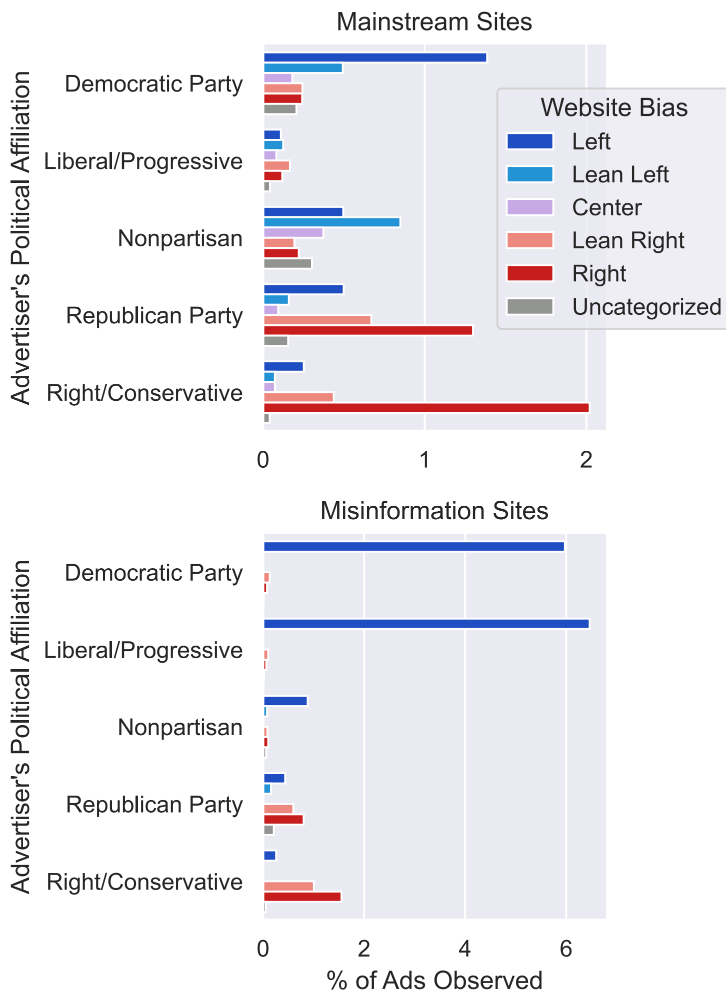


Figure 5.5: The percentage of ads observed on websites from advertisers of different political affiliations, by the political bias and misinformation label of the website. Advertisers tended to run ads on websites aligned with their politics.

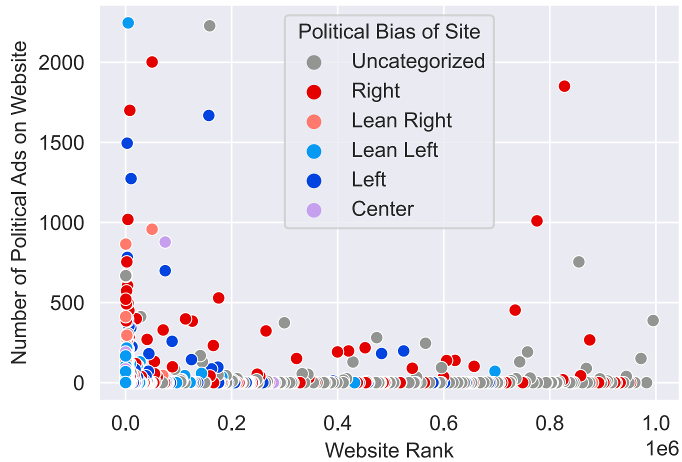


Figure 5.6: The total number of political ads observed on each site, by the site's Tranco rank. Though the largest outliers in terms of political ads tend to be popular sites, many popular sites show few if any political ads.

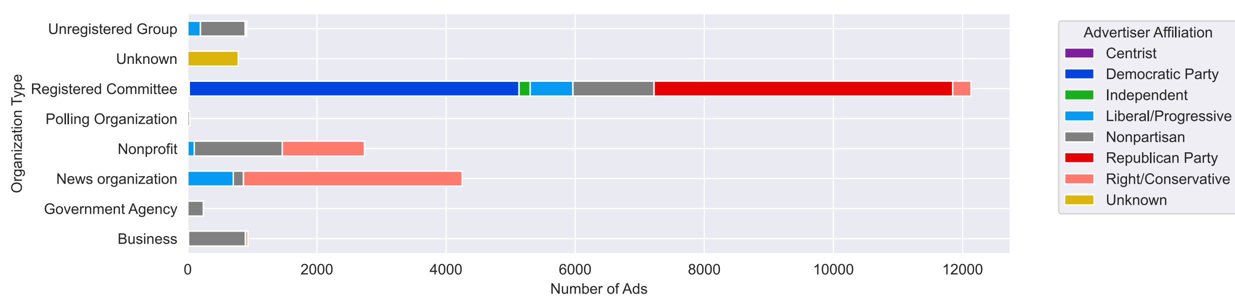


Figure 5.7: Campaign and advocacy ads by organization type of the advertiser, color-coded by the political affiliation of the advertiser. Ads from registered committees dominated, roughly evenly divided between Democratic and Republican ads, but ads from news organizations and nonprofits were more heavily conservative and nonpartisan respectively.

Registered Committees Most campaign ads (12,131, 55.1%) were purchased by registered committees (FEC or state PACs). These ads were roughly evenly split between Republican- and Democratic-affiliated committees, including official candidate committees, like Biden for President, as well as Hybrid PACs and party-affiliated Super PACs, such as the Progressive Turnout Project and the Trump Make America Great Again Committee. These also include candidate committees for other state, local, and federal offices.

Nonprofits We observed campaign ads from nonpartisan nonprofits, e.g., AARP (259 ads, 1.2%), ACLU (256 ads, 1.2%), as well as explicitly conservative ones, e.g., Judicial Watch (504 ads, 2.3%), Pro-Life Alliance (471 ads, 2.1%). Few explicitly liberal nonprofits ran ads under our categorization system. However, some may consider self-described nonpartisan organizations as liberal, e.g., issue organizations like the ACLU, or voting rights groups like vote.org.

News Organizations Some news organizations ran explicitly political ads to promote candidates or policies — these were mostly conservative-leaning organizations. The top advertisers in this group are not well-known, e.g., ConservativeBuzz (1,199 ads, 5.4%), UnitedVoice.com (800 ads, 3.6%), and rightwing.org (393 ads, 1.8%). ConservativeBuzz does not have a website, despite claiming to be a news source on their landing page; UnitedVoice and rightwing.org are ranked 248,997 and 539,506 on the Tranco Top 1m.

Other advertisers in this category are more well-known, e.g., Daily Kos, a liberal blog (690 ads, 3.1%, site rank 3,218); Human Events, a conservative newspaper (390 ads, 1.8%, rank 19,311); Newsmax, a conservative news network (117 ads, 0.5%, rank 2,441).

Unregistered Groups Unregistered groups ran a small number of ads. The top advertiser was "Gone2Shit", a campaign from the marketing firm MullenLowe, which ran 228 ads for a humorous voter turnout campaign. The U.S. Concealed Carry Association ran 162 ads. Beyond these top two, a number of "astroturfing" groups or other industry interest groups ran ads, such as "A Healthy Future" (lobbying against price controls on Rx drugs), "Clean Fuel Washington", and "Texans for Affordable Rx" (a front for the Pharmaceutical Care Management Association, based on investigating their website). Other top ads came from unregistered, left-leaning groups, such as "Progress North" and "Opportunity Wisconsin", which describe themselves as grassroots movements. We also saw a small number of groups consisting of coalitions of registered nonprofits, who collectively fund an ad campaign, such as "No Surprises: People Against Unfair Medical Bills" and "votewith.us".

Businesses and Government Agencies Some businesses, e.g., Levi's, Absolut Vodka, ran political ads: mostly nonpartisan ads for voter registration. State/local election boards also ran voter information ads, e.g. the NYC Board of Elections.

5.4.6 Misleading Political Polls

Focusing now on the content of ads in our campaign and advocacy category, rather than the advertisers, we highlight the use of polls, petitions, and surveys, many of which appear to contain misleading content, and manipulate users into providing their email addresses.

The purpose of many online political petitions and polls are to allow political actors to harvest personal details like email addresses, so that they can solicit donations, canvas, or advertise to those people in the future [172]. This phenomenon is present in our dataset. In a few cases (30 ads), ads we labeled as polls or petitions linked to nonpartisan public opinion polling firms such as YouGov and Civiqs, but most ads were from political groups, and had landing pages asking people to provide their email addresses.

We observe that poll and petition ads are more common from politically conservative advertisers. In Figure 5.8, we visualize the number of poll ads by the political affiliation of their advertisers.

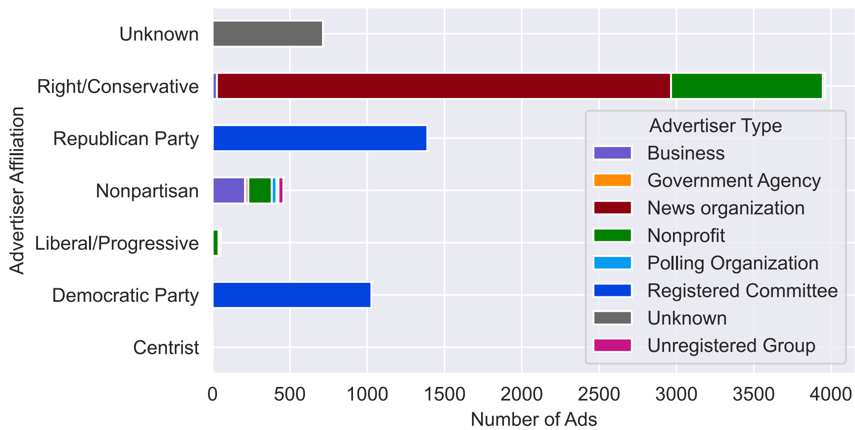
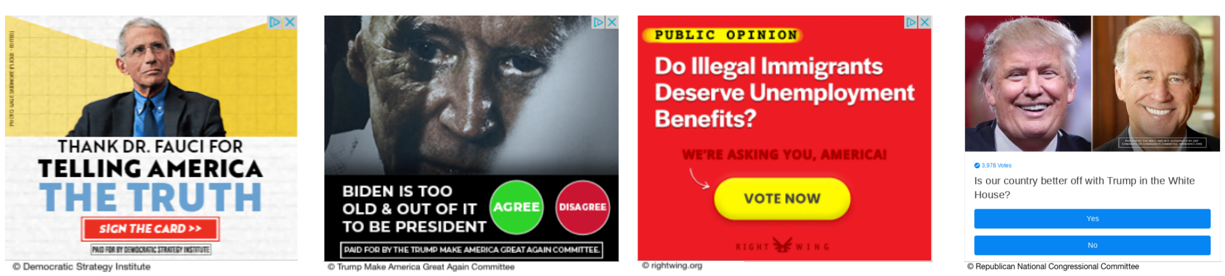


Figure 5.8: The political affiliation and organization types of poll/petition advertisers. These ads were primarily run by unaffiliated conservative advertisers, mostly news organizations and nonprofits.



(a) (b) (c) (d)

Figure 5.9: Examples of political ads purporting to be polls, including from: a Democratic-aligned PAC (a), the Trump campaign (b), a conservative news organization/email harvesting scheme (c), and a Republican-aligned PAC (d).

Non-affiliated conservative groups (mostly news organizations and nonprofits) ran the highest number of poll and petition ads (3,960 ads, 52% of total), followed by Republican party committees (1,389, 18.2%). Democratic committees ran fewer poll ads than their Republican counterparts (1,027 ads, 13.5%), while non-partisans and nonaffiliated liberals rarely use poll ads (458 ads, 6%; 53 ads, 0.6%).

Poll ads also made up a greater proportion of ads on right-leaning websites than other sites: 2.2% on Right and 1.1% on right-leaning websites were polls and petitions, compared to 1.1% on Left, 0.2% on left-leaning, and 0.2% on center sites.

Next, we describe several topics and manipulative tactics used by poll ads, which differ across political affiliations.

Democratic-Affiliated Groups Most poll or petition ads from Democratic-affiliated groups were for highly partisan issue-based petitions, e.g., "Stand with Obama: Demand Congress Pass a Vote-by-Mail Option", "Official Petition: Demand Amy Coney Barrett Resign - Add Your Name". However, some petitions used even more contrived scenarios, such as posing as a "thank you card" for important politicians (Figure 5.9a). These ads were run by affiliated PACs rather than party or candidate committees, such as the National Democratic Training Committee (290 ads), Progressive Turnout Project (282 ads), and Democratic Strategy Institute (215 ads).

Republican-Affiliated Groups The Trump campaign ran 906 ads with positive and neutral polls promoting President Trump and 479 ads with polls that attacked their opponent (e.g., Figure 5.9b). Other Republican committees, such as the NRCC, used the LockerDome ad platform to run generic-looking polls not clearly labeled as political (e.g., Figure 5.9d). Moreover, Lockerdome was also used by unaffiliated advertisers, e.g., "All Sears MD", rawconservativeopinions.com, to run nearly identical-looking ads that were used to sell political products; this homogenization makes it difficult for users to discern the nature of such ads. We also found 5 Lockerdome ads from the "Keep America Great Committee," whose operators turned out to be using it to commit fraud and pocket donations [137].

Conservative News Organizations The largest subgroup of advertisers that used polls were right-leaning news organizations, such as such as ConservativeBuzz, UnitedVoice, and rightwing.org. Some polls use neutral language, e.g., "Who Won the First Presidential Debate?", while others used more provocative language, e.g., "Do Illegal Immigrants Deserve Unemployment Benefits?" (Figure 5.9c).

Journalistic investigations have found that advertisers like ConservativeBuzz purport to be conservative news organizations but are actually run by Republican-linked digital marketing firms. Appearing as news, many of their stories are plagiarized and/or serve a political agenda. Their misleading poll ads are an entry point for harvesting email addresses for their mailing lists. They profit from these mailing lists by sending ads to their subscribers, including ads from political campaigns [136, 18].

Our data backs up these findings. We inspected poll ads from ConservativeBuzz, UnitedVoice, and rightwing.org, who comprise 55% of poll ads from Right/Conservative advertisers, and 29% of poll ads overall. The landing pages of their ads often asked for an email address to submit poll responses (Figre 5.10). We looked up these advertisers in the Archive of Political Emails to see the content of the emails that they send to subscribers [2]. We found that their emails often contained a mix of spam for various products (Subject: "This Toxic Vegetable Is The #1 Danger In Your Diet"), biased or inaccurate political news (Subject: "Fauci-Obama-Wuhan Connection Exposed in This Bombshell Report"), or a combination of the two (Subject: "URGENT – Think Trump Won? You need to see this...", selling a Trump mug).

[2] https://politicalemails.org/

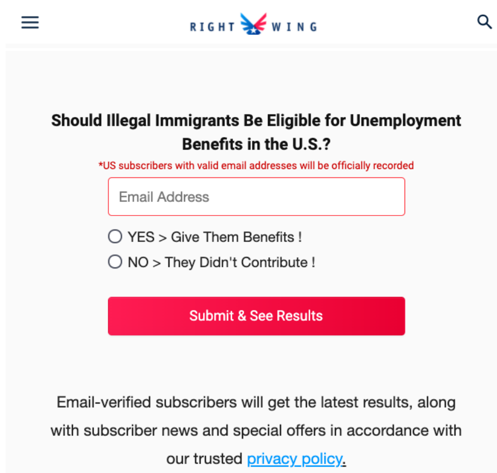
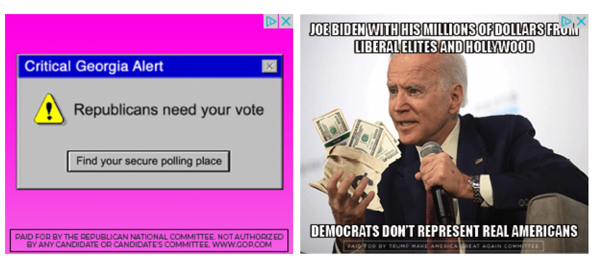


Figure 5.10: The landing page of the poll from Figure 5.9c. Viewers are asked to submit an email address to vote in the poll, and are signed up a newsletter. Prior reporting has shown this is typically a scheme to generate mailing lists and audiences for political campaigns to advertise to. © rightwing.org

Other Misleading Campaign Ads: Phishing Ads and Memes

Though many campaign and advocacy ads in the dataset were potentially misleading or factually incorrect, we highlight two types that appeared particularly egregious.

In December, the Republican National Committee ran ads that imitate a system alert popup, like an impersonation attack (Figure 5.11a). We found 162 ads of this style in our dataset. Though



(a) (b)

Figure 5.11: Other misleading campaign ads: an RNC ad imitates a system popup (a), and a Trump campaign meme-style ad attacking Biden (b). Images © Republican National Committee and © Trump Make America Great Again Committee.

the popup's style is outdated, it is generally misleading for ads and websites to imitate operating system dialogues or other programs.

Before the general election, the Trump Make America Great Again Campaign launched several attack ads in the style of an "image macro" meme. They featured (obviously) doctored photos of Joe Biden, holding Chinese flags, handfuls of cash, or depicting him approving of rioting (Figure 5.11b). We found 119 meme-style ads in our dataset. Though attack ads and smears are fairly normalized, we did not observe the use of memes for attacks by any other campaigns. These ads contrast with more polished ads placed by other campaigns, and could be misleading if users assume meme-style ads are placed by other users, not an official political campaign.

5.4.7 Political Product Ads

We now consider ads in our dataset that used political content to sell products, divided into three categories.

Ads for Memorabilia

We observed 3,186 ads for political memorabilia, including clothing with slogans, collectibles, and novelty items. These ads were placed by commercial businesses – none were affiliated with political parties. Our GSDMM model produced 45 topics for political memorabilia ads; Table 5.4 shows the top seven.

We observe that the majority of memorabilia ads are targeted towards conservative consumers. 2,175 advertisements (68.3% of memorabilia ads) contained "Donald" and/or "Trump". Seven of the top ten topics are directly related to Trump, selling items such as special edition $2 bills (Figure 5.12a), electric lighters, garden gnomes, and trading cards.

Some memorabilia ads targeting conservatives used potentially misleading practices. While

Topic	Weighted c-TF-IDF Terms	Ads
Trump wristbands and lighters	America, charger, USB, butane, require, vote, include	643
“free” Trump flags	dems, hate, give, foxworthynews, away, claim, flag	300
Trump electric lighters and garden decorations	spark, instantly, generate, one, click, open, light, garden	253
$2 bills and “currency”	legal, tender, authentic, official, Donald, USA, make	186
Israel support pins	Israel, request, pin, Jew, fellowship, Christian	172
Trump camo hats, bracelets, and coolers	camo, gray, anywhere, discreet, go, sale, way, bracelet	156
Trump coins and bills	left, gold, coin, Democrat, upset, hat, supporter, value	133

Table 5.4: Top Topics in Political Memorabilia Ads

some ads clearly advertised themselves as products, others disguised the memorabilia as “free” items, but requires payment to cover shipping and handling. Many ads did not clearly disclose the name of the advertiser. Some straddled the line between product ads and clickbait by making claims that the product “angered Democrats” or would “melt snowflakes.” We also observed many collectible bills and coins, advertised as "Legal U.S. Tender", by sellers such as Patriot Depot, making dramatic claims like "Trump Supporters Get a Free $1000 Bill."

We observed far fewer ads for left-leaning consumers; the first topic containing left-leaning products was the 15th largest at 71 ads. Ads targeting liberals include a pin for “flaming feminists” or a deck of cards themed around the 2020 Senate Impeachment Trial of former President Trump (Figure 5.12b).

Ads Using Political Context To Sell Something Else

We observed 1,258 ads that leveraged the political climate for their own marketing. Some of these ads were from legitimate companies, such as Capitol One advertising their alliance with the Black Economic Alliance to close opportunity gaps, or the Wall Street Journal promoting their market insight tools. However, many others were from relatively unknown advertisers

Topic (Context)	Weighted c-TF-IDF Terms	Ads
Hearing devices (congress action)	hearing, aidion, slash, price, health, hear, act, sign, Trump	266
Retirement finance (congress action)	sucker, punch, law, pension, even, rob, retire, IRA	205
Investing (election-time)	former, presidential, Stansberry, congressional, veteran	123
Seniors' mortgage (congress action)	amount, reverse, senior, Steve, calculate, tap, age	97
Banking (racial justice)	JPMorgan, Chase, advance, co, racial, important, equality	66
Portfolio finance (election-time)	inauguration, money, Jan, wonder, oxford, communique	63
Dating sites (for Republicans)	Republican, single, date, woman, wait, profile, view	54

Table 5.5: Top Topics in Ads About Nonpolitical Products Using Political Context

peddling get-quick-rich schemes, like stocks that would "soar" from Biden winning the election (Figure 5.12c) or election-proof security in buying gold.

Our GSDMM model found 29 topics for ads categorized as nonpolitical products using political context. Table 5.5 details the largest 7 topics. The most prominent political contexts used for these topics were Congress (e.g., legislation related to the product) and the 2020 election. Finance related topics in particular often cited market uncertainty around the election, e.g., referencing how a certain outcome might affect stocks and promoting their product as a hedge or chance to capitalize. Notably, three of the top four topics targeted older audiences: "hearing devices," "retirement finance," and "seniors' mortgage."

Where did political product ads appear?

We find that political product ads appeared much more frequently on right-of-center websites (Figure 5.13). This finding aligns with the qualitative content that we observed in these ads — a large amount of Trump memorabilia, and "scare" headlines about the election outcome. Two-sample Pearson Chi-Squared tests indicate a statistically significant association between the political bias of the site and the number of political product ads observed, both for mainstream news sites ($\chi^2(10, N = 1,150,676) = 4871.97, p < .0001$) and misinformation sites ($\chi^2(8, N = 206,559) =$

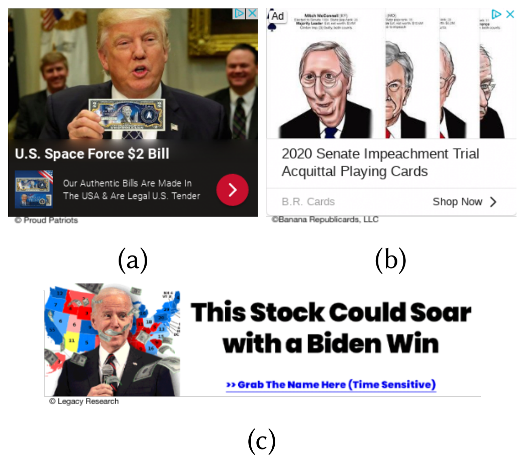


Figure 5.12: Examples of political product ads, including those selling memorabilia (a-b) and those using the political context to sell something else (c).

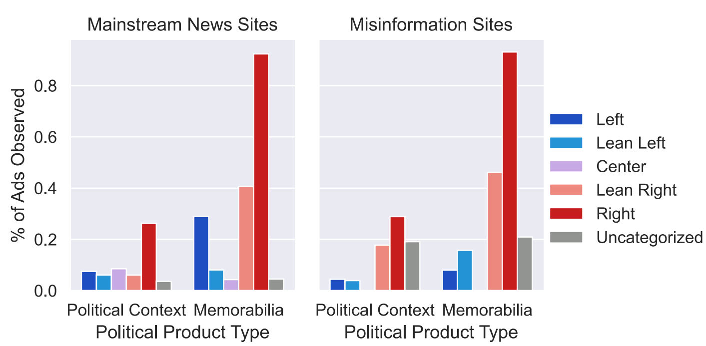


Figure 5.13: The percentage of ads observed that were for political products, by the political bias of the site. Right sites more frequently hosted ads for political products, both on misinformation and mainstream sites, and both for memorabilia or nonpolitical products using political contexts.

414.75, $p < .0001$). Pairwise comparisons using Pearson Chi-squared tests, corrected with the Holm-Bonferroni method, indicate that all pairs of website biases were significantly different ($p < .0001$), except for the following pairs on misinformation sites: (Lean Left, Lean Right), (Lean Left, Left), and (Lean Left, Uncategorized).

5.4.8 Political News and Media Ads

We observed 29,409 ads that were related to political news and media content. At 52.0% of all political ads, this was the most populous category and accounted for more than either of the other two categories. Unlike the product ads primarily selling goods or services, these ads advertised information or information-related services. We categorize these news and media ads into two groups: those that advertised specific political news articles, and those that advertised political outlets, events, or related media. Article ads contained a range of sensationalized, vacuous, or otherwise misleading content, especially with "clickbait-y" language that enticed people to click.

Sponsored Content / Direct Article Links

Overall, we find that most political news and media ads were sponsored content or links to articles (25,103 ads, 85.4%). Some of these ads reported substantive content, e.g., linking to a review of a documentary: "'All In: The Fight for Democracy' Tackles the Myth of Widespread Voter Fraud." Others were clickbait only using political themes for attention, e.g., "Tech Guru Makes Massive 2020 Election Prediction."

Misleading Ads and Headlines Given that our ads were primarily scraped from news and media websites, many appeared as native ads that blend into the other content, albeit with an inconspicuous "Sponsored content" or similar label. Further, the headline shown in a political article ad did not always align with the actual content on the clickthrough page. For example,

Word	Freq.
trump	1,050
biden	415
elect	314
read	235
new	219
top	215
articl	196
presid	176
thi	170
video	162

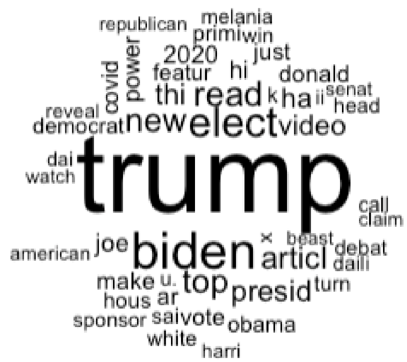


Figure 5.14: Frequencies of the top 10 words in political news article ads, and a word cloud showing the top 50. Ad text was deduplicated by ad, and then tokenized and lemmatized.

the ad shown in Figure 5.16a links (via a Zergnet aggregation page) to an article[3] that recounts Vanessa Trump's life before marrying Donald Trump Jr., instead of after, as the title suggests. Many Zergnet ads with headlines implying controversy were unsubstantiated by the linked article.

Unique Word Frequency Analysis We looked at the most common words in political article ads by first deduplicating ads (Section 5.3.2), then tokenizing and lemmatizing the ad text. The top 10 words and their frequencies, as well as a word cloud of the top 50 words, is shown in Figure 5.14. Among the top 50, we find frequent mentions of "trump" (1,050 times, more than double the next most common word, "biden"), as well as other politically relevant terms and names. Many of top 50 words reveal the general tone of these article ads, which often emphasize urgency, e.g., "new," "top," or scandal, e.g., "just," "claim," "reveal," "watch." The colloquialism "turn heads" was particularly common, e.g., "What Michigan's Governor Just Revealed May Turn Some Heads."

Ads Mentioning Top Politicians Overall, Trump and Biden were referenced in ads much more often than Pence and Harris (Figure 5.15). Within political news and media ads, "Trump" is

[3] https://www.thelist.com/161249/the-stunning-transformation-of-vanessa-trump/

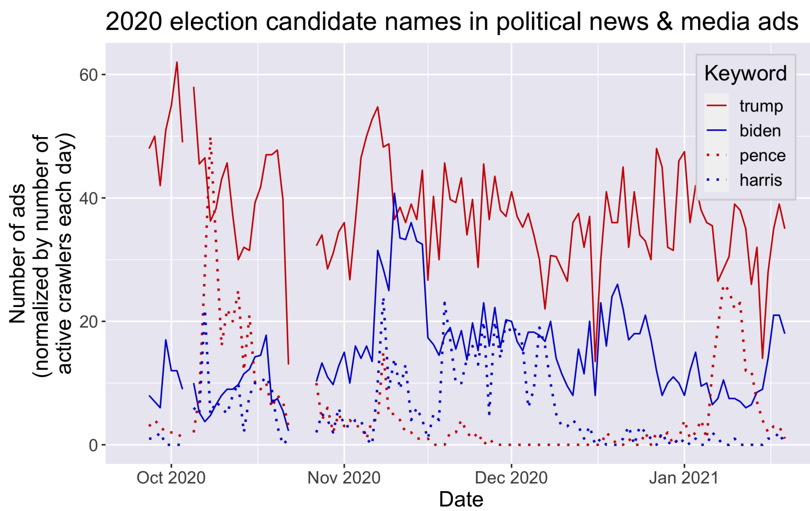


Figure 5.15: Number of ads that included the first and last names of the 2020 Presidential and Vice Presidential candidates over time.

referenced in ads 2.5x more than "Biden" (11,956 ads vs. 4,691, or 40.7% vs. 16.0%), even even after the election. Eight of the top ten ads mentioning Trump actually involve his family: e.g., "Trump's Bizarre Comment About Son Barron is Turning Heads" (1,377 ads, 4.7%), or "Eric Trump Deletes Tweet After Savage Reminder About His Father" (415 ads, 1.4%). The top 10 ads mentioning Biden imply scandals with his wife, e.g., Figure 5.16b (1,267 ads, 4.3%), and his health, e.g., "Ex-White House Physician Makes Bold Claim About Biden's Health" (423 ads, 1.4%).

Looking at the VP candidates, Pence is referenced in ads frequently during the run up to the election and immediately following the insurrection at the Capital, while a spike in the mentions of Harris occurs in late November and early December. Some of the top 10 ads mentioning Pence connect him to high-profile events, including the VP debate ("The Pence Quote from the VP Debate That Has People Talking," 143 ads, 0.5%) and the U.S. Capitol storming (Figure 5.16c). Some of the top 10 ads mentioning Harris highlight her ex ("Why Kamala Harris' Ex Doesn't Think She Should Be Biden's VP," 246 ads, 0.8%) as well as her gender ("Women's Groups Are Already Reacting Strongly to Kamala," 51 ads, 0.2%).

(a) (b) (c)

Figure 5.16: Examples of clickbait political news and media articles. © Zergnet

Frequent Re-Appearances of Sponsored Content Out of 25,103 political article ads, we counted only 2,313 unique ads, meaning that many political article ads were shown to our crawler multiple times. On average, a single (unique) political article ad appeared to our crawlers 9.9 times, compared to 9.3 times for campaign ads and 5.1 times for product ads. The frequent re-appearance of political article ads is likely an artifact of content farms' practice of producing high quantities of low-quality articles solely for revenue from clicks [36]. 79.4% of all political news articles were run by Zergnet, which accounted for 19,690 ads and only 1,388 unique ads. Other top ad platforms for political news articles were Taboola (10.0%), Revcontent (5.7%), and Content.ad (1.8%).

Political Outlets, Programs, Events, and Related Media

A small portion of political ads, just 4,306 (7%), advertised a political news outlet, event, or other media content. This includes ads run by well-known news organizations, e.g., Fox News, The Wall Street Journal, The Washington Post, that advertised their organizations at large, as well as highlighting specific events, such as CBS's coverage of the "Assault on the Capitol" (Figure 5.17a), or special programs about the presidential election. Ads were also run by less-well known news organizations advertising themselves or their events, e.g., The Daily Caller, a right-wing news and opinion site, or advocacy groups and nonprofits, e.g., Faith and Freedom Coalition (Figure 5.17b), a conservative 501(c)(4). We also observed ads about books, podcasts, movies, and more.

(a) (b)

Figure 5.17: Examples of political news and media ads about political outlets and events. Images © CBS and © Faith and Freedom Coalition

Where did political news and media ads appear?

Political news and media ads appeared more often on right-of-center sites, compared to center and left-of-center sites (Figure 5.18). Two-sample Pearson Chi-Squared tests indicate a statistically significant association between the political bias of the site and the number of political news and media ads, both for mainstream news sites ($\chi^2(10, N = 1,150,676) = 16729.34, p < .0001$) and misinformation sites ($\chi^2(8, N = 206,559) = 3985.43, p < .0001$). Pairwise comparisons using Pearson Chi-squared tests, corrected with the Holm-Bonferroni method, indicate that all pairs of website biases were significantly different ($p < .0001$). Nearly 5% of ads on both Right and Lean-Right sites are sponsored content, but only 3.9%, 2.2%, and 0.8% on Left, Lean Left, and Center sites.

5.5 Discussion

5.5.1 Concerns About Problematic Political Ads

Our investigation adds to a growing body of work studying potentially problematic content in online ads, political and otherwise (see Sec. 5.2). Here, we discuss further the potential harms

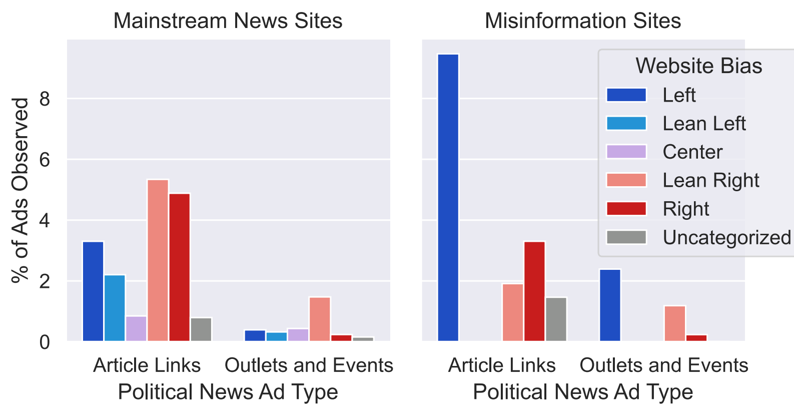


Figure 5.18: The number of political news ads observed per site, by the political bias of the site. Right sites more frequently host political news ads than others.

from the problematic political ads we found.

Manipulative Polls The most common manipulative pattern we observed in our political ads was the poll-style ad. We view these ads as problematic for two reasons. First, they manipulate people into clicking on ads by appealing to political motivations with (seemingly) clickable user interface elements. Second, once users click, they often ask users to provide personal information for further manipulation, e.g., to put them on manipulative email newsletters [140].

Political Clickbait We observed attention-grabbing news and media ads that were not official political ads and thus do not appear in political ad transparency libraries. However, these ads are misleading: they are often designed to looks like real news articles, but the political controversies they imply (e.g., "Viral Video Exposes Something Fishy in Biden's Speeches," Figs. 5.16a-5.16c) are not usually substantiated by the underlying articles. Though we believe these ads' goal is to entice clicks for ad revenue, we worry that the provocative political "headlines" contribute to a climate of hyper-partisan political communication and muddy the information ecosystem to which voters are exposed. We argue that this type of political-adjacent advertising requires additional scrutiny from ad platforms and the public.

Exploitative Product Ads Most ads aiming to make money through the sales of products and services are legitimate, identifiable as ads, and meet expectations of appropriateness [227]. However, we identified product ads that we would consider exploitative, e.g., that promise "free" products that turn out to not to be. Though such ads are not unique to political contexts, we observed many that leverage political controversy to attract potential buyers.

Misleading Political Organizations Online ads (particularly native ads) have been criticized for being potentially hard to identify as ads, and thus regulated to require disclosure [34, 69]. We observe that these issues are compounded in a political context, where the advertiser's identity — e.g., political leaning, official (or not) political organization — is (or should be) key to a user's assessment of the ad. Being mistaken for a legitimate, official political organization can benefit problematic advertisers (e.g., exploitative product sellers or the fraudulent "Keep America Great Committee" [137]).

Partisan Ad Targeting We observed more political ads, and more of the problematic ads that we discussed above, on more partisan websites, particularly right-leaning sites, as well as on low-quality and misinformation sites. Ad targeting in itself is not problematic, and naturally, political advertisers would wish to reach people with partisan alignments most likely to click on a given ad. However, we raise two concerns: first, the continued polarization of U.S. political discourse, reinforced by online ads; second, the risk that more vulnerable people are targeted with more manipulative and exploitative political ads.

5.5.2 Recommendations and Future Work

Recommendations for Ad Platforms and Policymakers Political ads are already strongly regulated due to its sensitivity. We argue that ad platforms (which make and enforce ad policies) and policymakers (e.g., the FTC or FEC) should also consider the potential harms from ads not currently violating of existing policies. Many of the problematic ads that we saw were *not* official political ads but leveraged political themes and could have political ramifications (e.g., spreading misinformation via clickbait headlines). Ad platforms and regulators should consider these ads alongside official political ads in transparency and regulation efforts.

It is worth noting that there were types of problematic political ads that we did *not* observe. In a preliminary qualitative analysis, we did not find ads providing false voter information, e.g., incorrect election dates, polling places, or voting methods. While that does not mean they did not exist, it nevertheless suggests that ad platforms are regulating the most egregiously harmful ads.

The extreme decentralization of the online ad ecosystem poses additional challenges for ad moderation. Though Google periodically banned political ads during our data collection, we continued to see political ads, including problematic political ads, placed by other ad platforms. Thus, we call for more comprehensive ad moderation standards (and perhaps regulation) across advertising platforms — while recognizing the complex financial and political incentives that may hamper the clear-cut adoption of regulation [96].

Future Research Future research should continue to audit ad content and targeting. While our study has focused on web ads appearing on news and media websites, the online ad ecosystem is large and requires analysis with different data collection and analysis methods. Future work should (continue to) consider political and other ads across various platforms — social media, mobile web and apps — and sites. Moreover, we focused on U.S. political ads, but future research should also critically study the role of online ads in non-U.S. political contexts or around other historical events.

Future work should also directly study people who view these ads, to better understand the actual impact of potentially problematic ads and for different user populations.

To enable other researchers to further analyze our collected ads, our dataset and codebook are available at: https://badads.cs.washington.edu/political.

5.6 Conclusion

We collected ads from 745 news and media sites around the time of the 2020 U.S. elections, including 55,943 political ads, which we analyzed using quantitative and qualitative methods. We identified the use of manipulative techniques and misleading content in both official and non-official political-themed ads, and we highlight the need for greater scrutiny by ad platforms and regulators, as well as further external study and auditing of the online ad ecosystem.

Chapter 6

What Factors Affect Targeting and Bids in Online Advertising?

This chapter takes a slight diversion from the focus on "bad" ad content, and instead investigates the extent to which ads are targeted, how much advertisers bid to run their ads, and whether these behaviors differ across demographics. These questions are largely opaque outside of the online advertising industry. However, understanding the phenomena of targeting and tracking is important for revealing the economics and practices of online advertising, which could shed light on the incentives and practices for "bad ads" as well.

This chapter presents field measurements of targeting in online advertising using an online panel of 286 participants. In the study, participants installed a browser extension collected data on display ads on a set of 10 websites, including screenshots and the value of the bid placed by the advertiser to show the ad, and participants' qualitative perceptions of targeting for a sample of their ads. The study analyzes trends in targeting and bidding across websites and ad topics. The dataset also uniquely enables investigation demographic factors in targeting and bidding, compared to crawler-based studies. Among the findings in the study, we observe large differences in bid values and ad topics between websites, but small-to-no differences between demographic groups. Confirming findings from prior work, we also find that high outliers in bid values (10x higher than baseline) may be indicative of retargeting, where advertisers micro-target previous visitors to their website. Our findings provides a rare empirical view of targeted online advertising *in situ*, and the connection between auction prices and targeting outcomes.

6.1 Introduction

Online advertising is an enormous and complex system, allowing millions of advertisers to reach billions of users across millions of websites, with the capability to target individual users based on their interests, online history, and personal information. On the web, this system is underpinned by a tangled ecosystem of ad tech companies, intermediaries who run the infrastructure for determining which ads are placed on which pages. This model is known as *programmatic advertising*, where for every web page that a user loads, advertisers compete in an automated, real-time bidding auction to determine who gets to place their ads on the page.

The complexity and scale makes it difficult for observers outside of the industry to answer broad and fundamental questions about the online advertising ecosystem. For example: How are different topics of ads targeted? How do advertisers determine the value of an ad? How do factors like behavioral profiles, demographics, and website context affect how users are valued or targeted by advertisers?

Though prior measurement work has provided some answers on these questions, such as work observing the existence of behavioral targeting and retargeting [37, 131, 150, 103], and measurements of winning bid values from real-time bidding and header bidding auctions [156, 162, 158, 44, 159], these studies collect their data through crawler-based experiments, or through field studies with non-representative convenience samples. In the case of crawler studies, statistics like proportions of targeted ads, or bid values, may thus not be representative of what end users actually experience on the web [225, 114]; or in the case of field studies with limited samples, studies may overlook differences in the user population due to demographics or other factors.

In this paper, our goal is to provide accurate measurements of targeting and bid values on the web that reflect what end users observe in the real world. We ask the following **research questions**:

1. How are ads on the web targeted at an individual, demographic, and contextual level?
2. How much do advertisers pay to show ads to people, and how do individual, demographic, and contextual factors affect their bids?
3. How much targeting do users perceive, and do those perceptions relate to bid values?

To measure the influence of individual, demographic, and website factors on targeting and bid values, we based our methodology on the following **measurement goals**:

- *In situ data collection:* To accurately measure behavioral targeting, which is based on browsing histories, we aimed to collect data directly from real users' browsers.
- *Demographic representativeness:* Convenience samples of the population, such as friends and colleagues, or unscreened online participant pools, may have skewed demographics, which affects the generalizability of results. Thus, we aimed to recruit a demographically representative sample of participants in the U.S.
- *Control for differences in websites:* In their daily lives, people likely browse different sets of websites. In a field study, this makes it difficult to compare data between participants. To ensure that data between participants is directly comparable, we aimed to collect data from a fixed set of websites for all participants.
- *Control for changes over time:* Market conditions, advertising campaigns, as well as user behaviors and preferences, may change over time, affecting results from participants who

collect data at different times. Thus, we aimed to collect data from small snapshot in time (about 1 week) to minimize longitudinal changes.

With these goals, we designed a carefully controlled field measurement study of online advertising. First, we recruited a representative sample of 286 U.S. participants, asking for demographic information to allow us to ensure representativeness and answer research questions about demographics. Participants installed a browser extension that collected the content and winning bid values (via header bidding) of the ads shown to them, meaning that the data collected would reflect targeting of participants' actual profiles. We also surveyed participants about the perceived level of targeting of a sample of the ads shown to them. Participants visited the same set of 10 websites, to control for differences in topics, popularity, and trackers across websites. In total, we collected 41,032 ads, including 7,117 with winning bid data.

The **contributions** of our measurements include:

- We provide some of the first empirical measurements of demographic targeting in the wild, showing differences in frequency of certain ads categories like apparel, beauty, education, and careers; across age, gender and ethnicity. We also measure differences in the distribution of ads across websites and individuals.
- We quantify the value of users to advertisers in the wild, using data from header bidding auctions. We observe little to no effect of demographic factors on bid values, but we do find variation in bid values across websites, individuals, ad categories, and ad networks.
- We find that ads with abnormally high winning bid values (up to 16x higher than average) typically promote products which participants report previously viewing, providing additional evidence that high bid values correlate with retargeting.
- Our findings complement and concur with findings from prior work measuring targeting and bid values, confirming in the field the same forms of targeting measured by crawlers,

and adding evidence that bid values are increasing over time.

6.2 Background

We provide background on how ad auctions in programmatic advertising operate, including real time bidding and header bidding. Then, we explain how programmatic ad auctions are the mechanism used to implement targeted advertising.

Real-Time Bidding Real-time bidding is an method for connecting advertisers, who want to buy ads, to publishers, who are selling spaces on their websites. When a user loads a webpage with an ad, a script on the page will contact one of the website's *demand partners* and request an ad. These demand partners are typically *supply side platforms* (SSP) or ad networks, which are entities whose primary purpose is to help websites place ads on their page. Upon receiving a bid request, SSPs will forward the request to an *ad exchange*, which runs an auction where advertisers can bid on the opportunity to run their ad in that slot (usually through another intermediary, i.e. a *demand side platform* (DSP) [214]. The ad that wins the auction is rendered on the user's page, and the advertiser pays the website (and intermediaries) the amount they bid [52]. The value of a bid is typically denoted in CPM, or cost per mille, which means the cost to show 1000 impressions of an ad. For example, a typical bid may be $1.50 CPM, or $0.0015 to show the ad to a single person.

Targeting and Bid Strategies To help decide how much to bid in RTB auctions, bidders are supplied with identifiers for the user, like cookies or fingerprints, which they can use in conjunction with data collected by web trackers and data brokers to find users' interests, browsing behavior, and real world behaviors [214]. Bidders have many strategies for choosing what to target, like targeting visitors of specific websites (contextual targeting) [161], users that appear to be interested in a topic based on past browsing history (behavioral targeting) [41], users that had previously visited their website (remarketing) [214], or people in specific geographical areas (geotargeting) [41]. Determining the exact bid value is an optimization problem where multiple factors are considered to determine the optimal bid value, such as the targeting parameters, budget and strategy of the ad campaign, and how well the ad matches the available information about the website and user [33, 40, 231, 108, 230].

Header Bidding To complicate matters, websites may partner with more than one company to place their ads. Websites can make requests to multiple ad networks or SSPs, like OpenX, Criteo, and Google Ads; or run ads via direct orders (a direct agreement with an advertiser). Each of these demand partners run their own RTB auctions, and offer different bids — and some exchanges may not provide a bid at all [159]. To decide on which demand partner to select for a given ad slot, websites previously used a static priority list, known as "waterfalling" [52], but this approach was not optimal if demand partners farther down the list sometimes offered higher prices.

To optimally decide on which demand partner to pick when filling an ad slot, many websites began using a technique called *header bidding*. Header bidding allows a website to solicit bids from multiple demand partners in parallel, and pick the highest bid from among them. Header bidding auctions often take place in a client-side JavaScript library, such as Prebid.js. Figure 6.1 shows a diagram of a header bidding auction.

Header bidding is advantageous for researchers, because it makes bids transparent. In RTB, bids could be observed through win notifications, but these are increasingly encrypted, making bid prices difficult to measure [162]. Header bidding is typically implemented as a JavaScript library (e.g. Prebid.js), which allows researchers to directly view bid responses by querying the header bidding script using an instrumented browser or browser extension.

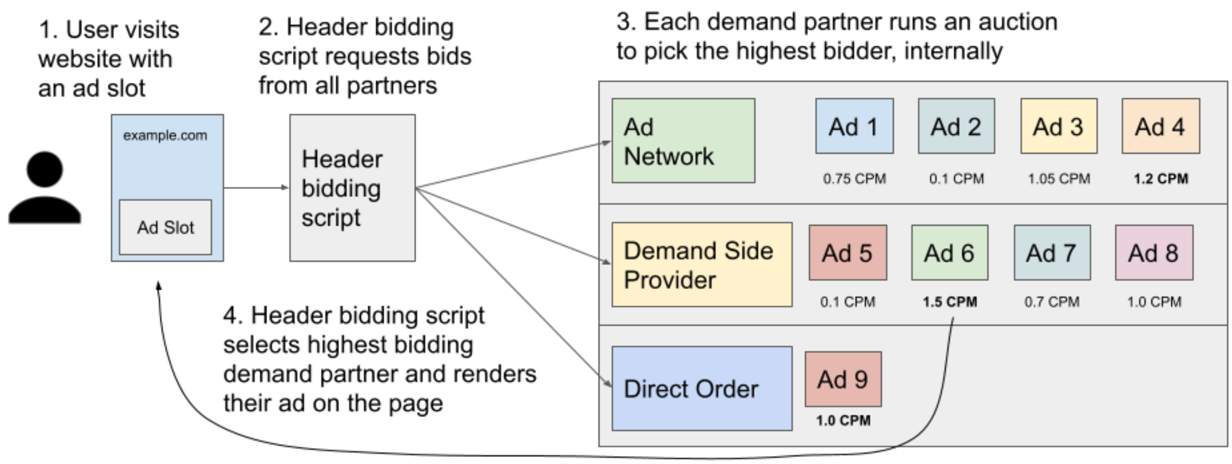


Figure 6.1: A diagram of a header bidding auction.

6.3 Related Work

There is a rich body of measurement research aiming to bring tranparency to the online advertising ecosystem.

Targeting Measurements Prior work has measured targeted ads from a variety of perspectives. Most commonly, studies use web crawlers to detect the presence of different modes of targeting. For example, several studies use synthetic profile or persona based crawls to detect behavioral targeting and contextual targeting. In the absence of having access to browsers with real user profiles, these studies crawl lists of websites intended to signal interest in a certain topic, and compare differences in ads between profiles. Crawler-based targeting studies have found that certain ad categories, and personas are more heavily targeted than others, such as health, travel, and shopping [37, 131, 150]. A similar study using fine-grained targeting detection also found that health ads were highly targeted in Gmail [125]. However, it is unclear how well these crawler profiles match real profiles in terms of how the ad ecosystem treats them [225, 114]

Other crawler-based case studies have examined potentially problematic targeting of specific types of ads, such as gender discrimination in the behavioral targeting of career ads [50], and contextual targeting of misleading political ads on politically partisan websites [228].

Few studies have measured targeting in field studies with real users. Parra-Arnau et al. collected field measurements to validate their targeting detection method, finding that retargeting was common, and that large firms were responsible for most behavioral targeting, but only used a small convenience sample of other researchers and friends [163]. Iordanou et al. developed a privacy-preserving methodology for detecting demographic-based targeting from crowdsourced data from real users, finding that women, older people, and middle income people were more likely to be targeted, but they did not collect data on the content of ads or websites [103].

Our work adds to this literature by investigating the effects of demographics on targeting, and by providing metrics on targeting that more accurately represent the magnitude of targeting experienced by end users in the real world.

Real-Time Bidding and Header Bidding Measurements Prior work has measured multiple aspects of ad auctions through real-time bidding (RTB) and header bidding (HB).

Most closely related to our work, a number of papers have measured bid values to quantify the value of users, and understand the factors that affect bids. Olejnik et al. and Papadopoulos et al. measured bid values from RTB auctions, using data collected from convenience samples of real users. They found that bid prices can be affected by multiple contextual and longitudinal factors, such as time of day and year, country, ad slot sizes, operating system, website category, ad category, and retargeting [156, 162]. Pachilakis et al. replicates this work to measure differences in bid values over a multi-year scale, they found increases in bid values due to cookie syncing, and analyzed the effect of gender and age, but did not obtain a demographically representative sample [158]. Other studies have measured bid values through HB using crawlers, finding differences due to ad slot sizes and crawling profiles [44, 159].

Other studies used bid responses as a mechanism for detecting and measuring other phenomena. Cook et al. utilized bid values from HB to learn tracker-advertiser relationships [44]. Iqbal et al. used header bidding as a signal to detect retargeted ads originating from queries to smart assistants [104]. Other measurements of ad auctions examine performance metrics, such as latency of bid responses and the bidding behaviors of ad networks in the auctions [15, 159, 223].

Our work adds to this literature by providing measurements of HB bid values from a demographically diverse sample of real users, providing insight into demographic effects on bid values, and by separating the effects of other factors such as site, demand partner, and individual variation.

Other Related Work Farther afield, other work has investigated issues with targeted ads on other platforms like Facebook, such as discrimination in ad delivery [9, 102], and targeting of harmful ads [8] and misinformation [174]. Other work has measured the prevalence of web trackers and fingerprinting which enable behavioral targeting on the web [176, 128, 63, 3, 2, 105, 17].

6.4 Field Study Methodology

In this section, we describe the methodology for our field study. As described in Section 6.1, our measurement goals were to collect data from a demographically representative sample of real users, and to control for differences across websites and time. This approach leverages users' existing advertising profiles, ensuring the data matches real world conditions, is representative, and is directly comparable between participants.

6.4.1 Participant Recruitment

We recruited a demographically representative sample of U.S. participants from Prolific. Because online panels are known to have skewed demographics, we used a two-part recruitment method.

Website	Topics	Site Rank
businessinsider.com	National and business news	137
weather.com	Weather forecasts and news	288
speedtest.net	Internet performance test	289
usnews.com	National news, college rankings	365
foodnetwork.com	Recipes and cooking content	1016
detroitnews.com	Local newspaper	2904
ktla.com	Local TV news	4626
phonearena.com	Tech news, smartphone reviews	4954
fashionista.com	Fashion and celebrity news	8773
oxfordlearnersdictionaries.com	Online dictionary	8903

Table 6.1: We selected websites across a range of topics and popularity for study participants to visit.

First we conducted a pre-screening survey, open to all U.S.-based Prolific users, where participants provided their age, gender, and ethnicity, primary browser, and whether they used an ad blocker. Optionally, we asked for participants' sexuality, income, and ZIP code.

Next, we filtered out all respondents except those who used either Google Chrome and Microsoft Edge, for compatibility with our extension, and to control for privacy features in other browsers that could affect participants' advertising profiles. We also filtered out participants who used ad blockers, which could similarly impact their profiles.

Then, we used stratified sampling to select a representative group of participants. We created quotas for each cross-section of the population by age, gender, and ethnicity, based on U.S. demographic data from the 2020 American Community Survey [31], aiming for 300 participants. We invited batches of participants to a second, private Prolific study, until all quotas were filled. However, we excluded 14 participants post-study due to anomalies in their data, e.g. they used an ad blocker, or could not load particular sites.

6.4.2 Study Procedure

Participants selected for the study were directed to our website with a consent form, and instructions to install our browser extension. Upon installing the browser extension, the extension opened a page asking the participant to sign in with their Prolific user ID, followed by an instructions page.

Website List After the instructions, participants were redirected to a page showing a list of 10 websites to scan using our extension (Table 6.1). All participants were asked to visit the same websites to control for contextual targeting, and in randomized order to control for ordering effects. We chose the websites by scanning the top 10,000 websites on the Tranco top sites list, filtering to sites which contained the prebid.js script. Then, we manually evaluated the sites, looking for websites that reliably received bid responses on load, sites spanning a range of topics, and a range of popularity.

Data Collection When a participant visited a site on our list, the extension's content script displayed a modal dialog box, asking them for permission to start a scan. When the scan is initiated, the extension uses CSS selectors from an ad blocker filter list (EasyList) to determine which elements on the page are ad slots.

For each ad, the extension scrolls it into view, and attempts to extract bid metadata from the Prebid.js header bidding script, which is accessible from the global JavaScript context. The extension's content script queries the following APIs: `getBidResponses()` which returns all bids received, `getAllWinningBids()` which returns winning bids for ads which were rendered on the page, and `getAllPrebidWinningBids()` which returns winning bids for ads which won their auction, but the site decided not to run on their page.[1] These calls return bid metadata for all ad slots on the page; so the extension attempts to match metadata items to the ad currently in view, by checking if the `id` of the ad slot's HTML element matches the `adUnitCode` field in each bid response. If bid data was matched for the ad, the extension then

[1] A reason why an ad could win a header bidding auction, but not appear on the page, is that the site has another demand partner that takes precedence over the header bidding result (i.e. waterfall prioritization [52])

takes a screenshot of the ad (stored locally) and sends the header bidding data to the study server. If a bid cannot be matched to an ad, then the ad is skipped.

After scanning all ads, the extension automatically refreshes the page and collects a second run of data, to increase the sample size of ads collected per site and participant.

Targeting Perceptions Survey

After visiting all 10 websites, participants were redirected to a survey, where participants rated how targeted they felt by the ads collected. The extension draws a deterministic sample of 8 ads to show the participant; by ranking the ads by winning bid value, and selecting ads at uniform intervals from the lowest to highest value ad. We chose this over random sampling to guarantee that the sample contained ads with a range of bid values.

For each ad in the sample, we asked the participant four questions about their perceptions of the targeting of the ad:

1. *(Relevance)* “How relevant is this ad to your interests?” (1-5 Scale)
2. *(Targeting)* “How personalized or targeted is this ad to you?” (1-5 Scale)
3. *(Likeliness to Click)* “How likely would you be to click on this ad?” (1-5 Scale)
4. *(Retargeting)* “Have you ever previously clicked on this ad, viewed the product or website featured in the ad, or bought the product in the ad?” (Yes/No/Not Sure)

Data Exclusion

Lastly, we provided a chance for participants to remove any screenshots of ads which they felt might be sensitive, e.g. if they felt that the ad was targeted and the screenshot would reveal unwanted information to us, the researchers. Participants were shown all of the ads we collected (and stored locally), and selected the ones they did not want to upload to our server.

6.4.3 Labeling Ad Categories

To enabled analysis of targeting, we assigned ads to categories using a mix of automated and manual approaches.

First, we used a topic model to automatically place ads into semantically similar clusters. We first used the Google Cloud Vision API to extract text from ad screenshots. We then used locality sensitive hashing to deduplicate ads. Then, we used the BERTopic topic modeling library [94], which combines several algorithms: the all-MiniLM-L12-v2 language model for generating embeddings, UMAP for dimensionality reduction, and HDBScan for clustering. We also evaluated other topic modeling algorithms, like LDA and GSDMM, but found that BERTopic produced the most qualitatively coherent topics. The topic model produced 311 topics.

We then manually audited the topics, finding overlapping topics, misclassified ads, and generally too many topics for analysis. We manually combined similar topics together into 52 categories of products, such as "medications", "home kitchen and bathroom products", and "electronics". We manually verified each category and moved misclassified ads.

Some ads were not assigned a category, either because the ad was blank, cut off by a popup, or in the middle of loading when the screenshot was taken, or because multiple ads were captured in the image, and we could not determine which ad the header bidding data corresponded to. These ads are excluded from our analysis.

6.4.4 Limitations

We can only approximate detection of behavioral targeting, because we do not have ground truth for the interests that advertisers inferred about participants. Participants' demographics are somewhat of a proxy for advertising interests, but do not encompass all of the possible variation.

We selected a limited set of 10 websites, to control for websites as a variable, and to keep the duration of the study short. However, the small sample size means that certain results may be specific to the sites chosen, such as the overall counts of ads by category, or the overall average bid values.

The sample size of ads with winning bids was smaller than expected, with only 7117 ads. In some cases, we lack the statistical power for certain advanced analyses, such as interactions between factors. For example, we did not have the sample size to analyze an interaction effect between ads categories and a demographic characteristic of a participant, when predicting bid values.

The time period when the ads were collected was approximately 1-2 weeks before Christmas in the United States, which may have had an effect on bids. Bid values may have been higher than usual, due to high demand for advertising during the Christmas shopping season.

6.4.5 Ethics

Our study was approved by our institutional review board, which determined that the study qualified for exempt status (Category 3).

Participants agreed to a consent form explaining the possible risks of the study before starting. Participants were compensated $0.25 for completing the pre-screening survey, and $8.00 for completing the browser extension study, a rate $15.00 per hour by our initial estimates for completion time. Some participants took much longer than expected due to technical issues; in these cases we provided bonus payments to compensate them for the additional time.

We took into consideration users' privacy and safety in multiple aspects of the design of our study and browser extension.

First, we designed the extension to require user input and consent before collecting data: rather than immediately taking control of the browser like a crawler, participants manually visited each

Gender	Female					*F-All*	Male					*M-All*	Non-binary		*NB-All*	*All*
Age	18-24	25-34	35-44	45-54	55+		18-24	25-34	35-44	45-54	55+		25-34	35-44		
Ethnicity																
Asian or Pacific Islander	2.45	1.05	0.35	0.35	0.00	*4.20*	2.45	2.10	1.05	1.05	0.00	*6.64*	0.00	0.00	*0.00*	10.84
Black or African American	1.75	2.10	1.40	0.70	0.35	*6.29*	0.35	1.75	1.40	0.35	0.00	*3.85*	0.00	0.00	*0.00*	10.14
Hispanic or Latino	4.90	1.40	1.75	0.00	0.00	*8.04*	1.05	2.10	0.35	0.00	0.70	*4.20*	0.35	0.00	*0.35*	12.59
Other	0.00	2.10	0.35	0.35	0.00	*2.80*	1.40	0.70	0.00	0.35	0.00	*2.45*	0.00	0.35	*0.35*	5.59
White or Caucasian	6.99	5.59	7.69	4.55	6.99	*31.82*	2.10	6.99	9.09	5.24	5.59	*29.02*	0.00	0.00	*0.00*	60.84
All	16.08	12.24	11.54	5.94	7.34	*53.15*	7.34	13.64	11.89	6.99	6.29	*46.15*	0.35	0.35	*0.70*	100.00

Table 6.2: Demographics of the 286 participants in our study. All values above are percentages.

site on our list. Then, upon opening a page on the list, the extension asked for permission to start scanning before starting the data collection procedure. For websites not on the list, the content script would not execute at all, meaning participants could continue using the site.

Second, we were aware that screenshots of ads could inadvertently expose information about participants, if the ads were targeted and revealed something sensitive that the did not want to share. Thus, we added an interface before the screenshots were uploaded to us, where participants could exclude any screenshots that they found too sensitive.

Third, we provided clear instructions for participants to remove the extension at the conclusion of our study, but the extension did not continue to collect any data if participants forgot to remove it.

6.5 Results

6.5.1 Dataset Description

Participant Demographics

In total, 286 participants successfully completed data collection for our study. Table 6.2 shows a summary of the demographic data of our study participants. Our dataset roughly approximates the U.S. population, but skews slightly younger and female. 267 participants used Google Chrome while 19 used Microsoft Edge.

Ads Overview

We collected 41,032 ads in total, or an average of 143.5 ads per participant, from 20 page loads each.

We were able to extract the winning bid in 25,764 of ads where a header bidding auction took place. Only in 7,117 ads of these ads was the winner actually rendered on the page — websites can choose not to use the winner of the header bidding auction, and instead choose an ad from another ad network to fill the slot instead.

Through topic modeling and manual qualitative analysis, we generated 52 categories describing the content of ads (see Section 6.4.3). We were able to assign categories to 35,681 ads, 5,351 ads were not assigned a category. Of the rendered winning bids, which we analyze in greater detail later, 5,851 out of 7,117 ads, or 82%, were assigned a category. Ads may not have been assigned categories if we detected anomalies (ads where popups or the extension UI accidentally covered the ad in the screenshot), if the ad was not fully loaded at screenshot time, or if multiple ads were in the screenshot.

In the study, we analyze four overlapping subsets of data:

- *Ads with categories* (35,681 ads). This subset contains the ads which we were able to assign a category to, either manually or automatically. We examine this subset in Section 6.5.2, where we analyze how the categories are distributed across demographics and sites.
- *Ads with rendered winning bids* (7,117 ads). These are ads for which we obtained the winning bid amount, and confirmed that the ad was rendered on the page. We examine this subset in Section 6.5.3.
- *Ads with user targeting perceptions* (1,744 ads). These are the ads which participants rated with their perceptions of targeting, and is a strict subset of the above subset. We examine targeting perceptions in Section 6.5.4.

- *Ads with non-rendered winning bids* (18,916 ads). Ads for which we have a winning bid amount, but the HB winner was not rendered on the page. We briefly discuss this subset in Section 6.5.1, but do not use this data for other analyses, because the screenshots captured do not correspond to the bid response.

Overall Winning Bid Values Averaged $5.47 per Thousand Impressions

How much did advertisers bid to show ads on the 10 sites in our dataset? The average winning bid had a mean value of $5.47 and median of $4.16 (IQR=$4.43). However, not all ads that won their header bidding auctions were rendered on the page. For ads where the header bidding winner was not rendered, the mean bid value was $3.60 CPM, and the median was $2.62 CPM (IQR = $3.25). Figure 6.2 shows the cumulative distribution functions for winning bids, separating ads that were rendered versus not rendered.

Though most bids won with a value less than $10, there is a substantial long tail of outliers. The top 10% most expensive winning bids were $10.62 CPM or above, and the top winning bid was $89.7 CPM, or nearly $0.09 to show a single ad. We perform a brief case study of these outliers in Section 6.5.5.

Summary of Ad Categories

Next, we summarize the categories of ad by content. Figure 6.3 shows the number of ads collected in each category, in the subset of all ads with a category (35,681 ads). Ads spanned a large variety of products, ranging from apparel, to home goods, and medications. The most common ads were for electronics (smartphones, computers, accessories), business ads (cloud computing, marketing services, office supplies, etc.), banking and finance ads (ads for mortgages, banks, investments), mixed native ads (a.k.a. content recommendation networks), and travel ads. Other notable categories specific to the time period when the measurements were conducted include

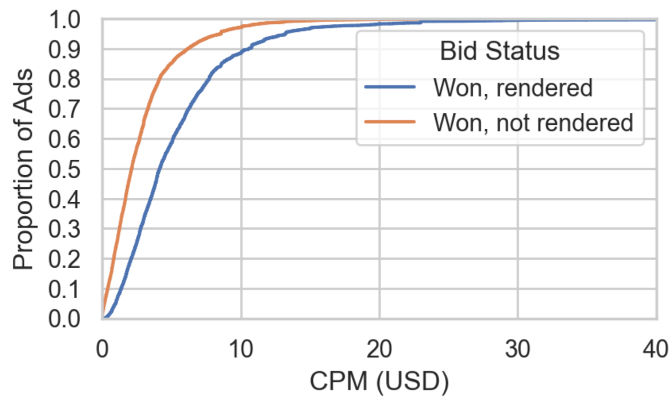


Figure 6.2: Cumulative distribution function (CDF) of all winning bid values in our dataset. Winning bid values for ads that were actually rendered on the page were higher than those that were not rendered.

COVID-19 related ads for vaccines, tests, and PSAs; and holiday-specific ads, such Christmas cards, gift wrap, and holiday sales (measurements were conducted in December, proximate to Christmas and other winter holidays in the United States).

Note that this distribution of ads by category is biased by the 10 sites we selected for the study; a different configuration of sites may result in a different category distribution. We discuss contextual targeting more in Section 6.5.2. We also observe some differences in the categories of ads in the subset with winning bid data, compared to the subset without bid data — see Appendix 6.5.1 for details.

Header Bidding Ad Categories

We compare the proportion of ads in each category between the subset of ads with rendered winning bids, and all other ads in Table 6.3. We find that the proportions of certain categories differ substantially while others are approximately equivalent. For example the rendered winning bid dataset has more medication ads (7.24% vs 1.8%), about the same number of banking and finance ads (5.72% vs. 6.23%), and substantially fewer career (0.48% vs. 3.83%) and native ad widgets (0.09%

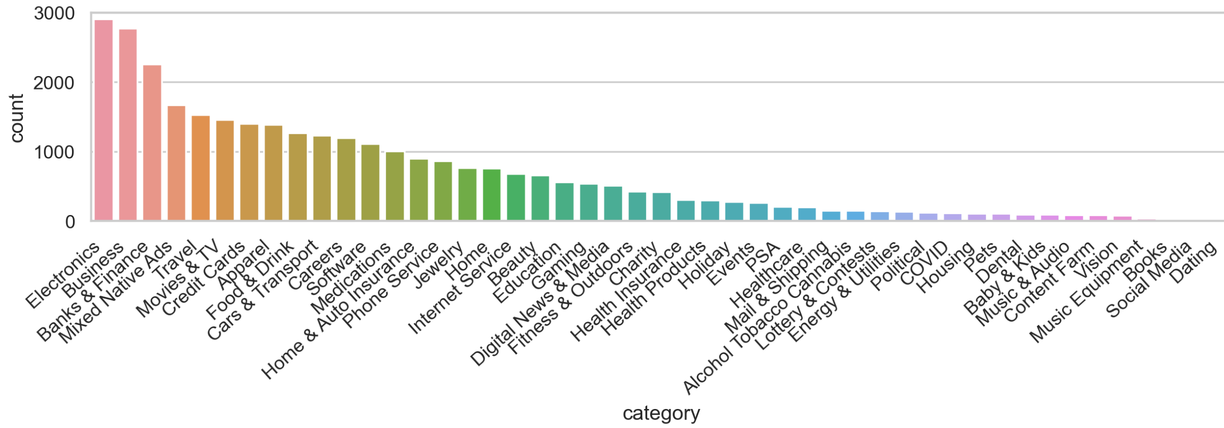


Figure 6.3: The number of ads in our dataset by category, including ads without winning bids associated with them.

vs. 5.47%). This suggests that the demand partners that advertisers prioritize over header bidding may have qualitatively different ad campaigns in their inventory than the demand partners in header bidding auctions.

6.5.2 How were ads targeted?

Next, we measure how ads are targeted by comparing the frequency of ads in each category, across demographic groups, websites, and individuals. For demographic and contextual factors of interest, we conducted an omnibus chi-square test of independence, to determine whether there is a significant association between ad category and the factor of interest. We adjusted the resulting p-values for multiple comparisons using the Bonferroni method. To identify which categories were more or less common than expected (based on the overall proportions of ads by category across the dataset) we calculated the standardized residuals (a measure of the difference between the observed and expected cell value), and conduct a post-hoc Z-test, with critical values adjusted with the Bonferonni method. For individuals, we use distributional inequality metrics to characterize how each category of ad is distributed across individuals.

	% of Ads with Winning Bid	% of All Other Ads	Residuals
Medications	7.24	1.80	24.19
Internet Service	3.50	1.50	10.78
Food & Drink	5.13	3.10	8.05
Apparel	5.09	3.49	6.10
Cars & Transport	4.45	3.11	5.42
Movies & TV	4.58	3.83	2.79
Banks & Finance	5.72	6.23	-1.54
Business	6.69	7.72	-2.84
Credit Cards	2.69	4.06	-5.21
Software	1.36	3.39	-8.60
Electronics	5.20	8.49	-8.84
Travel	2.05	4.59	-9.28
Careers	0.48	3.83	-13.73
Mixed Native Ads	0.09	5.47	-18.79

Table 6.3: Difference in the size of ad categories between data subsets (15 largest categories shown). The residuals column shows the standardized residuals between the two subsets; residuals larger than ±3.28 indicate significant differences ($p < 0.05$).

Strong Evidence of Contextual (Website-based) Targeting

We find that some categories are more common on specific websites than others, usually when the topic of the ad is relevant to the topic of the website — evidence of contextual targeting. A chi-squared test of independence found a significant association between website and category ($\chi^2(423, N = 31,407) = 37,155.82, p < 0.001$). Post-hoc Z-tests on the adjusted residuals indicated that 202 of 470 residuals exceeded the critical value of 3.70 ($p < 0.05$), indicating that a large number of the categories were over- or under-represented on specific sites.

Table 6.4 shows the percentage of ads from each category on each website, for the 24 most common categories overall. Qualitatively, we find that categories that are more common than expected (in bold) are often related to the website. For example, ads in the “education” category, which contain ads for college programs and online classes, are much more common on usnews.com (11.24%), a website best known for its college rankings. speedtest.net, a tool for measuring internet speeds, had a high percentage of ads for gaming (14.6%) and internet service (20.7%); two topics where bandwidth is important. Business ads, which include marketing services and cloud software, were common on businessinsider.com (25.47%), a business news site.

Category	businessinsider.com	detroitnews.com	fashionista.com	foodnetwork.com	ktla.com	oxfordlearnersdictionaries.com	phonearena.com	speedtest.net	usnews.com	weather.com
Apparel	*1.11*	**6.13**	6.11	**8.56**	**6.36**	**9.30**	*0.78*	*0.74*	*2.87*	5.50
Banks & Finance	6.01	7.54	*1.76*	*4.11*	*5.11*	5.85	*1.27*	*3.74*	**28.46**	7.88
Beauty	*0.66*	**3.20**	2.81	2.49	**3.09**	2.28	*0.49*	2.58	2.58	2.23
Business	**25.47**	*5.44*	7.59	*4.00*	*5.46*	**14.85**	*2.82*	10.80	*5.35*	*5.50*
Careers	**21.36**	*0.33*	*0.50*	*0.63*	*0.73*	*0.92*	*0.21*	*0.18*	*2.26*	*1.07*
Cars & Transport	4.11	**5.65**	*1.27*	4.32	**7.99**	3.39	*1.62*	*0.86*	3.42	3.83
Charity	*0.39*	1.43	**3.69**	0.88	1.70	1.66	*0.38*	0.43	**2.44**	1.75
Credit Cards	**13.16**	3.70	6.16	*1.58*	4.66	3.88	*1.22*	*2.27*	*2.15*	*2.67*
Digital News & Media	2.23	*0.19*	**17.55**	*0.53*	*0.10*	*0.00*	*0.07*	*0.06*	*0.73*	*0.92*
Education	*0.49*	*0.81*	*0.33*	*0.70*	1.22	2.53	*0.45*	1.04	**11.24**	1.28
Electronics	*0.98*	*3.70*	8.86	10.67	8.86	*5.61*	**35.49**	8.10	*2.66*	*3.93*
Fitness & Outdoors	*0.13*	*0.69*	0.44	**8.32**	1.18	0.37	*0.71*	*0.31*	1.24	0.85
Food & Drink	*1.07*	*2.76*	4.51	**9.72**	**6.53**	4.19	*1.15*	4.05	2.87	**6.74**
Gaming	*0.30*	*0.96*	0.99	1.89	*0.73*	2.28	*0.33*	**12.94**	1.24	2.28
Home	*0.26*	2.04	1.71	**5.51**	**4.00**	2.83	*0.85*	2.70	2.04	**3.97**
Home & Auto Insurance	*0.94*	*2.04*	*1.16*	3.93	**6.05**	**4.50**	*1.48*	*0.43*	2.40	**5.91**
Internet Service	*0.17*	*0.27*	1.60	*0.84*	*0.73*	*0.49*	*0.64*	**18.47**	**4.11**	**3.30**
Jewelry	**4.09**	*0.60*	**8.58**	**4.32**	2.36	1.11	*0.19*	*0.86*	2.00	2.57
Medications	*0.13*	**6.96**	2.75	2.42	*1.08*	2.16	*1.60*	2.33	2.26	**7.63**
Mixed Native Ads	*0.19*	**19.42**	*0.00*	*0.07*	*0.17*	*0.06*	**9.04**	*0.06*	*0.00*	**8.02**
Movies & TV	*2.98*	**6.73**	**6.99**	3.82	*1.43*	**9.37**	3.74	**10.00**	*2.00*	4.53
Phone Service	*0.39*	*0.23*	2.81	*1.23*	*1.46*	1.48	**13.74**	*1.29*	*0.80*	*1.38*
Software	*0.60*	*1.91*	*1.05*	*0.67*	*1.11*	**6.96**	**14.29**	**7.24**	*1.20*	*1.33*
Travel	**9.70**	*1.81*	*2.42*	**8.98**	**12.83**	3.94	*0.59*	*1.66*	*3.09*	*2.81*

Table 6.4: Percent of ads observed on each website from each category (top 24 categories only). Blue/bold cells indicate a significantly higher proportion than expected, and red/italic cells indicate a significantly lower proportion than expected, based on post-hoc Z-tests on the standardized residuals. Darker colors indicate larger differences.

Gender	Female	Male	Non-binary
Apparel	**5.39**	*3.30*	2.74
Banks & Finance	6.78	7.65	8.22
Beauty	**2.64**	*1.48*	0.91
Business	8.96	8.66	9.13
Careers	3.80	3.78	5.94
Cars & Transport	3.64	4.26	2.74
Charity	1.38	1.28	0.46
Credit Cards	4.50	4.42	5.94
Digital News & Media	*1.38*	**1.90**	3.20
Education	1.74	1.85	1.37
Electronics	8.84	9.71	13.24
Fitness & Outdoors	1.32	1.42	0.91
Food & Drink	4.17	3.92	2.28
Gaming	*1.31*	**2.17**	4.57
Home	2.66	2.13	2.28
Home & Auto Insurance	2.83	2.95	1.37
Internet Service	1.97	2.42	0.00
Jewelry	2.65	2.20	2.74
Medications	3.52	2.89	0.91
Mixed Native Ads	5.27	5.33	6.85
Movies & TV	4.43	4.86	5.48
Phone Service	*2.43*	**3.11**	4.11
Software	3.38	3.78	2.28
Travel	4.69	5.11	1.83

Table 6.5: The percentage of ads shown to people of each gender from each category, for the top 24 categories. Blue / bolded cells indicate a significantly higher proportion than expected, and red / italic cells indicate a significantly lower proportion than expected.

Targeting by Gender

We saw differences in the number of ads seen between genders in a small number of categories. A chi-squared test of independence found a significant association between gender and category ($\chi^2(92, N = 31,407) = 425.72, p < 0.001$). Post-hoc Z-tests on the adjusted residuals indicated that 12 of 72 residuals exceeded the critical value of 3.39 ($p < 0.05$). Table 6.5 shows the percentage of ads by category. We found that women tend to receive more ads for Apparel and Beauty, while men tended to receive more ads for Gaming, Digital News, and Phone Service. We did not have a large enough sample of non-binary participants to find significant differences.

Category	Asian	Black	Latino	Other	White
Apparel	4.51	3.29	4.68	*2.45*	4.71
Banks & Finance	7.03	7.42	6.86	7.30	7.23
Beauty	1.40	**3.20**	**3.00**	1.95	1.88
Business	8.49	7.55	8.83	9.53	9.02
Careers	3.20	3.04	4.31	4.52	3.87
Cars & Transport	5.02	4.09	3.10	4.01	3.85
Charity	1.07	1.12	0.76	0.72	**1.57**
Credit Cards	4.69	4.32	**5.70**	4.40	4.23
Digital News & Media	2.01	1.47	1.50	1.73	1.61
Education	**3.32**	1.98	1.50	1.78	*1.56*
Electronics	10.50	9.92	9.25	8.70	9.01
Fitness & Outdoors	1.55	1.41	1.18	1.06	1.39
Food & Drink	3.41	4.22	3.47	3.51	4.28
Gaming	1.37	1.86	1.45	**3.29**	1.68
Home	2.34	1.57	1.76	2.06	**2.73**
Home & Auto Insurance	2.71	3.55	2.34	3.40	2.85
Internet Service	2.83	2.56	1.94	1.84	2.06
Jewelry	1.67	**3.77**	2.00	2.23	2.47
Medications	*2.16*	2.56	3.94	2.56	3.41
Mixed Native Ads	4.96	5.44	5.49	5.85	5.26
Movies & TV	5.24	5.60	5.44	5.07	*4.18*
Phone Service	2.95	2.50	3.00	3.40	2.66
Software	2.65	3.65	3.36	2.73	3.80
Travel	6.00	4.64	4.89	**6.97**	*4.50*

Table 6.6: The percentage of ads shown to people of each ethnicity from each category, for the top 24 categories. Blue / bolded cells indicate a significantly higher proportion than expected, and red / italic cells indicate a significantly lower proportion than expected.

Targeting by Ethnicity

We saw significant differences in the number of ads seen between ethnicities in a small number of categories. A chi-squared test of independence found a significant association between ethnicity and category ($\chi^2(184, N = 31,407) = 690.03, p < 0.001$). Post-hoc Z-tests on the adjusted residuals indicated that 23 of 235 residuals exceeded the critical value of 3.52. Table 6.6 shows the percentage of ads by category shown to people by ethnicity. Among the significant examples, Black and Latino participants were shown more Beauty ads, Latino participants were shown more Credit Card ads, White participants were shown more Charity and Home ads, and Asian participants were shown more Education ads.

Age Range	18-24	25-34	35-44	45-54	55+
Apparel	**5.46**	*3.54*	3.75	5.11	4.96
Banks & Finance	6.46	7.65	6.85	6.84	8.44
Beauty	2.51	2.00	2.34	1.80	1.40
Business	8.49	8.70	8.89	9.17	9.21
Careers	*2.89*	3.43	**4.58**	4.28	4.28
Cars & Transport	3.81	3.96	3.73	3.93	4.33
Charity	*0.88*	1.38	1.55	1.25	1.67
Credit Cards	4.82	4.58	4.12	4.74	4.06
Digital News & Media	1.56	1.72	1.80	1.20	1.67
Education	1.93	1.79	2.13	1.35	1.33
Electronics	9.50	9.51	9.47	8.82	8.42
Fitness & Outdoors	1.17	1.40	1.39	1.33	1.60
Food & Drink	3.60	**5.00**	3.83	3.76	3.53
Gaming	1.57	1.92	**2.30**	1.38	*0.89*
Home	2.24	2.11	2.27	**3.23**	2.83
Home & Auto Insurance	2.47	2.98	3.33	2.66	2.71
Internet Service	1.89	2.35	2.08	1.50	**3.07**
Jewelry	2.85	2.00	*1.78*	**3.78**	2.59
Medications	3.45	2.85	3.46	3.93	2.39
Mixed Native Ads	5.17	5.27	5.51	5.26	5.32
Movies & TV	5.31	5.24	4.43	*3.36*	3.85
Phone Service	2.93	2.61	2.91	1.95	3.24
Software	3.45	3.92	2.95	3.58	4.09
Travel	**5.95**	4.27	4.98	5.26	*3.56*

Table 6.7: The percentage of ads shown to an age group from each category, for the top 24 categories. Blue / bolded cells indicate a significantly higher proportion than expected, and red / italic cells indicate a significantly lower proportion than expected.

Targeting by Age

We saw differences in the number of ads seen across age ranges in a small number of categories. A chi-squared test of independence found a significant association between gender and category ($\chi^2(184, N = 31,407) = 735.93, p < 0.001$). Post-hoc Z-tests on the adjusted residuals indicated that 20 of 235 residuals exceeded the critical value of 3.52 ($p < 0.05$). Table 6.7 shows the percentage of ads by category, across age ranges. 18-24 year olds saw more ads for apparel and travel, and fewer for careers, 25-34 year olds saw more ads for food and drink, 35-44 year olds saw more ads for careers, 45-54 year olds saw more ads for jewlery, and 55+ year olds saw more ads for internet service.

Individual Targeting

Next, we characterize the amount of variation in ads seen by individuals, due to possible behavioral targeting. Theoretically, if there are no differences in the ads seen by different people visiting the same sites, we would expect equal quantities of ads from each category in our study. However, with the presence of individual targeting, a few participants may account for a large proportion of the ads in a category.

Figure 6.4 shows Lorenz curves for each ad category, which describe the level of distributional inequality [122] in who sees ads from each category. If a category of ads were distributed equally across participants, the line would be diagonal; the lower the curve, the more unequally the ads are distributed.

We find that ad categories had varying levels of distributional disparities. Some ads, like Mixed Native Ads, and Electronics ads, were shown roughly equally: the top 5% of participants saw 7.4% and 11% of the ads in those categories (if totally equal, the top 5% would have seen 5% of ads). On the other hand, ads for Charity ads and Fitness ads were much more unequally distributed; the

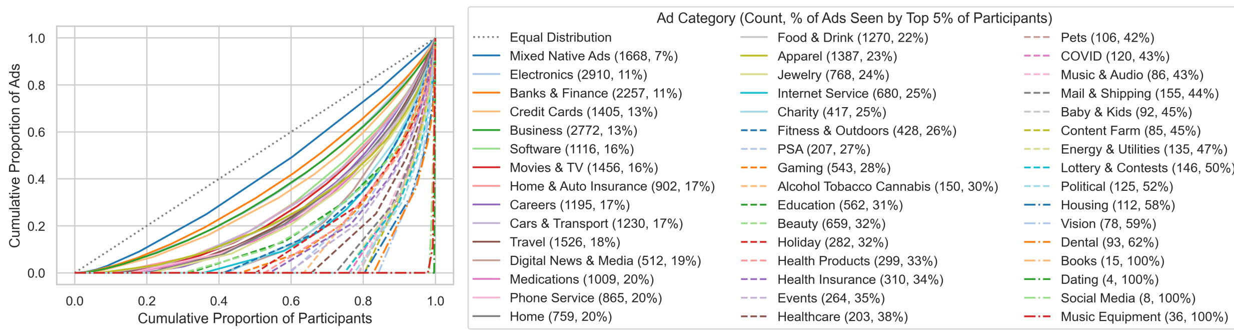


Figure 6.4: Lorenz curve showing the cumulative fraction of ads as a function of the cumulative fraction of participants, for each category. Curves closer to the diagonal line represent ad categories that are more evenly distributed across participants.

top 5% of participants saw 24.7% and 26% of ads respectively. Though ads that were more common overall were generally more evenly distributed, this was not a perfect correlation: Apparel ads were less evenly distributed than Movies & TV (23% vs. 16% shown to the top 5% of participants), even though both categories contained around 1400 ads.

Category	Website	Top 5% on Site	Top 5% Overall
Business	businessinsider.com	14.22	13.06
Careers	businessinsider.com	13.45	16.85
Electronics	phonearena.com	13.45	11.31
Phone Service	phonearena.com	16.78	19.63
Education	usnews.com	17.44	31.39
Banks & Finance	usnews.com	9.97	11.43
Internet Service	speedtest.net	13.67	24.53

Table 6.8: Percent of ads seen by the top 5% of participants, for select categories and website combinations that we identified as contextually targeted, compared to all websites. Ads were distributed more equally when looking specifically at the site than overall, suggesting that behavioral targeting does not amplify differences seen in contextual targeting.

We also investigate whether behavioral targeting at the individual level might amplify contextual targeting. In Table 6.8, we compare the percent of ads seen by the top 5% of participants in contextually targeted categories on specific sites, with the percent of ads seen by the top 5% participants over the whole dataset. We find that within websites, ads likely to be contextually targeted were distributed *more* equally than in the overall dataset. Thus, in our sample, we do not see evidence of behavioral-contextual amplification.

6.5.3 What influences winning bid values?

In an ad auction, bidders consider many factors to determine the value of the ad, including the user's inferred interests, demographics, the website the ad appears on, and the targeting and budget parameters of the ads. To estimate the influence of each of these factors on bid values simultaneously, we used a linear mixed effects model to predict rendered winning bid values (response variable) as a function of the user's age, gender, and ethnicity (fixed effects/explanatory variables), as well as the website the ads appeared on, the bidder, the individual, and the category of the ad (random effects).

We selected our model using the top-down method suggested by Zuur et al. [232]: we started with a full specified model, including all of the above fixed and random effects, as well as other optional demographics we collected (sexuality, income, and children), and other labels we generated, such as whether ads used a native format, and labels based on our contextual targeting results. We did not include interaction effects, like gender and ad category, because we did not have enough data to estimate the number of parameters. We then experimented with removing random effects and fixed effects to improve the fit of models, using the REML Akaike information criterion (AIC) (when removing random effects) and maximum likelihood AIC (when removing fixed effects) to measure the goodness of fit. Our final model included all random effects but only included age, gender, and ethnicity as fixed effects. The final model's REML criterion was 42141.3. We show the raw regression estimates in Table 6.9.

Fixed Effects				
Effect	Estimate	Std. Error	t	Pr(>|t|)
Intercept	5.903	0.889	6.639	>0.000***
Age	0.003	0.010	0.291	0.771
Gender — Male	-0.582	0.250	-2.324	0.021*
Gender — Nonbinary	-2.068	1.716	-1.205	0.229
Ethnicity — Asian	-0.231	0.427	-0.542	0.588
Ethnicity — Black	0.315	0.424	0.743	0.458
Ethnicity — Latino	0.746	0.392	1.900	0.059
Ethnicity — Other	0.807	0.567	1.424	0.156
Random Effects				
Groups	Effect	Variance	Std.Dev.	
Website	Intercept	3.748	1.936	
Bidder	Intercept	3.639	1.908	
Participant	Intercept	3.255	1.807	
Ad Category	Intercept	1.719	1.311	
Residual		19.977	4.470	

Table 6.9: Fixed effects estimates and random effects structures for a linear mixed model with winning bid values as the outcome variable, fixed effects of age, gender and ethnicity, and random effects for website, individuals, bidder, and ad category. p-values estimated via t-tests using the Satterthwaite approximations to degrees of freedom. Male participants received slightly lower winning bid values (-$0.58 CPM). 38% of the variance that demographics did not account for are explained by variation in websites, bidders, individual participants, and ad categories, though 62% of the variance remains unexplained.

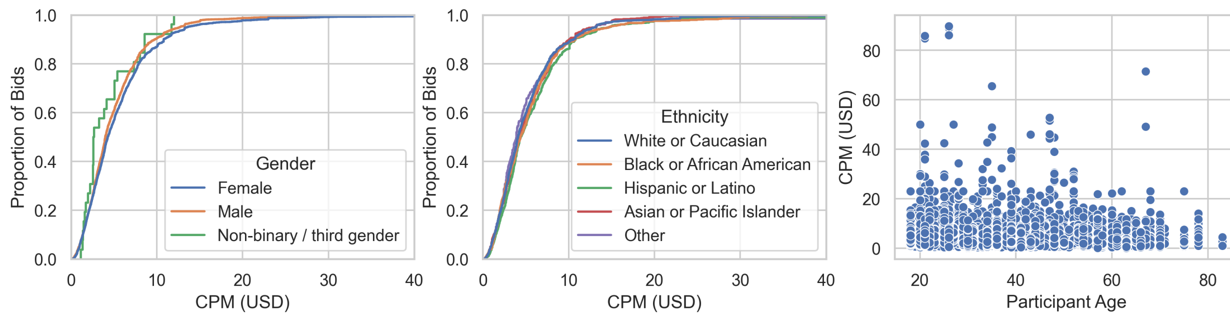


Figure 6.5: Distribution of bid values across gender, ethnicity, and age. Demographic factors explained little of the differences in bid values; we only detected a significant effect of gender on bid value, with an estimated difference of \$0.58 CPM between women and men.

Demographics: Advertisers Bid Slightly Higher for Women

Overall, we did not see that rendered winning bid values were strongly affected by demographic factors. Bid values for male participants were estimated to be \$0.58 CPM lower than women. However, we did not detect any effect of age or ethnicity on bid values. A linear mixed model analysis of variance indicated a statistically significant effect on bid values of gender ($F(2, 329) = 3.25, p = 0.040$) but no statistically significant effect of ethnicity ($F(4, 277) = 1.589, n.s.$) or age ($F(1, 281) = 0.085, n.s.$). We also did not detect an effect of optional demographic factors (sexuality, income, children) on bid values; these variables did not improve the fit of the model, and were excluded from the final analysis. Figure 6.5 shows cumulative distribution functions for bid values by gender and ethnicity, and a scatter plot of age and bid values.

This finding suggests that in the online advertising markets, no particular demographic groups are in substantially higher or lower demand than others, overall. However, this does not mean that people are not being targeted by demographics. A likelier explanation is that there is relatively even demand in the market to sell to people of all ages, ethnicities, and genders, and advertiser demand for one demographic group may be canceled by demand for another.

Individual Variation: Winning Bids Differed Between Participants

Though all participants visited the same set of websites, the same number of times, the mean value of the bids seen by each participant ranged from as low as $1.15, and as high as $17.35. The median of the mean bid value for each participant ranged was $4.96 (IQR = 2.34). Participants' median bid values were slightly lower than the mean; the median of the median values was $4.39 (IQR = 2.35), indicating that outliers skewed means upwards. The mixed model predicts a slightly smaller amount of variation than the raw averages (by controlling for other factors): the median random intercept for participant was -$0.23, with an IQR of $1.61. The variance of the participant random effect was 3.266, which explains 10.1% of the variance in the model.

Website: Winning Bid Values Differed Across Websites

Among the 10 websites in our study, we found differences in the winning bid values. Table 6.10 shows the average winning bid values for each domain. For example, we saw that speedtest.net had the highest mean winning bid at $6.43 CPM, while ktla.com had the lowest at $1.11 CPM. Mixed model estimates for the effect of website range from $3.66 to -$2.62. The variance of the website random effect was 3.748, accounting for 11.6% of the total variance. Higher winning bids did not appear to correlate with site rank; for example, phonearena.com had the 7th highest site rank, but the 2nd highest mean winning bid value.

These results suggest that some sites are in higher demand from advertisers than others. Perhaps certain sites signal greater intent to certain types of advertisers; e.g. phonearea.com may have higher demand from smartphone manufacturers and wireless carriers because visitors are more likely to purchase their products, while news sites like ktla.com may provide little information to most advertisers.

	Mean	Std.Dev.	# Ads	Estimate
Website				
speedtest.net	9.95	6.07	508	3.66
businessinsider.com	7.95	6.09	289	2.34
phonearena.com	7.87	3.42	313	0.84
foodnetwork.com	6.03	6.11	873	0.57
weather.com	5.39	5.28	834	-0.17
oxfordlearnersdictionaries.com	5.40	5.75	671	-0.22
fashionista.com	4.88	5.50	369	-1.29
usnews.com	3.83	3.29	589	-1.50
detroitnews.com	4.97	4.96	2033	-1.60
ktla.com	2.44	1.68	638	-2.62
Ad Category (Top 25)				
Medications	6.95	3.17	463	1.14
Beauty	7.27	9.83	184	1.12
Health Insurance	6.37	10.54	73	1.12
Gaming	5.40	6.94	67	0.93
Holiday	6.31	6.47	64	0.67
Jewelry	6.70	6.32	83	0.48
Business	5.80	7.04	428	0.36
Internet Service	6.18	5.95	224	0.29
Banks & Finance	4.19	2.96	366	-0.05
Home	4.63	5.01	177	-0.05
Cars & Transport	5.53	4.03	285	-0.09
Movies & TV	6.43	5.98	293	-0.14
Health Products	4.89	3.14	46	-0.22
Phone Service	6.33	3.96	135	-0.25
Software	4.66	4.68	87	-0.31
Travel	4.94	3.45	131	-0.34
Electronics	5.19	7.61	333	-0.36
Credit Cards	4.92	4.09	172	-0.37
Home & Auto Insurance	4.10	2.88	167	-0.38
Education	4.05	3.58	84	-0.55
Healthcare	3.86	3.70	49	-0.78
Alcohol Tobacco Cannabis	4.21	2.16	70	-0.80
Food & Drink	4.41	3.99	328	-0.86
Apparel	4.90	3.64	326	-0.87
Charity	2.99	2.56	69	-1.89
Demand Partner				
consumable	18.04	20.92	12	5.27
trustx	9.42	12.57	133	3.90
districtm	11.29	7.35	31	1.19
appnexus	7.38	6.62	791	1.15
colossusssp	5.53	6.99	36	0.81
aol	10.85	6.59	25	0.75
pubmatic	7.12	7.25	718	0.55
rubicon	5.87	5.25	684	0.48
sonobi	5.34	2.64	215	0.09
teads	3.71	2.06	376	-0.13
criteo	7.53	5.40	123	-0.22
openx	5.31	3.62	284	-0.55
verizon	2.41	1.29	173	-0.65
kargo	4.51	2.13	157	-0.67
ix	5.05	4.79	803	-0.75
onemobile	4.46	3.29	323	-1.09
pulsepoint	2.43	2.35	30	-1.57
triplelift	3.36	3.47	765	-1.99
medianet	5.38	3.18	132	-2.08
nobid	6.76	2.56	28	-2.33

Table 6.10: Summary of winning bid values by website, ad category, and demand partner. We show the mean and standard deviation of bid values, the number of ads in the group, and the estimated difference from the predicted baseline bid (random intercept).

Bidders: Winning Bid Values Differed Across Demand Partners

Bid values varied between the demand partners: the ad networks, supply side providers, or other entities placing the bid on the behalf of the advertiser. Table 6.10 shows the average winning bid values for each demand partner. Based on estimated intercepts from the mixed model that control for other factors, the highest bidding demand partners were Consumable (mean bid value of $18.04), TrustX ($9.42), and District M ($11.29), while the lowest bidders were NoBid ($6.76), MediaNet ($5.38), and TripleLift ($3.36).

To understand the potential underlying reasons for these differences, we investigated the public facing websites of these bidders. Though many made similar claims about the power and reach of their technology, we noticed some qualitative differences. The highest bidders (Consumable and TrustX), focused their message on "premium" content and advertisers, and improving users' experience, meaning they likely work with higher profile websites and brands, involving higher budgets. The lowest bidders (NoBid and MediaNet), described their products in terms of "maximizing revenue" and filling "unfilled and undervalued inventory", suggesting that their strategy is to win auctions where demand is lowest, and bidding at low amounts.

Ad Categories: Winning Bids Differed Across Ad Categories

How did bid values vary for different categories of ads? Table 6.10 summarizes winning bid price for ads of each category. The ads with the highest bid values came from the "mail & shipping" category, which included US Postal Service ads and home delivery services ($13.03), beauty ($7.27), and medications ($6.95). Categories with low values included charity ($2.99), healthcare ($3.86), and live events ($3.04). However, the size of the categories suggest that some differences may be due to outlier bids. For example, the mail and shipping category contains 32 ads, and two outliers with winning bid values over $80, and a standard deviation of 19.61, which suggests that

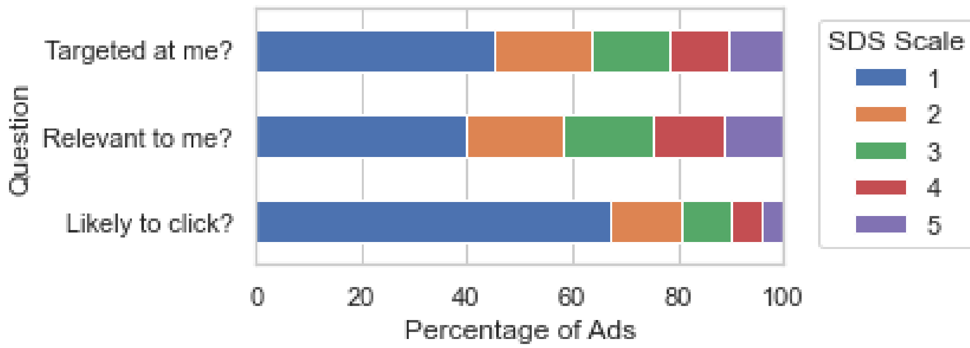


Figure 6.6: Summary of the targeting perceptions survey. Responses are on a semantic differential scale (i.e., 1 means “not relevant at all”, 5 means “very relevant”). Participants said that a majority of ads were not relevant to them, targeted at them, and that they were unlikely to click on them.

the presence of outliers in a small sample is skewing the overall figures.

6.5.4 Self-Reported Targeting Perceptions

What proportion of ads did participants themselves perceive as targeted? In this section, we report on results of the self reported targeting perceptions survey. We also investigate whether targeting perceptions correlate with bid values.

Each participant rated a sample of 8 ads that they saw with their perceptions of how targeted each was. We used a deterministic sample of 8 ads with winning bids, uniformly selected across the range of bid values, to ensure we had data on high and low bids for each participant. We received responses for 1746 ads from 286 participants, an average of 6.1 per participant. Some participants were not able to submit responses for all 8 ads for several possible reasons: because the ad screenshots were blank or obscured (215 participants, affecting 449 ads), because they did not receive 8 rendered winning bids in total (16 participants, affecting 61 ads), or because of other unknown technical issues with the extension (32 ads).

Most Ads Were Not Relevant to Participants

Figure 6.6 shows the distribution of participants' responses to the targeting perceptions survey. Most ads were perceived as not relevant to participants: over 40% of ads received the lowest score of 1 for relevance, targeting and click likelihood, while 10% or less scored the highest score of 5. Comparing the distributions for each question, participants perceived ads as relevant and targeted at similar proportions, but were less likely to click on ads. We also asked participants whether they had previously visited the website of the advertiser or product, which could indicate if the ad was retargeted. Participants responded "Yes" for 18.3% of ads, "No" for 76.6% of ads, and "Not Sure" for 5% of ads. We expected a somewhat even distribution to these responses, because an even number of ads with low and high bid values were sampled, these results still skew towards low relevance, indicating that participants did not perceive much targeting broadly.

Self-Reported Retargeted Ads had Higher Winning Bid Values

Next, we investigate whether participants' targeting perceptions correlate with winning bid values. To determine which factors may be related to bid values, we fit a linear mixed effects model to the subset of 1746 ads with survey responses. Winning bid price was the outcome variable, with fixed effects for perceptions of relevance, targeting, likeliness to click, and retargeting. Additionally, we include the fixed and random effects from the final model in Section 6.5.3: fixed effects of age, gender, and ethnicity, random intercepts for website, participant, bidder, and ad category. Coefficient estimates are reported in Table 6.11.

Ads where participants reported previously visiting the advertiser's site had a median CPM of \$4.50 (IQR = \$5.08), and ads not perceived as retargeted had a median of \$3.90 (IQR = \$4.32). A linear mixed model analysis of variance found a statistically significant effect of self-reported visits on winning bid value ($F(2, 1645) = 6.064$, $p = 0.002$), with an estimated increase of \$1.07 for ads with

Fixed Effects				
Effect	Estimate	Std. Error	t	Pr(>|t|)
Intercept	5.998	1.058	5.668	<0.001***
Age	-0.008	0.012	-0.676	0.500
Gender — Male	-0.391	0.300	-1.306	0.193
Gender — Nonbinary	-1.857	2.368	-0.784	0.433
Ethnicity — Asian	-0.483	0.510	-0.947	0.345
Ethnicity — Black	0.321	0.505	0.636	0.525
Ethnicity — Latino	1.213	0.484	2.506	0.013*
Ethnicity — Other	-0.707	0.667	-1.060	0.290
Retargeted — Yes	1.074	0.386	2.783	0.005**
Retargeted — Not Sure	1.447	0.566	2.557	0.011*
Perceived Relevance	-0.073	0.163	-0.451	0.652
Perceived Targeting	0.239	0.163	1.468	0.142
Likely to Click	-0.212	0.157	-1.354	0.176
Random Effects				
Groups	Effect	Variance	Std.Dev.	
Website	Intercept	5.891	2.427	
Bidder Code	Intercept	2.587	1.608	
Participant	Intercept	2.307	1.519	
Ad Category	Intercept	1.530	1.237	
Residual		22.583	4.752	

Table 6.11: Fixed effects estimates and random effects structures for a linear mixed model with winning bid values as the outcome variable, and fixed effects of demographic factors and targeting perceptions. p-values estimated via t-tests using the Satterthwaite approximations to degrees of freedom. Ads that participants perceived as retargeted had higher winning bid values (+$1.36 CPM).

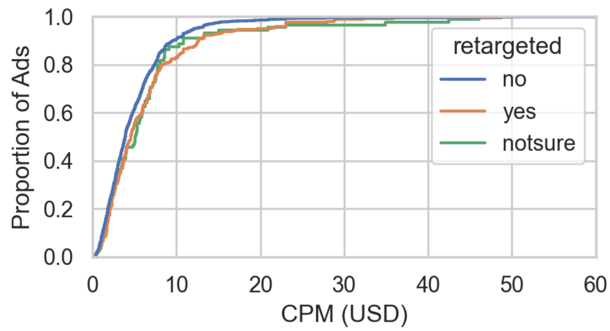


Figure 6.7: CDF of winning bid values, for ads that participants self reported as retargeted or not retargeted. In aggregate, retargeted ads had higher bid values than non-retargeted ads.

"yes" responses, and $1.45 for "not sure". However, no effect was detected for perceived targeting, relevance, and likelihood to click. Figure 6.7 shows the CDF for bid values, across participants' responses to whether they visited the advertiser's site. These findings concur with the findings of Olejnik et al., who found in a crawler-based study that retargeted ads had substantially higher bid values [156].

6.5.5 Case Study: Extreme Outliers in Bid Values

Though the average bid value was $3.55 CPM, we observed many examples of bids an order of magnitude higher, as high as $89.00 CPM. What explains these extremely high bids? In this section, we perform a case study of the ads that we observed in this range, to try to understand what may explain these bid prices. We examine the subset of ads with a winning bid values greater than $20 CPM, which encompasses 127 ads, or the top 1.8% of ads by bid value. This subset of ads came from 66 participants.

Outliers were distributed among individuals roughly evenly; the data is not dominated by one or more individuals. The mean number of outliers seen by an individual was 1.9, 92% participants saw 1-3 outliers, making up 81% of the data, and five participants had 8, 7, 5, 4, and 4 ads.

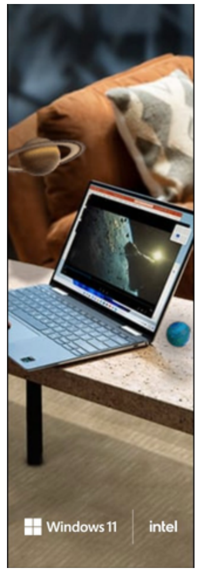


Figure 6.8: Ad seen by Participant 639 two times. This is the highest valued ad in our dataset, at $89.75 CPM, or almost $0.09 for a single impression.

Individual Examples We start by looking at some examples from individual participants, to illustrate exactly what these outliers look like.

Participant 639 had the highest bid values in the dataset, with two ads with bids of $89.75 and $89.09 CPM each. Both ads appeared on oxfordlearnersdictionaries.com, and were ads from Microsoft for Intel-based laptops with Windows 11.

Participant 719 had 7 ads in the outlier subset, with values ranging from $44.96 to $65.57 CPM. All ads were for the same product — a perfume from Yves Saint Laurent — and all appeared on detroitnews.com (Figure 6.9). The participant reported that the ad was something they visited the website for previously, and responded with the maximum score for targeting perception, relevance, and likeliness to click. These pieces of evidence strongly suggest that these ads were targeted at the particular individual.

Participant 535 had four ads with bids ranging from $44.31 to $52.80, all appearing on food-network.com, and all from SurveyMonkey, an online survey platform. The participant reported

Figure 6.9: Ad seen by Participant 719 seven times, with bid values $44.96-$65.57 CPM.

that they hadn't been to the SurveyMonkey site before, and only rated it with a 2 for perceived relevance, targeting, and likeliness to click.

Participant 414 had four ads with bids ranging from $21.74 to $31.00. All ads were from Jewelry Television, a TV channel specializing in selling jewelry, three appearing on businessinsider.com and one appearing on speedtest.net. The participant reported going to this site in the past, and scored the relevance, targeting, and likeliness to click 4, 5, and 4.

Targeting Survey Responses to Outliers Participants perceived the ads in this subset to be more targeted than the remainder of the dataset, but not overwhelmingly so. 40 of 127 ads had relevance survey responses. The average SDS scores for relevance, targeting perception, and likelihood to click were 2.65, 2.65, and 1.73 respectively, compared to 2.36, 2.22, and 1.66 for all other ads (scores range from 1-5). 40.0% of participants said they had visited the website of the ad previously, compared to 14.1% for ads outside this subset.

Though these values suggest that ads with significantly higher bid value are more likely to be perceived to be targeted by participants than others, around half ads in the dataset are still not seen as targeted. Because the data is self-reported, we cannot know for sure whether this is because the ads were not targeting individuals, or if they were simply poorly targeted for their actual interests.

Repeat Ads In many cases, the same advertiser would show multiple high-value ads to the same participant. We manually inspected the advertiser of these ads, and found that 24 of 33 participants who received more than one outlier received multiple ads from the same advertiser. Often times, these repeat ads appeared on the same website.

Demographics The subset of participants in the outlier subset were skewed younger and more female than the overall sample of participants. 66% of participants were female; 60% were white; 30% were aged 18-24, 29% were aged 25-34, 21% were aged 35-44, 11% were aged 45-54, and 9% were aged 55+.

Website and Ad Category Outliers appeared on some sites more than others. weather.com, speedtest.net, and detroitnews.com, hosted 39, 33, and 18 ads each, while fashionista.com, phonearena.com, and usnews.com only hosted 2 ads each. Outliers cover a range of topics: for example, beauty (7), business (11), electronics (6) gaming (3), health insurance (4), home (8) movies and TV (12, the maximum), etc. No particular category is notably overrepresented.

Demand Partner We observed Pubmatic and Rubicon had substantially more ads in the outlier subset than all other demand partners. 43 ads were from Rubicon, (34%), and 40 were from Pubmatic (31%). The remaining demand partners had between 1 and 9 ads in the subset. This suggests that these two demand partners are more aggressive in their bidding strategies.

6.6 Discussion and Conclusion

6.6.1 Summary: What Factors Affect Bids and Targeting in the Wild?

Evidence of All Forms of Targeting We measured all common types of targeting on the web: contextual targeting, behavioral targeting, demographic targeting, and retargeting, all occuring in concert. We found that contextual targeting is extremely clear on some specific websites, such as a high proportion of electronics ads on a smartphone review site. We also saw evidence suggesting behavioral targeting through differences between demographics in categories, like differences in the number of apparel ads seen across genders, through unequal distributions of ad categories across the participant pool, and through self-reported observations of retargeting by participants.

Behavioral and Contextual Factors Explain Bid Values Our analysis shows that variation in winning bid values results from many factors. Demographic factors have either small or no effect on bid values; we only detected a small effect of gender on bid values. The website the ad appears on, the demand partner placing the bid, and the category of the ad each have an impact on the bid value, but a large amount of variance is either explained by individual variation in participants or remains unexplained. One clear signal we found was that ads that participants perceived as retargeted had substantially higher bids. Thus, we conclude that bid values can be partially described as the combination of the average observed bid value for the website, ad network, and ad category, and the remainder of the variance can be attributed to behavioral targeting.

6.6.2 Comparison to Prior Work

Value of a User Our study finds higher winning bid values than past studies on real-time bidding and header bidding. We observed a median winning bid value of around $4.16 CPM, which is higher than prior work. Prior work from 2019-2020 measured median bids in header bidding ranging from <$0.10 CPM [159]), to $2.00 CPM [44] (both using crawlers). RTB studies also found lower bids, ranging from $0.36 CPM [156] (2013) to $0.273 CPM [162] (2017). Some methodological factors may explain these differences: the ten sites in our study were relatively high ranked, demand for ads may have been high during our study, due to the December holiday shopping season, and bid values for real users with extensive browsing profiles may be higher than for synthetic profiles or stateless crawls. We also speculate that bid prices are rising over time, which concurs with other recent measurements [158].

Differences in Bid Values We concur with other results finding that women receive higher bids than men overall, but did not observe statistically significant effect of age [158]. Our finding that self-reported retargeting was associated with substantially higher bids aligns with other studies finding a link between previous visits to sites and higher bid values [159, 156]. Our results on the average bid values of different demand partners differed in rank order differed from the header bidding study of Pachilakis et al. [159], suggesting that bidding behaviors of individual advertisers may not be stable over time or specific collection methodologies.

6.6.3 Implications and Future Work

How Users Are Valued, and Implications for Privacy What is it ultimately that advertisers "value" about a user? Our results suggest that the signals that affect value most strongly are the host website, and past visits to the advertiser's site. This suggests that these factors communicate intent to purchase most reliably to an advertiser. This also suggests that more general forms of behavioral targeting (e.g. inferred interests) may not be as valuable as a signal to advertisers. If this is the case, then it suggests that privacy-enhancing proposals to limit cross-site tracking, such as removing third-party cookies, may have smaller-than-expected costs for advertisers. A potential approach to reforming web privacy that would strongly improve user privacy while minimally impacting advertisers could be to implement privacy-preserving mechanisms to allow advertisers to continue retargeting (e.g., Google Chrome's FLEDGE proposal [54]), but to remove third-party cookies and mitigate fingerprinting without replacing them with alternatives (e.g., FLoC/Topics [55]).

Ad Quality and Bid Values What is the economic model behind low-quality, misleading, or other ads that are bad for user experience? Prior work [226] has shown that such ads are common, especially on news websites. Our work does not address this question directly, as we did not find many examples of such ads with header bidding metadata. Though our data suggests that some SSPs, like NoBid, specialize in filling cheap, low-demand ad slots, we have no data on the incentives for low quality ads themselves. Do they mainly fill low-demand ad slots? Or do they outbid other ads? Future work may require mechanisms besides header bidding to measure the value of these ads.

In conclusion, our work presents a controlled field study of ad targeting and pricing data with real users in the wild, helping shed light on the inner workings of the ubiquitous yet opaque online advertising ecosystem.

Chapter 7

Conclusion

In this conclusion, I will summarize the main contributions of the work in this dissertation, the challenges and lessons from doing this research, and unanswered questions and future work.

7.1 Contributions

7.1.1 General Findings on the Advertising Ecosystem

First, this work provides new insights into several aspects of the online advertising ecosystem, including users' perceptions of the content of online ads, and empirical data on targeting and bidding practices of advertisers.

Chapter 3 presents a taxonomy of people's positive and negative reactions to ads, which provides insight into how bad ad content harms user experience. My hope is that the taxonomy will be broadly useful for helping researchers classify ads and other content that negatively impacts user experience, and will provide a framework for drawing up guidelines or content moderation policies for organizations interested in ensuring quality user experiences with advertising and other user generated content.

Chapter 6 presents new empirical data on how ads are targeted, and how advertisers value users, which provides insight into the economics and practices of the ad tech industry. In somewhat surprising results, we find that demographic factors, such as age, gender, and ethnicity, have a relatively small impact on targeting and bid values, compared to contextual and individual factors, such as the website ads appear on, and whether users previously shopped for a product. Though these findings are encouraging in that the data does not show broad-based demographic disparities in advertiser behavior, it leaves the door open to problematic targeting or bidding practices at the individual level. These results help provide more accurate measurements of how the online advertising ecosystem operates, and their incentives.

7.1.2 Problematic Practices in Online Advertising

Second, this work identifies specific case studies of problematic practices of online advertisers: the use of native advertising on news websites, and misleading political advertising.

Chapter 4 provides quantitative evidence to back up anecdotal observations that native advertising is responsible for a higher amount of deceptive ads than standard display ads. This study also provides evidence that native advertising is common across all types of news websites and misinformation websites, not just the low quality sites, and that it is an endemic problem to the digital media industry. We recommend that native ad networks either make serious changes to their content policies, that websites stop using these networks, or that regulatory action is taken against these kinds of deceptions.

Chapter 5 delivers new results on how political ads are targeted, and identifies new types of misleading political advertising used in the 2020 Elections. We find that political ads are contextually targeted at more partisan sites, including misleading ones like the political poll ads meant to harvest email addresses. This suggests that political advertisers are target people who consume a lot of partisan political content, especially for ads whose purpose is to get voter data or collect fundraising. We also identify types of misleading content in political ads; such as the deceptive poll ads, political clickbait ads, and products using political content. Our results suggest the need for more aggressive regulatory action against deceptive political advertising, more expansive criteria for political advertising at ad platforms, and the need for additional transparency for political ads on the web.

7.1.3 Methods for Auditing Online Advertising

Lastly, this work presents a suite of methods for auditing the practices of online advertisers.

Chapters 4 and 5 describe a crawler-based method for detecting differences in ads and ad-serving infrastructure between websites, providing insight into contextual targeting and choices made by websites. In these projects, I developed an ad crawling infrastructure based on the Puppeteer browser automation library, focused on scraping content from ads (compared to other measurement tools like OpenWPM that focus on measuring web trackers [63]). Crawls for each domain are run in a separate, stateless Docker container, ensuring that no browser history or state is shared between sites. This enables analyses that look for differences in the types of ads between sites, such as differences in the number of ads in a particular topic, or differences in the ad platforms used by each site.

Chapter 3 demonstrates methods for qualitative analysis of subjective labels and perceptions of online ads (and other content) in the presence of disagreement. In this work, an explicit goal was to embracing the potential diversity of opinions that people may have about ads. To this end, I collected responses from multiple participants per ad in my dataset. To analyze this data, I employed Population Label Distribution Learning to preserve the subjectivity in the dataset, and perform clustering while considering disagreements among participants. This enabled a ground-up approach to identifying classes of problematic ad content based on subjective perceptions, even without agreement and definitive labels.

Chapter 5 and 6 demonstrate the use of language models for large scale clustering and classification of ad content. Most of the findings described above depend on accurate labels for the content of the ads. Given the sizes of these datasets, manual labeling is often impractical. However, modern language models can dramatically speed up labeling, while still providing high quality outputs. We found success using a BERT-based classifier to identify political ads based on text extracted from ad images, and sentence-level models such as MiniLM for unsupervised clustering and topic modeling for labeling ad topics.

7.2 Recommendations for Improving the Online Advertising Ecosystem

Regulations and Content Moderation This dissertation has surfaced many instances of online ads that are deceptive, misleading, potentially harmful, and detrimental to users' experience. These ads often exploit gray areas in the rules and policies set by ad platforms, by coming close to, but not directly violating guidelines for acceptable content. Thus, I recommend that ad platforms increase standards for harmful advertising experiences in their content policies, and ensure that content moderation processes are sufficient to prevent these ads from reaching users.

I also believe there is the possibility for more aggressive regulatory action on deceptive ads. For example, in Chapter 5, I found numerous examples of ads who used misleading polls and other tactics to trick people into giving money or the personal information to political campaigns and other actors. These kinds of practices may fall under the FTC's authority to regulate unfair and deceptive practices, without impinging on the rights of advertisers and political campaigns to free speech.

Transparency and External Auditing The responsibility for reforming and improving the online advertising ecosystem ultimately rests with policy and engineering teams at ad platforms, like Google, Facebook, Taboola, Pubmatic, and all of the other intermediaries involved in delivering ads. These companies have the power to create policies for what kinds of ads acceptable, and have the access and resources to investigate and enforce them. In fact, many of the methods and findings described in this dissertation may have been used and discovered internally at these companies. However, their incentives are not always aligned in the interests of users and the web; financial incentives from customers and shareholders may override concerns about "bad" ads, or lead to these concerns to be deprioritized.

Thus, to provide an external check on the practices of ad platforms, it is critical that researchers are able to access data on online advertising, so that they are able to conduct independent audits of their practices. Though this work, and other related work, have studied aspects of this ecosystem by using scrapers and by directly purchasing ads, there is a substantial amount that remains unknown about online ads, particularly when it comes to how ads are targeted, and bad ads outside of the web platform (such as ads on TikTok or Instagram). While advertisers claim to practice self-regulation, allowing researchers much more detailed access to data, and *not* actively preventing researchers from doing their work [25], would enable more meaningful independent oversight of the advertising industry. And further cooperation on this front could lead to more proactive efforts to eliminate problematic practices, such as scams or algorithmic discrimination.

7.3 Future Work

Automated Identification of "Bad Ads" First, I hope to extend the conceptual contributions on characterizing bad ads towards automatically and proactively identifying "bad ads". Many of the contributions in this dissertation depend on manual labeling and qualitative analysis, limiting the scope of much of this research. However, state-of-the-art large language models such as GPT-3 and BERT promise strong performance on tasks such as text classification. These models could be used to detect characteristics of "bad" ads identified in this work, which could enable measurement studies of bad ads at a dramatically larger scale. This would allow truly internet-scale studies of deceptive advertising, such as studies spanning tens of thousands of websites, long-running longitudinal studies, and field studies with large participant pools. Studies at this scale could provide more precise answers questions about how "bad" ads are targeted, where they appear, and exactly how prevalent they are. These language-model based approaches to studying bad ads could potentially be extended to other platforms, such as social media ads and video ads, through transfer learning.

Economics of "Bad Ads" Second, I hope future work can address targeting and economics of "bad ads" in greater depth. In Chapter 6, there were a number of limitations in our data collection methods that prevented us from collected a large enough sample size of misleading advertising (including their bid values), such as the limited set of websites and the types of ads present in header bidding auctions.

However, with different methods, it would be interesting to study the targeting and bidding strategies of advertisers who run "bad" ads. For example, are these ads targeted at specific demographics or behaviors? If so, this could be a problematic and predatory practice that regulators like the FTC could address through enforcement actions. And do these advertisers tend bid high or low values, compared to other ads? Answering these questions could help identify what types of websites or other features attract problematic ads. This could inform changes that ad platforms and websites could make to prevent exploitation by dishonest advertisers, and potential modifications to ad auction algorithms that could reduce the prevalence of these ads on the web, in the vein of other work on algorithmic fairness in ad auctions [38, 209].

Ad Transparency Tools for Users Lastly, I think it would be useful to develop user-facing tools to surface some of the phenomena studied in this dissertation to people in their daily lives. A browser extension, or the browser itself, could display information about the presence, deceptiveness, or targeting of a particular ad. For example, this tool could visually highlight or warn users about deceptively formatted ads such as native ads, to make it clear to users which content is paid for by an advertiser. A text summarization model could be used to help users interpret and understand the claims made by advertisers, and potentially help them learn about and avoid scams or other deceptions [78]. And certain phenomena in header bidding, such as abnormally high bid values, could be used to inform users when an ad is likely being targeted at them through a retargeting campaign.

From a research perspective, it would be interesting to study whether interventions like this improve people's understandings of targeted advertising, or deception in advertising (such as prior work on transparency for web tracking [217]). From an impact perspective, it would be exciting to see if finished products such as this would build broader support among the public for greater regulation and transparency for online advertising.

7.4 Closing Thoughts

I hope that my work has laid the groundwork for better criteria and methods for defining and identifying “bad” ads, enabling better policies and enforcement against deceptive and misleading advertising in the future. I am optimistic that this line of research will lead to better ads and a better web, because the tools available to researchers have improved substantially in the past few years. Rapid improvements in tools such as large language models and measurement infrastructures will allow researchers to detect and quantify problematic ads at larger scales than ever before. Meanwhile, I believe that the techniques used by online advertisers to deceive and mislead users are not improving or evolving nearly as fast, as they primarily rely on social engineering tactics that have not substantially changed over time. My ultimate hope is that research, moderation, and regulation of online ads will create an environment where ads respect people’s privacy, security, and browsing experiences, allowing the web to remain free and sustainable.

Bibliography

[1] AARP. Dietary supplement scams, December 2019. https://www.aarp.org/money/scams-fraud/info-2019/supplements.html.

[2] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, page 674–689, New York, NY, USA, 2014. Association for Computing Machinery.

[3] Gunes Acar, Marc Juarez, Nick Nikiforakis, Claudia Diaz, Seda Gürses, Frank Piessens, and Bart Preneel. Fpdetective: Dusting the web for fingerprinters. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS '13, page 1129–1140, New York, NY, USA, 2013. Association for Computing Machinery.

[4] Google Ads. Reach a larger or new audience with google display network targeting. https://ads.google.com/intl/en_id/home/resources/reach-larger-new-audiences/, 2022.

[5] Lalit Agarwal, Nisheeth Shrivastava, Sharad Jaiswal, and Saurabh Panjwani. Do not embarrass: Re-examining user concerns for online tracking and advertising. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, New York, NY, USA, 2013. Association for Computing Machinery.

[6] Alexa Web Information Service. Alexa - Top Sites by Category - Top/News/Alternative. https://www.alexa.com/topsites/category/Top/News/Alternative. Accessed on 2019-07-24.

[7] AlgoAware. An overview of the Programmatic Advertising Ecosystem: Opportunities and Challenges. https://platformobservatory.eu/app/uploads/2019/10/An-overview-of-the-Programmatic-Advertising-Ecosystem-Opportunities-and-Challenges.pdf, june 2019. Observatory on the Online Platform Economy.

[8] Muhammad Ali, Angelica Goetzen, Piotr Sapiezynski, Elissa Redmiles, and Alan Mislove. All things unequal: Measuring disparity of potentially harmful ads on facebook. In *6th Workshop on Technology and Consumer Protection (ConPro '22)*. IEEE, 2022.

[9] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), November 2019.

[10] Digital Citizens Alliance. Digital bait: How content theft sites and malware are exploited by cybercriminals to hack into internet users' computers and personal data. https://www.digitalcitizensalliance.org/clientuploads/directory/Reports/digitalbait.pdf, 2015.

[11] AllSides. Balanced news via media bias ratings for an unbiased news perspective. https://www.allsides.com/unbiased-balanced-news.

[12] Michelle A. Amazeen and Bartosz W. Wojdynski. Reducing native advertising deception: Revisiting the antecedents and consequences of persuasion knowledge in digital news contexts. *Mass Communication and Society*, 22(2):222–247, 2019.

[13] Amazon. Alexa Web Information Service API. https://awis.alexa.com/.

[14] Mimi An. Why People Block Ads (And What It Means for Marketers and Advertisers. https://web.archive.org/web/20200831221906/https://www.upa.it/static/upload/why/why-people-block-ads/why-people-block-ads.pdf, 2016. HubSpot Research.

[15] Waqar Aqeel, Debopam Bhattacherjee, Balakrishnan Chandrasekaran, P. Brighten Godfrey, Gregory Laughlin, Bruce Maggs, and Ankit Singla. Untangling header bidding lore. In Anna Sperotto, Alberto Dainotti, and Burkhard Stiller, editors, *Passive and Active Measurement*, pages 280–297. Springer International Publishing, 2020.

[16] Anocha Aribarg and Eric M. Schwartz. Native advertising in online news: Trade-offs among clicks, brand recognition, and website trustworthiness. *Journal of Marketing Research*, 57(1):20–34, 2020.

[17] Pouneh Nikkhah Bahrami, Umar Iqbal, and Zubair Shafiq. Fp-radar: Longitudinal measurement and early detection of browser fingerprinting. *Proceedings on Privacy Enhancing Technologies*, 2022(2):557–577, 2022.

[18] Samantha Baker. How GOP-linked PR firms use Google's ad platform to harvest email addresses. Engadget, November 2019. https://www.engadget.com/2019-11-11-google-political-ads-polls-email-collection.html.

[19] Andrew O. Ballard, D. Sunshine Hillygus, and Tobias Konitzer. Campaigning Online: Web Display Ads in the 2012 Presidential Campaign. *PS: Political Science & Politics*, 49(3):414–419, 2016.

[20] Ballotpedia. Presidential election, 2020. https://ballotpedia.org/Presidential_election,_2020.

[21] Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson. Tracing Information Flows between Ad Exchanges Using Retargeted Ads. In *Proceedings of the 25th USENIX Conference on Security Symposium*, SEC'16, page 481–496, USA, 2016. USENIX Association.

[22] Muhammad Ahmad Bashir, Sajjad Arshad, and Christo Wilson. Recommended For You: A First Look at Content Recommendation Networks. In *Proceedings of the Internet Measurement Conference (IMC 2016)*, Santa Monica, CA, November 2016.

[23] Chris Bentzel. Under the hood: How chrome's ad filtering works. https://blog.chromium.org/2018/02/how-chromes-ad-filtering-works.html, Feb 2018. Chromium Blog.

[24] Andrew Blustein. Publishers May Lower Their Standards on Ad Quality in Search of Revenue. https://www.adweek.com/programmatic/publishers-may-lower-their-standards-on-ad-quality-in-search-of-revenue/, March 2020. Adweek.

[25] Shannon Bond. Nyu researchers were studying disinformation on facebook. the company cut them off. https://www.npr.org/2021/08/04/1024791053/facebook-boots-nyu-disinformation-researchers-off-its-platform-and-critics-cry-f, August 2021. NPR.

[26] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. Tales from the dark side: Privacy dark strategies and privacy dark patterns. *Proceedings on Privacy Enhancing Technologies*, 2016(4):237–254, jul 2016.

[27] Giorgio Brajnik and Silvia Gabrielli. A review of online advertising effects on the user experience. *International Journal of Human–Computer Interaction*, 26(10):971–997, 2010.

[28] Joshua A. Braun and Jessica L. Eklund. Fake news, real money: Ad tech platforms, profit-driven hoaxes, and the business of journalism. *Digital Journalism*, 7(1):1–21, 2019.

[29] K. A. Braun and E. Loftus. Advertising's misinformation effect. *Applied Cognitive Psychology*, 12(6), 1998.

[30] Harry Brignull. Dark patterns, 2019. https://www.darkpatterns.org/.

[31] United States Census Bureau. American community survey (acs). https://www.census.gov/programs-surveys/acs, 2020.

[32] Moira Burke, Anthony Hornof, Erik Nilsen, and Nicholas Gorman. High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. *ACM Transactions on Computer-Human Interaction*, 12(4):423–445, December 2005.

[33] Han Cai, Kan Ren, Weinan Zhang, Kleanthis Malialis, Jun Wang, Yong Yu, and Defeng Guo. Real-time bidding by reinforcement learning in display advertising. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, page 661–670, New York, NY, USA, 2017. Association for Computing Machinery.

[34] Colin Campbell and Pamela E. Grimm. The Challenges Native Advertising Poses: Exploring Potential Federal Trade Commission Responses and Identifying Research Needs. *Journal of Public Policy & Marketing*, 38(1):110–123, 2019.

[35] Margaret C. Campbell. When attention-getting advertising tactics elicit consumer inferences of manipulative intent: The importance of balancing benefits and investments. *Journal of Consumer Psychology*, 4(3), 1995.

[36] Ashley Carr. Low quality websites: Content farms: What is a content farm?, 2021. https://researchguides.austincc.edu/contentfarms.

[37] Juan Miguel Carrascosa, Jakub Mikians, Ruben Cuevas, Vijay Erramilli, and Nikolaos Laoutaris. I always feel like somebody's watching me: <i>measuring online behavioural advertising</i>. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, CoNEXT '15, New York, NY, USA, 2015. Association for Computing Machinery.

[38] Elisa Celis, Anay Mehrotra, and Nisheeth Vishnoi. Toward controlling discrimination in online ad auctions. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4456–4465. PMLR, 09–15 Jun 2019.

[39] A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly. Stop clickbait: Detecting and preventing clickbaits in online news media. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016.

[40] Ye Chen, Pavel Berkhin, Bo Anderson, and Nikhil R. Devanur. Real-time bidding algorithms for performance-based display ad allocation. In *KDD '11: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, page 1307–1315, New York, NY, USA, 2011. Association for Computing Machinery.

[41] Ye Chen, Dmitry Pavlov, and John F. Canny. Large-scale behavioral targeting. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, page 209–218, New York, NY, USA, 2009. Association for Computing Machinery.

[42] Chris Cillizza. Why the delayed election results prove the system is actually working, November 2020. https://www.cnn.com/2020/11/04/politics/donald-trump-joe-biden-2020-election-results/index.html.

[43] Cloudflare. What is ad fraud? https://www.cloudflare.com/learning/bots/what-is-ad-fraud/, 2022.

[44] John Cook, Rishab Nithyanand, and Zubair Shafiq. Inferring tracker-advertiser relationships in the online advertising ecosystem using header bidding. *Proceedings on Privacy Enhancing Technologies*, 2020(1):65–82, 2020.

[45] Adam W. Craig, Yuliya Komarova Loureiro, Stacy Wood, and Jennifer M.C. Vendemia. Suspicious minds: Exploring neural processes during exposure to deceptive advertising. *Journal of Marketing Research*, 49(3):361–372, 2012.

[46] Matthew Crain and Anthony Nadler. Political Manipulation and Internet Advertising Infrastructure. *Journal of Information Policy*, 9:370–410, 2019.

[47] Henriette Cramer. Effects of ad quality & content-relevance on perceived content quality. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 2231–2234, New York, NY, USA, 2015. Association for Computing Machinery.

[48] L. G. Crovitz. How Amazon, Geico and Walmart fund propaganda. The New York Times, January 2020. https://www.nytimes.com/2020/01/21/opinion/fake-news-russia-ads.html.

[49] Peter R. Darke and Robin J. B. Ritchie. The defensive consumer: Advertising deception, defensive processing, and distrust. *Journal of Marketing Research*, 44(1):114–127, 2007.

[50] Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. *ArXiv*, abs/1408.6491, 2014.

[51] Paresh Dave and Christopher Bing. Russian disinformation on YouTube draws ads, lacks warning labels: researchers, June 2019. https://www.reuters.com/article/us-alphabet-google-youtube-russia/russian-disinformation-on-youtube-draws-ads-lacks-warning-labels-researchers-idUSKCN1T80JP.

[52] Stylianos Despotakis, R. Ravi, and Amin Sayedi. First-price auctions in online display advertising. *Journal of Marketing Research*, 58(5):888–907, 2021.

[53] Jack Dorsey. We've made the decision to stop all political advertising on Twitter globally, 2019.

[54] Sam Dutton. Fledge api. https://developer.chrome.com/docs/privacy-sandbox/fledge/, 2022. Google Chrome Developers.

[55] Sam Dutton. The topics api. https://developer.chrome.com/docs/privacy-sandbox/topics/, 2022. Google Chrome Developers.

[56] Irina Dykhne. Persuasive or Deceptive - Native Advertising in Political Campaigns. *Southern California Law Review*, 91:339, 2018.

[57] Easylist Filter List Project. Easylist. https://easylist.to.

[58] Benjamin Edelman. False and deceptive display ads at yahoo's right media. https://www.benedelman.org/rightmedia-deception/, 2009.

[59] Benjamin Edelman. The pathologies of online display advertising marketplaces. *SIGecom Exch.*, 9(1), jun 2010.

[60] Laura Edelson, Tobias Lauinger, and Damon McCoy. A Security Analysis of the Facebook Ad Library. In *IEEE Symposium on Security and Privacy (Oakland)*, 2020.

[61] Laura Edelson, Shikhar Sakhuja, Ratan Dey, and Damon McCoy. An Analysis of United States Online Political Advertising Transparency. arXiv:1902.04385, February 2019.

[62] Steven M. Edwards, Hairong Li, and Joo-Hyun Lee. Forced exposure and psychological reactance: Antecedents and consequences of the perceived intrusiveness of pop-up ads. *Journal of Advertising*, 31(3):83–95, 2002.

[63] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis. In *ACM Conference on Computer and Communications Security (CCS)*, 2016.

[64] Motahhare Eslami, Sneha R. Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. Communicating algorithmic process in online behavioral advertising. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 2018.

[65] Kenya Evelyn. Capitol attack: the five people who died, 2021. https://www.theguardian.com/us-news/2021/jan/08/capitol-attack-police-officer-five-deaths.

[66] FactCheck.org. Misinformation directory. https://www.factcheck.org/2017/07/websites-post-fake-satirical-stories/.

[67] Irfan Faizullabhoy and Aleksandra Korolova. Facebook's Advertising Platform: New Attack Vectors and the Need for Interventions. In *Workshop on Consumer Protection (ConPro)*, 2018.

[68] Federal Trade Commission. Blurred Lines: Advertising or Content? An FTC Workshop on Native Advertising, December 2013. https://www.ftc.gov/system/files/documents/public_events/171321/final_transcript_1.pdf.

[69] Federal Trade Commission. .com Disclosures: How to Make Effective Disclosures in Digital Advertising, March 2013. https://www.ftc.gov/system/files/documents/plain-language/bus41-dot-com-disclosures-information-about-online-advertising.pdf.

[70] Federal Trade Commission. Commission Enforcement Policy Statement on Deceptively Formatted Advertisements. https://www.ftc.gov/public-statements/2015/12/commission-enforcement-policy-statement-deceptively-formatted, December 2015.

[71] Federal Trade Commission. Native Advertising: A Guide for Businesses. https://www.ftc.

gov/tips-advice/business-center/guidance/native-advertising-guide-businesses, December 2015.

[72] Federal Trade Commission. An Exploration of Consumers' Advertising Recognition in the Contexts of Search Engines and Native Advertising, December 2017. https://www.ftc.gov/reports/blurred-lines-exploration-consumers-advertising-recognition-contexts-search-engines-native.

[73] Federal Trade Commission. The truth behind weight loss ads, August 2019. https://www.consumer.ftc.gov/articles/truth-behind-weight-loss-ads.

[74] Federal Trade Commission. Fraud reports. https://public.tableau.com/app/profile/federal.trade.commission/viz/FraudReports/FraudFacts, 2021. Losses & Contact Methods.

[75] Sara Fischer. Scoop: Google to block election ads after election day, 2021. https://www.axios.com/google-to-block-election-ads-after-election-day-4b60650d-b5c2-4fb4-a856-70e30e19af17.html.

[76] Sara Fischer. Scoop: Google to lift political ad ban put in place following capitol siege, 2021. https://www.axios.com/capitol-siege-google-political-ad-ban-5000245d-35d6-4448-b7b2-daa7ccfe816a.html.

[77] Coalition for Better Ads. The better ads standards. https://www.betterads.org/standards/, 2022.

[78] Saadia Gabriel, Skyler Hallinan, Maarten Sap, Pemi Nguyen, Franziska Roesner, Eunsol Choi, and Yejin Choi. Misinfo reaction frames: Reasoning about readers' reactions to news headlines. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3108–3127, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[79] Stacia Garlach and Daniel Suthers. 'i'm supposed to see that?'adchoices usability in the mobile environment. In *Proceedings of the 51st Hawaii International Conference on System Sciences*, Honolulu, HI, 2018. University of Hawaii at Manoa.

[80] Global Disinformation Index. Cutting the Funding of Disinformation: The Ad-Tech Solution, May 2019. https://disinformationindex.org/wp-content/uploads/2019/05/GDI_Report_Screen_AW2.pdf.

[81] Global Disinformation Index. The Quarter Billion Dollar Question: How is Disinformation Gaming Ad Tech?, September 2019. https://disinformationindex.org/wp-content/uploads/2019/09/GDI_Ad-tech_Report_Screen_AW16.pdf.

[82] Benjamin Goggin. 7,800 people have lost their jobs so far this year in a media landslide. https://www.businessinsider.com/2019-media-layoffs-job-cuts-at-buzzfeed-huffpost-vice-details-2019-2, December 2019.

[83] Daniel G. Goldstein, R. Preston McAfee, and Siddharth Suri. The cost of annoying ads. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, page 459–470, New York, NY, USA, 2013. Association for Computing Machinery.

[84] Gustavo Gomez-Mejia. “fail, clickbait, cringe, cancel, woke”: Vernacular criticisms of digital advertising in social media platforms. In *HCII 2020: Social Computing and Social Media. Participation, User Experience, Consumer Experience, and Applications of Social Computing*, pages 309–324, Cham, CH, 2020. Springer International Publishing.

[85] Dan Goodin. Here’s why the epidemic of malicious ads grew so much worse last year. https://arstechnica.com/information-technology/2018/01/malvertising-factory-with-28-fake-agencies-delivered-1-billion-ads-in-2017/, Jan 2018. Ars Technica.

[86] Google. Ad Manager and Ad Exchange program policies - Prevent malware in ad content. https://support.google.com/admanager/answer/181490?hl=en&ref_topic=28145.

[87] Google. Google Ads policies. https://support.google.com/adspolicy/answer/6008942.

[88] Google. Puppeteer. https://developers.google.com/web/tools/puppeteer/.

[89] Google. About responsive display ads, 2020. https://support.google.com/google-ads/answer/6363750?hl=en.

[90] Google. Political advertising in the United States. https://transparencyreport.google.com/political-ads/region/US, 2021. Google Transparency Report.

[91] Megan Graham. Pinterest says it will no longer allow ads on elections-related content, employees get time off to vote, 2020.

[92] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. The dark (patterns) side of UX design. In *CHI Conference on Human Factors in Computing Systems*, 2018.

[93] Elizabeth Grieco, Nami Sumida, and Sophia Fedeli. About a third of large U.S. newspapers have suffered layoffs since 2017. https://www.pewresearch.org/fact-tank/2018/07/23/about-a-third-of-large-u-s-newspapers-have-suffered-layoffs-since-2017/, July 2018.

[94] Maarten Grootendorst. BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics, 2020.

[95] Maarten Grootendorst. Creating a class-based TF-IDF with Scikit-Learn. https://towardsdatascience.com/creating-a-class-based-tf-idf-with-scikit-learn-caea7b15b858, October 2020. Towards Data Science.

[96] Katherine Haenschen and Jordan Wolf. Disclaiming responsibility: How platforms deadlocked the Federal Election Commission’s efforts to regulate digital political advertising. *Telecommunications Policy*, 43(8):101824, 2019.

[97] Lynn Hasher, David Goldstein, and Thomas Toppino. Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1):107–112, 1977.

[98] Chris Herbert. The Fake News Codex. http://www.fakenewscodex.com, December 2018.

[99] Weiyin Hong, James Y.L. Thong, and Kar Yan Tam. How do web users respond to non-banner-ads animation? the effects of task type and user experience. *Journal of the American Society for Information Science and Technology*, 58(10):1467–1482, 2007.

[100] Michael C Hughes and Erik B Sudderth. bnpy: Reliable and scalable variational inference

for bayesian nonparametric models. In *Proceedings of the NIPS Probabilistic Programming Workshop, Montreal, QC, Canada*, pages 8–13, Cambridge, MA, 2014. MIT Press.

[101] David A. Hyman, David J. Franklyn, Calla E. Yee, and Mohammad Hossein Rahmati. Going Native: Can Consumers Recognize Native Advertising? Does it Matter? *Yale Journal of Law and Technology*, 19(2), 2018.

[102] Basileal Imana, Aleksandra Korolova, and John Heidemann. Auditing for discrimination in algorithms delivering job ads. In *Proceedings of the Web Conference 2021*, WWW '21, page 3767–3778, New York, NY, USA, 2021. Association for Computing Machinery.

[103] Costas Iordanou, Nicolas Kourtellis, Juan Miguel Carrascosa, Claudio Soriente, Ruben Cuevas, and Nikolaos Laoutaris. Beyond content analysis: Detecting targeted ads via distributed counting. In *Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies*, CoNEXT '19, page 110–122, New York, NY, USA, 2019. Association for Computing Machinery.

[104] Umar Iqbal, Pouneh Nikkhah Bahrami, Rahmadi Trimananda, Hao Cui, Alexander Gamero-Garrido, Daniel Dubois, David Choffnes, Athina Markopoulou, Franziska Roesner, and Zubair Shafiq. Your echos are heard: Tracking, profiling, and ad targeting in the amazon smart speaker ecosystem, 2022.

[105] Umar Iqbal, Steven Englehardt, and Zubair Shafiq. Fingerprinting the fingerprinters: Learning to detect browser fingerprinting behaviors. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 1143–1161, 2021.

[106] Mark Irvine. Google ads benchmarks for your industry. https://www.wordstream.com/blog/ws/2016/02/29/google-adwords-industry-benchmarks, 2021. Wordstream: Online Advertising Made Easy.

[107] Caroline Jack. Lexicon of lies: Terms for problematic information. Data & Society, August 2017.

[108] Junqi Jin, Chengru Song, Han Li, Kun Gai, Jun Wang, and Weinan Zhang. Real-time bidding with multi-agent reinforcement learning in display advertising. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, page 2193–2201, New York, NY, USA, 2018. Association for Computing Machinery.

[109] Gita Venkataramani Johar. Consumer involvement and deception from implied advertising claims. *Journal of Marketing Research*, 32(3):267–279, 1995.

[110] Gita Venkataramani Johar. Intended and unintended effects of corrective advertising on beliefs and evaluations: An exploratory analysis. *Journal of Consumer Psychology*, 5(3):209–230, 1996.

[111] Gita Venkataramani Johar. Mistaken inferences from advertising conversations: A modest research agenda. *Journal of Advertising*, 45(3):318–325, 2016.

[112] Gita Venkataramani Johar and Anne L. Roggeveen. Changing false beliefs from repeated advertising: The role of claim-refutation alignment. *Journal of Consumer Psychology*, 17(2):118–127, 2007.

[113] Joshua Gillin. The more outrageous, the better: How clickbait ads make

money for fake news sites. https://www.politifact.com/punditfact/article/2017/oct/04/more-outrageous-better-how-clickbait-ads-make-mone/, October 2017.

[114] Jordan Jueckstock, Shaown Sarker, Peter Snyder, Aidan Beggs, Panagiotis Papadopoulos, Matteo Varvello, Benjamin Livshits, and Alexandros Kapravelos. Towards realistic and reproducible web crawl measurements. In *Proceedings of the Web Conference 2021*, WWW '21, page 80–91, New York, NY, USA, 2021. Association for Computing Machinery.

[115] A-Reum Jung and Jun Heo. Ad disclosure vs. ad recognition: How persuasion knowledge influences native advertising evaluation. *Journal of Interactive Advertising*, 19(1):1–14, 2019.

[116] Young Mie Kim, Jordan Hsu, David Neiman, Colin Kou, Levi Bankston, Soo Yun Kim, Richard Heinrich, Robyn Baragwanath, and Garvesh Raskutti. The Stealth Media? Groups and Targets behind Divisive Issue Campaigns on Facebook. *Political Communication*, 35(4):515–541, 2018.

[117] Kim LaCapria. Snopes' field guide to fake news sites and hoax purveyors. Snopes, January 2016. https://www.snopes.com/news/2016/01/14/fake-news-sites/.

[118] Sara Kingsley, Clara Wang, Alex Mikhalenko, Proteeti Sinha, and Chinmay Kulkarni. Auditing Digital Platforms for Discrimination in Economic Opportunity Advertising, 2020.

[119] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 03 1951.

[120] Laura Kloot. Native Ads vs. Display Ads: What are the differences? https://www.outbrain.com/blog/native-ads-vs-display-ads/, July 2018.

[121] Joe Lazauskas. Fixing Native Ads: What Consumers Want From Publishers, Brands, Facebook, and the FTC. Contently, 2016. https://the-content-strategist-13.docs.contently.com/v/fixing-sponsored-content-what-consumers-want-from-brands-publishers-and-the-ftc.

[122] Tomo Lazovich, Luca Belli, Aaron Gonzales, Amanda Bower, Uthaipon Tantipongpipat, Kristian Lum, Ferenc Huszar, and Rumman Chowdhury. Measuring disparate outcomes of content recommendation algorithms with distributional inequality metrics, 2022.

[123] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In *26th Network and Distributed System Security Symposium (NDSS)*, 2019.

[124] Mathias Lécuyer, Guillaume Ducoffe, Francis Lan, Andrei Papancea, Theofilos Petsios, Riley Spahn, Augustin Chaintreau, and Roxana Geambasu. Xray: Enhancing the web's transparency with differential correlation. In *USENIX Security Symposium*, 2014.

[125] Mathias Lecuyer, Riley Spahn, Yannis Spiliopolous, Augustin Chaintreau, Roxana Geambasu, and Daniel Hsu. Sunlight: Fine-grained targeting detection at scale with statistical confidence. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, page 554–566, New York, NY, USA, 2015. Association for Computing Machinery.

[126] Dami Lee. The FTC is cracking down on influencer marketing on YouTube, Instagram, and TikTok. https://www.theverge.com/2020/2/12/21135183/ftc-influencer-ad-sponsored-tiktok-youtube-instagram-review, February 2020. The Verge.

[127] Jonathan Lemire, Zeke Miller, and Will Weissert. Biden defeats Trump for White House, says 'time to heal', 2021. https://apnews.com/article/joe-biden-wins-white-house-ap-fd58df73aa677acb74fce2a69adb71f9.

[128] Ada Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In *25th USENIX Security Symposium*, 2016.

[129] Zhou Li, Kehuan Zhang, Yinglian Xie, Fang Yu, and XiaoFeng Wang. Knowing Your Enemy: Understanding and Detecting Malicious Web Advertising. In *ACM Conference on Computer and Communications Security (CCS)*, 2012.

[130] Ad Lightning. The 2020 State of Ad Quality Report. Obtained at https://blog.adlightning.com/the-2020-state-of-ad-quality-report, January 2020.

[131] Bin Liu, Anmol Sheth, Udi Weinsberg, Jaideep Chandrashekar, and Ramesh Govindan. Adreveal: Improving transparency into online targeted advertising. In *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks*, HotNets-XII, New York, NY, USA, 2013. Association for Computing Machinery.

[132] Tong Liu, Akash Venkatachalam, Pratik Sanjay Bongale, and Christopher Homan. Learning to predict population-level label distributions. In *Companion Proceedings of The 2019 World Wide Web Conference*, WWW '19, page 1111–1120, New York, NY, USA, 2019. Association for Computing Machinery.

[133] Sapna Maheshwari and John Herrman. Publishers are rethinking those 'around the web' ads. New York Times, October 2016. https://web.archive.org/web/20200331090649/https://www.nytimes.com/2016/10/31/business/media/publishers-rethink-outbrain-taboola-ads.html.

[134] John Mahoney. A Complete Taxonomy of Internet Chum. The Awl, June 2015. https://www.theawl.com/2015/06/a-complete-taxonomy-of-internet-chum/.

[135] Matthew Malloy, Mark McNamara, Aaron Cahn, and Paul Barford. Ad blockers: Global prevalence and impact. In *Proceedings of the 2016 Internet Measurement Conference*, IMC '16, page 119–125, New York, NY, USA, 2016. Association for Computing Machinery.

[136] Lachlan Markay. Pro-trump 'news sites' are harvesting your emails to sell to the campaign. https://www.thedailybeast.com/pro-trump-news-sites-are-harvesting-your-emails-to-sell-to-the-campaign, August 2020. The Daily Beast.

[137] Lachlan Markay. Pro-Trump scam PAC operator hit with wire fraud charge. Axios, April 2021. https://www.axios.com/pro-trump-scam-pac-operator-wire-fraud-charge-d888876f-680c-4b47-83d8-d72da709ce8b.html.

[138] Arunesh Mathur, Gunes Acar, Michael Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. Dark patterns at scale: Findings from a crawl of 11k shopping websites. *Proceedings of the ACM on Human-Computer Interaction (CSCW)*, 2019.

[139] Arunesh Mathur, Arvind Narayanan, and Marshini Chetty. Endorsements on Social Media: An Empirical Study of Affiliate Marketing Disclosures on YouTube and Pinterest. *Proceedings*

of the ACM on Human-Computer Interaction (CSCW), 2, November 2018.

[140] Arunesh Mathur, Angelina Wang, Carsten Schwemmer, Maia Hamin, Brandon M. Stewart, and Arvind Narayanan. Manipulative tactics are the norm in political emails: Evidence from 100K emails from the 2020 U.S. election cycle. https://electionemails2020.org, 2020.

[141] Mary L. McHugh. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*, 22(3):276–82, October 2012.

[142] Media Bias/Fact Check Team. Media Bias/Fact Check: The Most Comprehensive Media Bias Resource. https://mediabiasfactcheck.com/fake-news/.

[143] Cristina Miranda. Can you trust that ad for a dietary supplement? Federal Trade Commission, November 2017. https://www.consumer.ftc.gov/blog/2017/11/can-you-trust-ad-dietary-supplement.

[144] Shekhar Misra. Deceptive political advertising: Some new dimensions. *Journal of Legal, Ethical and Regulatory Issues*, 18:71–80, 2015.

[145] Brian Morrissey. How the banner ad was born. https://digiday.com/marketing/how-the-banner-ad-was-born/, April 2013. DigiDay.

[146] Mozilla. mozilla/readability. https://github.com/mozilla/readability.

[147] Michael Muller. *Ways of Knowing*, chapter Curiosity, Creativity, and Surprise as Analytic Tools: Grounded Theory Method. Springer, 2014.

[148] John Murphy. SSPs aren't created equally: Insights into which SSPs drove the most problematic impressions in Q2, 2020. https://www.confiant.com/resources/blog/ssp-performance-q4-2019, August 2020. Confiant.

[149] Arvind Narayanan, Arunesh Mathur, Marshini Chetty, and Mihir Kshirsagar. Dark patterns: Past, present, and future. *Communications of the ACM*, 63(9):42–47, August 2020.

[150] Suman Nath. Madscope: Characterizing mobile in-app targeted ads. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '15, page 59–73, New York, NY, USA, 2015. Association for Computing Machinery.

[151] Terry Nelms, Roberto Perdisci, Manos Antonakakis, and Mustaque Ahamad. Towards Measuring and Mitigating Social Engineering Software Download Attacks. In *Proceedings of the 25th USENIX Conference on Security Symposium*, SEC'16, page 773–789, USA, 2016. USENIX Association.

[152] Casey Newton. You might also like this story about weaponized clickbait. The Verge, apr 2014. https://www.theverge.com/2014/4/22/5639892/how-weaponized-clickbait-took-over-the-web.

[153] Nick Nikiforakis, Alexandros Kapravelos, Wouter Joosen, Christopher Krügel, Frank Piessens, and Giovanni Vigna. Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting. In *IEEE Symposium on Security and Privacy*, pages 541–555, 2013.

[154] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In *CHI Conference on Human Factors in Computing Systems*, 2020.

[155] Abby Ohlheiser. This is how Facebook's fake-news writers make money. Washington Post, November 2016. https://www.washingtonpost.com/news/the-intersect/wp/2016/11/18/this-is-how-the-internets-fake-news-writers-make-money/.

[156] Lukasz Olejnik, Tran Minh-Dung, and Claude Castelluccia. Selling off privacy at auction. In *Network and Distributed System Security (NDSS) Symposium*, 2014.

[157] OpenSources Contributors. Opensources. https://github.com/OpenSourcesGroup/opensources, April 2017.

[158] Michalis Pachilakis, Panagiotis Papadopoulos, Nikolaos Laoutaris, Evangelos P. Markatos, and Nicolas Kourtellis. Youradvalue: Measuring advertising price dynamics without bankrupting user privacy. *Proc. ACM Meas. Anal. Comput. Syst.*, 5(3), dec 2021.

[159] Michalis Pachilakis, Panagiotis Papadopoulos, Evangelos P. Markatos, and Nicolas Kourtellis. No more chasing waterfalls: A measurement study of the header bidding ad-ecosystem. In *Proceedings of the Internet Measurement Conference*, IMC '19, page 280–293, New York, NY, USA, 2019. Association for Computing Machinery.

[160] Elleen Pan, Jingjing Ren, Martina Lindorfer, Christo Wilson, and David R. Choffnes. Panoptispy: Characterizing audio and video exfiltration from android applications. In *Proceedings of Privacy Enhancing Technologies Symposium*, 2018.

[161] Abhimanyu Panwar, Iosif-Viorel Onut, and James Miller. Towards real time contextual advertising. In Boualem Benatallah, Azer Bestavros, Yannis Manolopoulos, Athena Vakali, and Yanchun Zhang, editors, *Web Information Systems Engineering – WISE 2014*, 2014.

[162] Panagiotis Papadopoulos, Nicolas Kourtellis, Pablo Rodriguez Rodriguez, and Nikolaos Laoutaris. If you are not paying for it, you are the product: How much do advertisers pay to reach you? In *Proceedings of the 2017 Internet Measurement Conference*, pages 142–156, 2017.

[163] Javier Parra-Arnau, Jagdish Prasad Achara, and Claude Castelluccia. Myadchoices: Bringing transparency and control to online advertising. *ACM Transactions on the Web*, 11(1), mar 2017.

[164] Michel Tuan Pham and Gita Venkataramani Johar. Contingent processes of source identification. *Journal of Consumer Research*, 24(3):249–265, 1997.

[165] Angelisa C. Plane, Elissa M. Redmiles, Michelle L. Mazurek, and Michael Carl Tschantz. Exploring user perceptions of discrimination in online targeted advertising. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 935–951, Vancouver, BC, August 2017. USENIX Association.

[166] Politifact. Fact-checking U.S. politics. https://www.politifact.com/.

[167] Martin Potthast, Tim Gollub, Matthias Hagen, and Benno Stein. The clickbait challenge 2017: Towards a regression model for clickbait strength, 2018.

[168] PropOrNot Team. Is it propaganda or not?: Your friendly neighborhood propaganda identification service, since 2016! http://www.propornot.com/p/home.html.

[169] Niels Provos, Panayiotis Mavrommatis, Moheeb Abu Rajab, and Fabian Monrose. All your iframes point to us. In *Proceedings of the 17th Conference on Security Symposium*, SS'08, page 1–15, USA, 2008. USENIX Association.

[170] Enric Pujol, Oliver Hohlfeld, and Anja Feldmann. Annoyed users: Ads and ad-block usage in the wild. In *Proceedings of the 2015 Internet Measurement Conference*, IMC '15, page 93–106, New York, NY, USA, 2015. Association for Computing Machinery.

[171] Emilee Rader. Awareness of behavioral tracking and information privacy concern in facebook and google. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 51–67, Menlo Park, CA, July 2014. USENIX Association.

[172] Christy Rakoczy. The sneaky reason you should never sign petitions or answer surveys online. Mic.com, May 2017. https://www.mic.com/articles/175333/want-to-save-money-heres-the-surprising-reason-why-you-should-never-sign-an-online-petition.

[173] Vaibhav Rastogi, Rui Shao, Yan Chen, Xiang Pan, Shihong Zou, and Ryan Riley. Are these Ads Safe: Detecting Hidden Attacks through the Mobile App-Web Interfaces. In *Network and Distributed System Security Symposium (NDSS)*, 2016.

[174] Filipe N. Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P. Gummadi, and Elissa M. Redmiles. On microtargeting socially divisive ads: A case study of russia-linked ad campaigns on facebook. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 140–149, New York, NY, USA, 2019. Association for Computing Machinery.

[175] Travis N. Ridout, Erika Franklin Fowler, and Michael M. Franz. Spending Fast and Furious: Political Advertising in 2020. *The Forum*, 18(4), 2021. https://www.degruyter.com/document/doi/10.1515/for-2020-2109/html.

[176] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. Detecting and Defending Against Third-Party Tracking on the Web. In *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2012.

[177] Anne L. Roggeveen and Gita Venkataramani Johar. Perceived source variability versus familiarity: Testing competing explanations for the truth effect. *Journal of Consumer Psychology*, 12(2):81–91, 2002.

[178] Christian Rohrer and John Boyd. The rise of intrusive online advertising and the response of user experience research at yahoo! In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, page 1085–1086, New York, NY, USA, 2004. Association for Computing Machinery.

[179] Matthew Rosenberg. Ad Tool Facebook Built to Fight Disinformation Doesn't Work as Advertised, July 2019. https://www.nytimes.com/2019/07/25/technology/facebook-ad-library.html.

[180] Rahul Roy-Chowdhury. Improving advertising on the web. https://blog.chromium.org/2017/06/improving-advertising-on-web.html, June 2017. Chromium Blog.

[181] Navdeep S. Sahni and Harikesh Nair. Sponsorship disclosure and consumer deception: Experimental evidence from native advertising in mobile search, March 2018. https://ssrn.com/abstract=2737035.

[182] Johnny Saldana. *The Coding Manual for Qualitative Researchers.* SAGE Publications, 2012.

[183] Iskander Sanchez-Rola, Matteo Dell'Amico, Davide Balzarotti, Pierre-Antoine Vervier, and

Leyla Bilge. Journey to the Center of the Cookie Ecosystem: Unraveling Actors' Roles and Relationships. In *IEEE Symposium on Security and Privacy*, 2021.

[184] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *ArXiv*, https://arxiv.org/abs/1910.01108v4, 2020.

[185] Michael S. Schmidt and Luke Broadwater. Officers' Injuries, Including Concussions, Show Scope of Violence at Capitol Riot, 2021. https://www.nytimes.com/2021/02/11/us/politics/capitol-riot-police-officer-injuries.html.

[186] Elena Schneider. Google will end political ad ban this week, 2021. https://www.politico.com/news/2020/12/09/google-end-political-ad-ban-443882.

[187] D. Sculley, Matthew Eric Otey, Michael Pohl, Bridget Spitznagel, John Hainsworth, and Yunkai Zhou. Detecting adversarial advertisements in the wild. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2011.

[188] Sergio Serra. Ad Quality: The Other Side Of The Fence. https://www.inmobi.com/blog/2019/05/10/ad-quality-the-other-side-of-the-fence, May 2019. Inmobi.

[189] Sharon Hurley Hall. Native Ads vs. Display Ads. https://blog.taboola.com/native-ads-vs-display-ads/, November 2019.

[190] Alex Shellhammer. The need for mobile speed. https://web.archive.org/web/20200723132931/https://www.blog.google/products/admanager/the-need-for-mobile-speed/, September 2016. Google Official Blog.

[191] Irina Shklovski, Scott D. Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 2347–2356, New York, NY, USA, 2014. Association for Computing Machinery.

[192] Craig Silverman and Laurence Alexander. How Teens In The Balkans Are Duping Trump Supporters With Fake News, November 2016. https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo#.tcR3aZa5r.

[193] Alexander Smith and Vladimir Banic. Fake News: How a Partying Macedonian Teen Earns Thousands Publishing Lies. NBC News, December 2016. https://www.nbcnews.com/news/world/fake-news-how-partying-macedonian-teen-earns-thousands-publishing-lies-n692451.

[194] Vera Sosnovik and Oana Goga. Understanding the Complexity of Detecting Political Ads. In *The Web Conference (WWW)*, 2021.

[195] Marianna Spring. 'Stop the steal': The deep roots of Trump's 'voter fraud' strategy, 2021. https://www.bbc.com/news/blogs-trending-55009950.

[196] Politifact staff. Politifact's guide to fake news websites and what they peddle. Politifact, April 2017. https://www.politifact.com/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/.

[197] Statista. U.s. digital advertising industry - statistics & facts. https://www.statista.com/topics/1176/online-advertising/, May 2022.

[198] Eric Van Steenburg. Areas of research in political advertising: a review and research agenda. *International Journal of Advertising*, 34(2):195–231, 2015.

[199] Jennifer Stromer-Galley, Jeff Hemsley, Patricia Rossini, Brian McKernan, Nancy McCracken, Sarah Bolden, Ania Korsunska, Shloak Gupta, Jay Kachhadia, Wenzhe Zhang, and Feifei Zhang. The Illuminating Project: Helping Journalists Cover Social Media in the Presidential Campaign, 2020. https://illuminating.ischool.syr.edu/.

[200] Michael Swart, Ylana Lopez, Arunesh Mathur, and Marshini Chetty. Is this an ad?: Automatically disclosing online endorsements on youtube with adintuition. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–12, New York, NY, USA, 2020. Association for Computing Machinery.

[201] Michael Swart, Arunesh Mathur, and Marshini Chetty. Is This An Ad? Help Us Identify Misleading Content On YouTube. Freedom to Tinker, July 2019. https://freedom-to-tinker.com/2019/07/09/is-this-an-ad-help-us-identify-misleading-content-on-youtube/.

[202] Latanya Sweeney. Discrimination in online ad delivery. *Communications of the ACM*, 56(5):44–54, 2013.

[203] Taboola. Advertising content policies overview. https://help.taboola.com/hc/en-us/articles/115007287467-Advertising-Content-Policies-Overview.

[204] AdStage Team. Google Display Ads CPM, CPC, & CTR Benchmarks in Q1 2018. https://blog.adstage.io/google-display-ads-cpm-cpc-ctr-benchmarks-in-q1-2018, 2018. AdStage.

[205] BBC Visual Journalism Team. Capitol riots: A visual guide to the storming of Congress, 2021. https://www.bbc.com/news/world-us-canada-55575260.

[206] James Temperton. We need to talk about the internet's fake ads problem. Wired, March 2017. https://www.wired.co.uk/article/fake-news-outbrain-taboola-hillary-clinton.

[207] Kaitlyn Tiffany. A mysterious gut doctor is begging americans to throw out "this vegetable" now. but, like, which? Vox, May 2019. https://www.vox.com/the-goods/2019/5/8/18537279/chum-box-weird-sponsored-links-gut-doctor.

[208] Barbara Trish. From recording videos in a closet to Zoom meditating, 2020's political campaigns adjust to the pandemic. https://theconversation.com/from-recording-videos-in-a-closet-to-zoom-meditating-2020s-political-campaigns-adjust-to-the-pandem 2020. The Conversation.

[209] Eray Turkel. Regulating online political advertising. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 3584–3593, New York, NY, USA, 2022. Association for Computing Machinery.

[210] Joseph Turow, Jennifer King, Chris Jay Hoofnagle, Amy Bleakley, and Michael Hennessy. Americans reject tailored advertising and three activities that enable it. *Available at SSRN 1478214*, 2009.

[211] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart,

useful, scary, creepy: Perceptions of online behavioral advertising. In *Symposium on Usable Privacy and Security (SOUPS)*, 2012.

[212] Giridhari Venkatadri, Athanasios Andreou, Yabing Liu, Alan Mislove, Krishna P. Gummadi, Patrick Loiseau, and Oana Goga. Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface. In *IEEE Symposium on Security and Privacy*, 2018.

[213] Paul Vines, Franziska Roesner, and Tadayoshi Kohno. Exploring ADINT: Using Ad Targeting for Surveillance on a Budget - or - How Alice Can Buy Ads to Track Bob. In *Workshop on Privacy in the Electronic Society (WPES)*, 2017.

[214] Jun Wang, Weinan Zhang, and Shuai Yuan. Display advertising with real-time bidding (rtb) and behavioural targeting. *Found. Trends Inf. Retr.*, 11:297–435, 2016.

[215] Tharindu Cyril Weerasooriya, Tong Liu, and Christopher M. Homan. Neighborhood-based pooling for population-level label distribution learning, 2020.

[216] Miranda Wei, Madison Stamos, Sophie Veys, Nathan Reitinger, Justin Goodman, Margot Herman, Dorota Filipczuk, Ben Weinshel, Michelle L. Mazurek, and Blase Ur. What twitter knows: Characterizing ad targeting practices, user perceptions, and ad explanations through users' own twitter data. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 145–162, Austin, TX, August 2020. USENIX Association.

[217] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L. Mazurek, and Blase Ur. Oh, the places you've been! user reactions to longitudinal transparency about third-party web tracking and inferencing. In *ACM Conference on Computer and Communications Security (CCS)*, 2019.

[218] Bartosz W. Wojdynski. The deceptiveness of sponsored news articles: How readers recognize and perceive native advertising. *American Behavioral Scientist*, 60(12):1475–1491, 2016.

[219] Bartosz W. Wojdynski and Nathaniel J. Evans. Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising. *Journal of Advertising*, 45(2):157–168, 2016.

[220] Xinyu Xing, Wei Meng, Byoungyoung Lee, Udi Weinsberg, Anmol Sheth, Roberto Perdisci, and Wenke Lee. Understanding Malvertising Through Ad-Injecting Browser Extensions. In *24th International Conference on World Wide Web (WWW)*, 2015.

[221] Yaxing Yao, Davide Lo Re, and Yang Wang. Folk models of online behavioral advertising. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, page 1957–1969, New York, NY, USA, 2017. Association for Computing Machinery.

[222] Jianhua Yin and Jianyong Wang. A Dirichlet Multinomial Mixture Model-Based Approach for Short Text Clustering. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, page 233–242, New York, NY, USA, 2014. Association for Computing Machinery.

[223] Shuai Yuan, Jun Wang, and Xiaoxue Zhao. Real-time bidding for online advertising: Measurement and analysis. In *Proceedings of the Seventh International Workshop on Data Mining*

for Online Advertising, ADKDD '13, New York, NY, USA, 2013. Association for Computing Machinery.

[224] A. Zarras, A. Kapravelos, G. Stringhini, T. Holz, C. Kruegel, and G. Vigna. The dark alleys of Madison Avenue: Understanding malicious advertisements. In *ACM Internet Measurement Conference*, 2014.

[225] David Zeber, Sarah Bird, Camila Oliveira, Walter Rudametkin, Ilana Segall, Fredrik Wollsén, and Martin Lopatka. *The Representativeness of Automated Web Crawls as a Surrogate for Human Browsing*, page 167–178. Association for Computing Machinery, New York, NY, USA, 2020.

[226] Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. Bad News: Clickbait and Deceptive Ads on News and Misinformation Websites. In *Workshop on Technology and Consumer Protection*, 2020.

[227] Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. What Makes a "Bad" Ad? User Perceptions of Problematic Online Advertising. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[228] Eric Zeng, Miranda Wei, Theo Gregersen, Tadayoshi Kohno, and Franziska Roesner. Polls, clickbait, and commemorative $2 bills: Problematic political advertising on news and media websites around the 2020 u.s. elections. In *Proceedings of the 21st ACM Internet Measurement Conference*, IMC '21, page 507–525, New York, NY, USA, 2021. Association for Computing Machinery.

[229] Wei Zha and H. Denis Wu. The impact of online disruptive ads on users' comprehension, evaluation of site credibility, and sentiment of intrusiveness. *American Communication Journal*, 16:15–28, 2014.

[230] Weinan Zhang, Shuai Yuan, and Jun Wang. Optimal real-time bidding for display advertising. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, page 1077–1086, New York, NY, USA, 2014. Association for Computing Machinery.

[231] Jun Zhao, Guang Qiu, Ziyu Guan, Wei Zhao, and Xiaofei He. Deep reinforcement learning for sponsored search real-time bidding. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '18, page 1021–1030, New York, NY, USA, 2018. Association for Computing Machinery.

[232] Alain F Zuur, Elena N Ieno, Neil J Walker, Anatoly A Saveliev, Graham M Smith, et al. *Mixed effects models and extensions in ecology with R*, volume 574. Springer, 2009.

Appendix A

Chapter 3

A.1 Survey 1 Protocol

In this survey, we are trying to learn about how you think about online advertisements. In particular, we want to know what kinds of online ads you like and dislike, and why. First, we have a few questions about your attitudes towards ads in general. After these questions, we will show you some examples of ads, and have you tell us what you think about them.

1. Think about the ads you see when browsing social media or news, on your computer or your phone. What kinds of ads do you like seeing, if any? (Free response)
2. What kinds of ads do you dislike the most, and why? Here are some optional prompts to guide your answer: Are there specific ads that you remember disliking? Is there a type/genre of ad that you dislike in general? Do you see more ads that you dislike on certain apps or websites? (Free response)
3. Do you use an ad blocker? (AdBlock, AdBlock Plus, uBlock Origin, etc.) (Yes/No/Not Sure)

The following questions will ask about the advertisement shown below. (Repeated 4 times)

4. What is your overall opinion of this ad? (7 point Likert Scale, Extremely Positive — Extremely Negative)
5. What parts of this ad, if any, did you like? And why? (Free Response)
6. What parts of this ad, if any, did you dislike? And why? (Free Response)
7. What words or phrases would you use to describe the style of this ad, and your emotions/reactions when you see this ad?
8. Have you seen this ad before, or ads similar to this one? (Free Response)
9. What do you like and/or dislike about ads similar to this one? (Free Response)

Now that you've seen some examples of ads, we'd like you to think one more time about the questions we asked at the beginning of the survey.

10. Think about the ads you see when browsing social media or news, on your computer or your phone. What kinds of ads do you like seeing, if any? (Free Response)

11. What kinds of ads do you dislike the most, and why? (Free Response)
12. Do you have anything else you'd like to tell us that we didn't ask about, regarding how you feel about online ads? (Free Response)

A.2 Survey 2 Protocol

Below is the text of the survey protocol we used in survey 2, to gather opinion labels and other data from 1025 participants. A screenshot of the ad-labeling interface is included in Figure A.1.

In this survey, we are trying to learn about what kinds of online advertisements you like and dislike, and why. First, we have a few questions about your attitudes towards ads in general.

1. When visiting websites (like news websites, social media, etc.), how much do you like seeing ads? (7-point Likert scale, Extremely Dislike — Extremely Like)
2. Do you use an ad blocker? (e.g. AdBlock, AdBlock Plus, uBlock Origin) (Yes/No/Not Sure)

In this survey, we will be asking you to look at 5 online ads and provide your opinion of each of them.
For each ad, we will first ask you to rate your overall opinion of the ad, on a scale ranging from extremely negative to extremely positive.
Please provide your honest opinion about how you feel about these ads. You might find some of them to be interesting or benign, and others to be annoying or boring, for example. Depending on the ad, your answers might be different from your opinion of online ads in general.
(For each ad, repeated 5 times:)

3. What is your overall opinion of this ad? (7 point Likert scale, Extremely Negative — Extremely Positive)
4. Which of the following categories would you use to describe your opinion of this ad? (Note that participants were also provided with the full category definitions shown verbatim in Table 3.2.)
    - Boring, Irrelevant
    - Cheap, Ugly, Badly Designed
    - Clickbait
    - Deceptive, Untrustworthy
    - Don't Like the Product or Topic
    - Offensive, Uncomfortable, Distasteful
    - Politicized
    - Pushy, Manipulative
    - Unclear
    - Entertaining, Engaging

- Good Style and Design
- Interested in the Product or Topic
- Simple, Straightforward
- Trustworthy, Genuine
- Useful, Interesting, Informative

5. How strongly do you agree with each of the categories you picked, on a scale of 1-5? Where 1 means “a little” and 5 means “a lot”. (1-5 scale, for each chosen category)
6. Are there other reasons you like or dislike this ad not covered by these categories? (optional, free response)

Before we let you go, we have two last questions about you to help us understand how people feel about political ads.

7. When it comes to politics, do you usually think of yourself as a Democrat, a Republican, an Independent, or something else?
8. Lastly, we want to ensure that you have been reading the questions in the survey. Please select the "Somewhat Negative" option below. Thank you for paying attention!

A.3 Ad Content Codes

Tables A.1 and A.2 list the content codes used to describe the semantic content of ads in Survey 2.

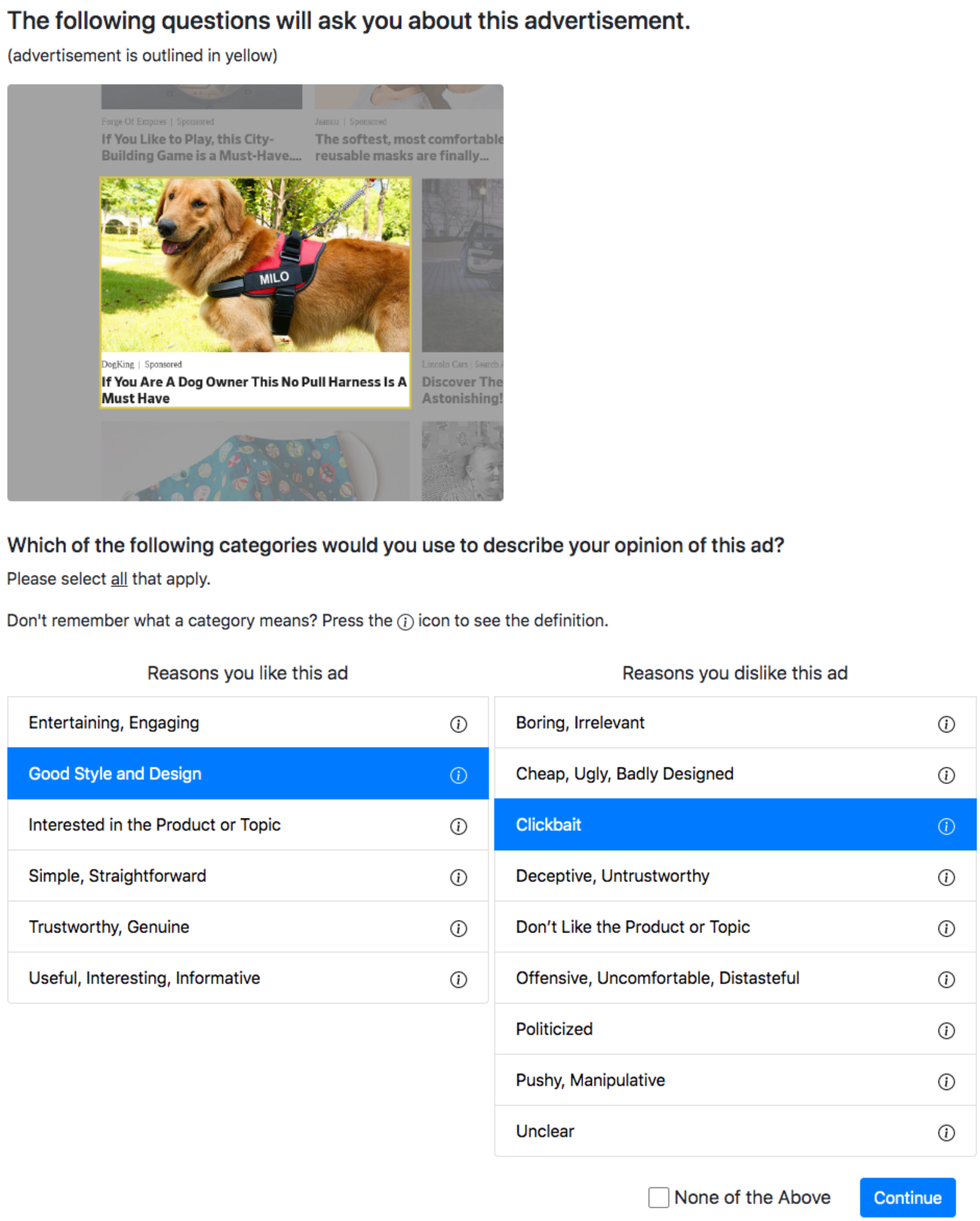


Figure A.1: A screenshot of the survey interface in survey 2. For each ad, participants were able to pick multiple reasons for why they liked/disliked an ad. These responses were used as opinion labels in our analysis.

Category	Content Code	Definition
Ad Formats	Image	Standard banner ads where the advertiser designs 100% of the ad content.
	Native	Ads that imitate first party site content in style and placement, such as ads that look like news article headlines.
	Sponsored Content	Ads for articles and content on the host page, that are explicitly sponsored by (and possibly written by) an advertiser.
	Google Responsive	Google Responsive Display Ads [89], a Google-specific ad format, where advertisers provide the text, pictures (optional), and a logo (optional), and Google renders it (possibly in different layouts). We highlight this format because it is common and has a distinctive visual style (e.g., the fonts and buttons in Figure 3.6b), and it is similar to native ads in terms of the ease for advertisers to create an ad.
	Poll	Ads that are interactive polls (not just an image of a poll).
Misleading Techniques	Advertorial	Ad where the landing page looks like a news article, but is selling a product.
	Decoy	A phishing technique, where advertisers place a large clickable button in the ad to attract/distract users from the page, imitating other buttons or actions on a page, like a "Continue" or "Download" button [151]
	Listicle	An ad where the headline promises a list of items e.g., "10 things you won't believe", and/or if the landing page is a list of items or slideshow.
	Political Poll	An ad that appears to be polling for a political opinion, but may have a different true purpose, like harvesting email addresses [18].
	Sponsored Search	An ad whose landing page is search listings, rather than a specific product
Topics	Apparel	Ads for clothes, shoes, and accessories
	B2B Products	Ads for any product intended to be sold to businesses
	Banking	Financial services that banks provide to consumers, financial advisors, brokerages
	Beauty Products	Cosmetics and skincare products
	Cars	Automobiles and motorcycles
	Cell Service	Mobile phone plans
	Celebrity News	Ads for articles about celebrities; gossip
	Consumer Tech	Smartphones, laptops, smart devices; accessories for consumer electronics
	Contest	Ads for giveaways, lotteries, etc.
	COVID Products	Masks, hand sanitizer, or other health measures for COVID
	Dating	Dating apps and services
	Education	Ads for colleges, degree programs, training, etc.
	Employment	Job listings
	Entertainment	Ads for entertainment content, e.g., TV, books, movies, etc.
	Food and Drink	Anything food related, e.g., recipes and restaurants
	Games and Toys	Video games, board games, mobile games, toys
	Genealogy	Ads for genealogy services/social networks
	Gifts	Ads for gifts, gift cards
	Health and Supplements	Ads for supplements and wellness advice, excludes medical services
	Household Products	Ads for furniture, home remodeling, any other home products
	Humanitarian	Ads for charities and humanitarian efforts, public service announcements
	Human Interest	Ads for articles that are generic, evergreen, baseline appealing to anyone
	Insurance	Ads for any kind of insurance product — home, car, life, health, etc.

Table A.1: Content codes that we (the researchers) used to label the content of ads in the Survey 2 dataset. Continues on next page...

Category	Content Code	Definition
Topics (cont.)	Investment Pitch	An ad promoting a specific investment product, opportunity, or newsletter
	Journalism	Ads from journalistic organizations — programs, newsletters, etc.
	Legal Services	Ads for law firms, lawyers, or lawyers seeking people in specific legal situations
	Medical Services and Prescriptions	Ads for prescription drugs, doctors and specific medical services
	Mortgages	Ads for mortgages, mortgage refinancing, or reverse mortgages
	Pets	Ads for pet products
	Political Campaign	Ads from an official political campaign
	Political Memorabilia	Ads for political souvenirs/memorabilia, like coins
	Public Relations	An ad intended to provide information about a company to improve public perceptions
	Real Estate	Ads for property rentals/sales
	Recreational Drugs	Ads for alcohol, tobacco, marijuana, or other drugs
	Religious	Ads for religious news, articles, or books
	Social Media	Ads for social media services
	Software Download	Ad promoting downloadable consumer software
	Sports	Ad with anything sports-related - sports leagues, sports equipment, etc.
	Travel	Ad for anything travel related - destinations, lodging, vehicle rentals, flights
	Weapons	Ad for firearms or accessories like body armor
	Wedding Services	Any services or products specifically for weddings, like photographers

Table A.2: Content codes that we (the researchers) used to label the content of ads in the Survey 2 dataset. (Continued from previous page).

Appendix B

Chapter 6

B.1 Data Collection Extension Screenshots

Figures B.1-B.5 show screenshots of the user interface of the browser extension that participants used to collect data.

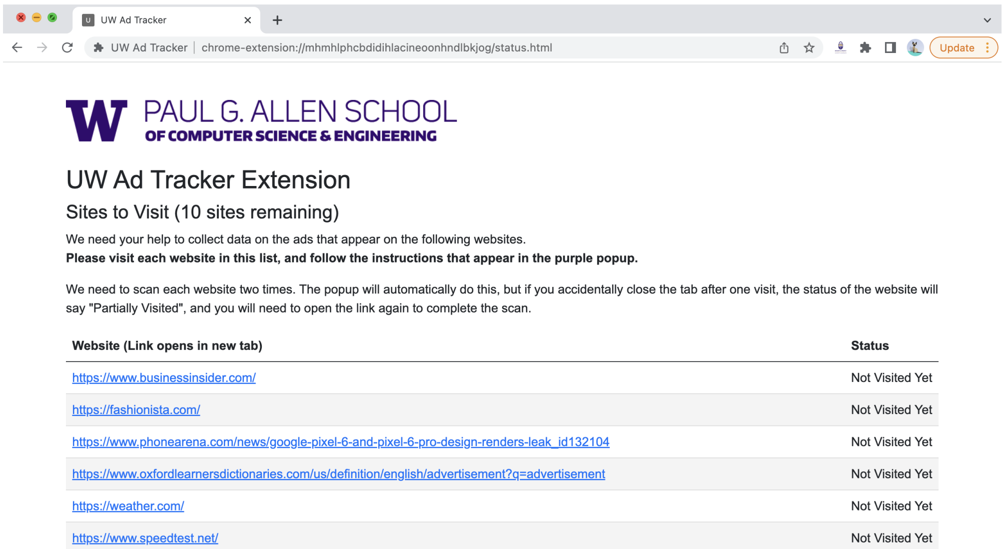


Figure B.1: After registering the data collection extension, participants are instructed to visit all 10 websites in this list. (Certain elements of page are redacted for anonymization)

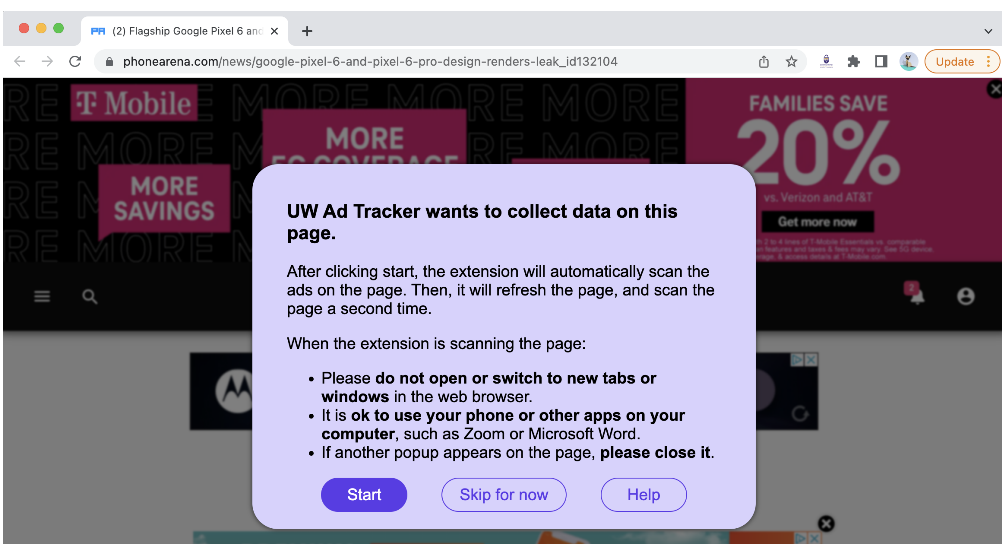


Figure B.2: On visiting a site from the list, participants are asked for permission to collect data.

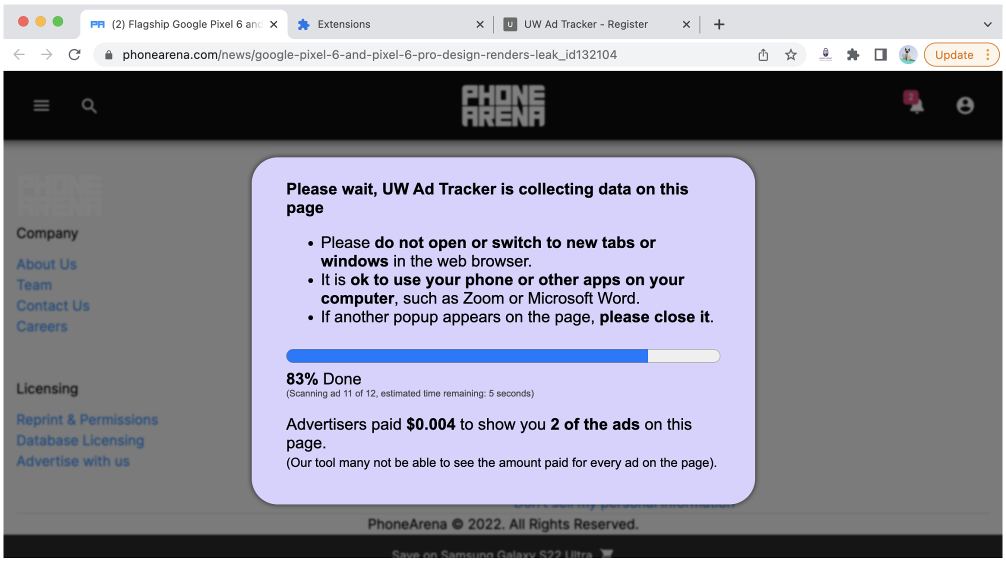


Figure B.3: The extension scans the page from top to bottom, one ad at a time. During this time the participant is instructed to not navigate from the page or open other tabs, which interferes with the screenshot process.

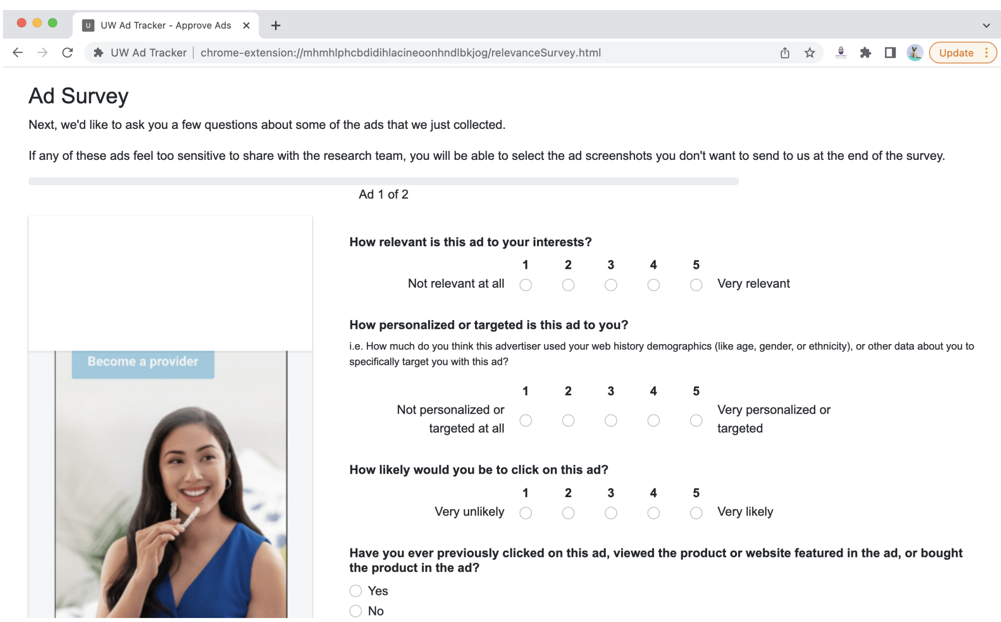


Figure B.4: After all data is collected, for a sample of their ads, participants are asked about how targeted they perceived the ad to be.

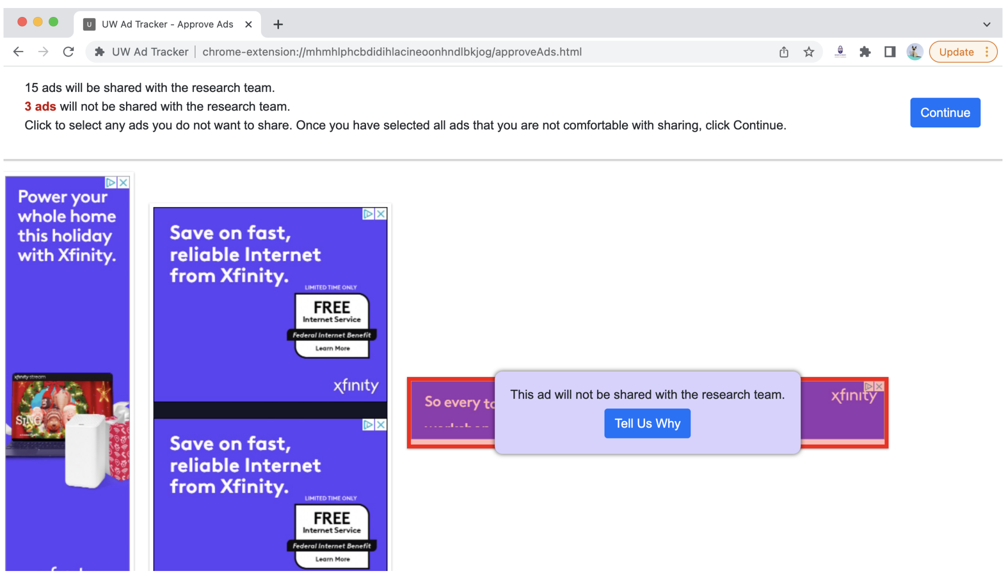


Figure B.5: Lastly, participants can opt out of sending any screenshots that they did not want to share with us.