What Makes a “Bad” Ad? User Perceptions of Problematic Online Advertising

Eric Zeng
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, WA, USA
ericzeng@cs.washington.edu

Tadayoshi Kohno
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, WA, USA
yoshi@cs.washington.edu

Franziska Roesner
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, WA, USA
franzi@cs.washington.edu

ABSTRACT

Online display advertising on websites is widely disliked by users, with many turning to ad blockers to avoid “bad” ads. Recent evidence suggests that today’s ads contain potentially problematic content, in addition to well-studied concerns about the privacy and intrusiveness of ads. However, we lack knowledge of which types of ad content users consider problematic and detrimental to their browsing experience. Our work bridges this gap: first, we create a taxonomy of 15 positive and negative user reactions to online advertising from a survey of 60 participants. Second, we characterize 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. Among our findings, we report that users consider a substantial amount of ads on the web today to be clickbait, untrustworthy, or distasteful, including ads for software downloads, listicles, and health & supplements.

CCS CONCEPTS

• **Information systems → Online advertising**; **Display advertising**; • **Social and professional topics → Commerce policy**; • **Human-centered computing** → *User studies.*

KEYWORDS

online advertising, deceptive advertising, dark patterns

ACM Reference Format:
Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. 2021. What Makes a “Bad” Ad? User Perceptions of Problematic Online Advertising. In *CHI Conference on Human Factors in Computing Systems (CHI ’21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/3411764.3445459

1 INTRODUCTION

Online display advertising is a critical part of the modern web: ads sustain websites that provide free content and services to consumers, and many ads inform people about products and services that they are interested in. Still, 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 [5] — for instance, a 2016 study estimated that 18% of U.S. internet users and 37% of German internet users used an ad blocker [69], 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 [5] or their impact on the load times of websites [92]. Users are also concerned about the privacy impacts of ads: research in computer security and privacy has revealed extensive ecosystems of tracking and targeted advertising (e.g., [9, 28, 30, 61, 62, 64, 76, 84, 97, 98]), which users often find to be creepy and privacy-invasive (e.g., [29, 96, 100, 101]). The specific *content* of ads can also cause direct or indirect harms to consumers, ranging from material harms in the extreme (e.g., scams [1, 34, 72], malware [65, 74, 104, 105], and discriminatory advertising [3, 57]) to simply annoying techniques that disrupt the user experience (e.g., animated banner ads [16, 38, 45]).

In this work, 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 [4, 7, 22, 39, 52, 63, 68, 71, 75, 90, 93, 103, 106]. 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?

CHI ’21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445459

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 [18, 33, 63], 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 [106].

Research Questions. In this paper, 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 [2], 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 [26, 87]) 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.

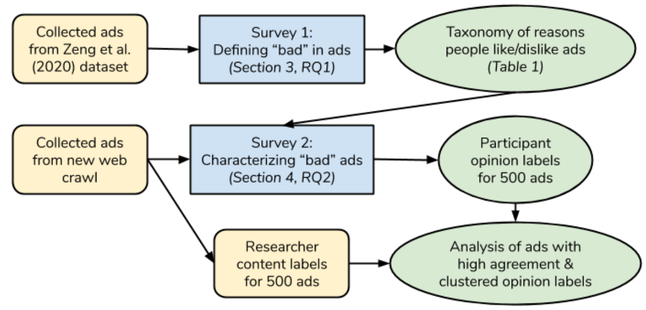

Figure 1: An overview of our work and contributions.

Contributions. Figure 1 shows an overview of the different components of our work and our resulting outputs and contributions. Specifically, our contributions include:

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, 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 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 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 advertising ecosystem.

2 BACKGROUND, RELATED WORK, AND MOTIVATION

2.1 Background and Related Work

Broadly speaking, related work has studied (1) problematic content and techniques in ads directly and/or (2) people's perception of ads. Our work is inspired by (1) and expands on (2). In this section, we break down prior work based on types of concerns with ads.

Computer Security Risks and Discrimination. Online ads have often been leveraged for malicious and harmful purposes. For example, prior work in computer security has studied the use of ads to spread malware, clickfraud, and phishing attacks (e.g., [65, 74, 82, 104, 105]).

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

Researchers have also surfaced concerns about how ads may be targeted at users in potentially discriminatory ways (e.g., [3, 57]), such as by (intentionally or unintentionally) serving ads for certain employment opportunities disproportionately to certain demographic groups.

Deceptive Ads. Another class of problematic ads is those which are explicitly deceptive — either in terms of the claims that they make, or in their appearance as advertisements at all. 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. [14, 50, 51]), that people infer false claims not directly stated in ads and misattribute claims to incorrect sources (e.g., [44, 48, 79, 85]), and that people’s awareness of specific deceptive ads can harm their attitudes towards those brands [49] as well as towards advertising *in general* [19, 23].

More recently on the web, there has been significant concern about “native” advertisements which are designed to blend into the primary content of the hosting website (e.g., sponsored search results or social media posts, or 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., [4, 9, 47, 53, 53, 59, 89, 102, 103]) — though people may do better after more experience [53], or with different disclosure designs (e.g., [47, 103]). Deceptive ads may affect user behavior even when identified [89]. Prior work suggests that native ads can reduce user perceptions of the credibility of the hosting site, even if the ads are rated as high quality in isolation [22]. Recent work has also raised concerns about unclear affiliate marketing and/or endorsements on social media [71, 93, 108]. Beyond outright scams, much of the U.S. Federal Trade Commission’s recent enforcement surrounding web ads has focused on ad disclosures for such native ads of various types [31–33, 63].

Annoying and Disruptive Ads. Even when ads are not explicitly malicious or causing material harms, many users still dislike them. Traditionally, a common reason that people dislike web ads is that they are annoying and disruptive — either in general or due to their specific designs — leading in part to the development and widespread adoption of ad blockers [5, 69, 81]. 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 [12, 27, 38, 86]. 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 [12, 17, 107].

Clickbait and Other Low-Quality Content. The concerns around annoying and disruptive ads in the previous paragraph stem largely from the design of ads. In addition, many modern web ads contain low-quality content that walks a fine line between “good” and “bad” ads. Recent anecdotal and scientific evidence suggests that there is a wide range of problematic, distasteful, and misleading content in online ads (or on the websites to which they lead) — including low-quality “clickbait”, content farms, misleading or deceptive claims, mis/disinformation, and voter suppression [8, 9, 24, 36, 37, 52, 52, 54–56, 68, 75, 78, 90, 94, 95].

Our previous work measured the prevalence of these types of ads on news and misinformation sites, finding that large fractions of ads on both types of sites contained content that was potentially problematic (based on a researcher-created definition of “problematic”) [106]. Though anecdotes suggest that users recognize and dislike such ads (for example, a recent qualitative study of French-speaking Twitter discussions surfaced user criticisms of social media ads using terms such as “Fail”, “Clickbait”, and “Cringe” [39]), user perceptions of these types of ads — which seem not to directly violate the policies of ad providers or regulators today — have not been systematically studied.

Further afield but related are broader discussions of “dark patterns” [15], e.g., on websites [70] and in mobile apps (e.g., [11, 43, 77]), though none of these works considered web ads.

Ad Targeting and Privacy. Finally, a unique aspect of the online advertising ecosystem is the ability to track and target specific individuals. Significant prior work has revealed and measured the privacy implications of these tracking and targeting capabilities (e.g. [28, 30, 61, 62, 64, 76, 84, 97, 98]), which end users may find “creepy”, insufficiently transparent, or otherwise distasteful [6, 29, 96, 100, 101]. We note that such ad targeting may increase the impacts of problematic ad content if such content is delivered to particularly susceptible users (e.g., [83]).

2.2 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 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 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

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 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.

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.

3.1 Survey 1 Methodology

3.1.1 Survey Protocol. We curated a set of 30 ads found on the web (described below in Section 3.1.2). 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 for the full survey protocol.

3.1.2 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 [106].

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"; 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 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 [41].

3.1.3 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 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.

3.1.4 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

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

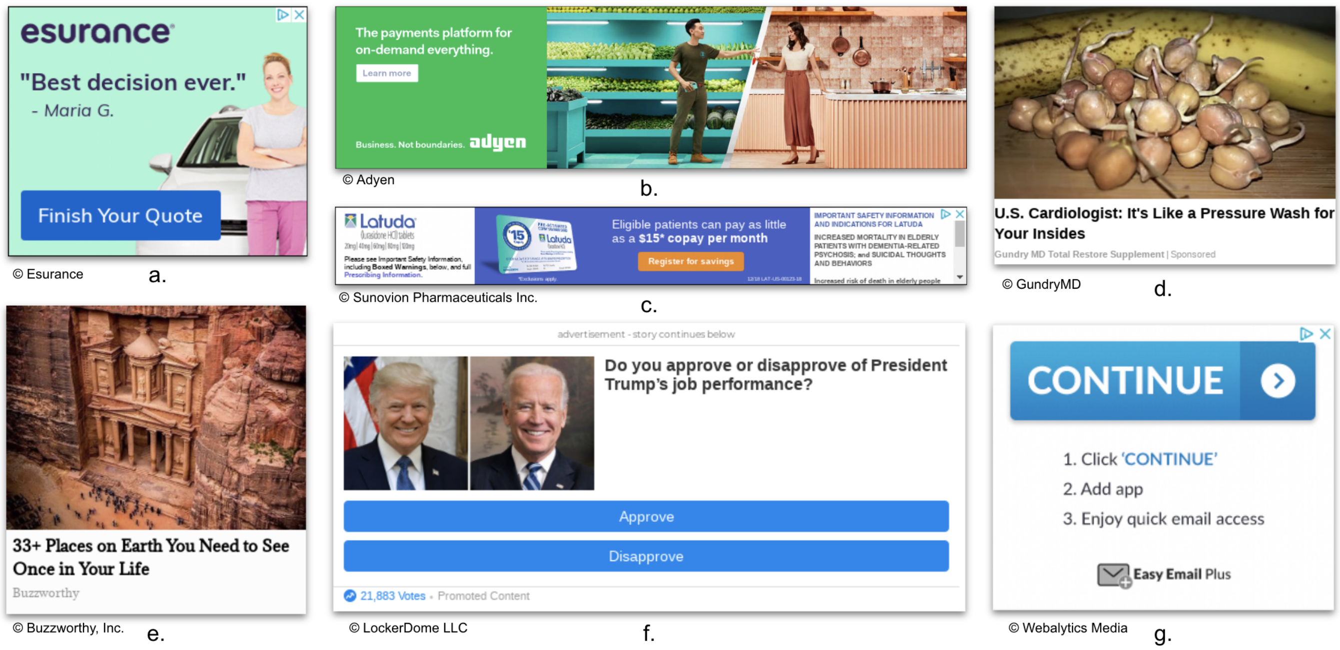

Figure 2: A sample of ads shown to participants in Survey 1, selected from the dataset of our prior work [106]. In that study, ads a-c were categorized as “benign”, and were each coded as “Product”. Ads d-f were categorized as “problematic”, using the following codes: d) Supplement, e) Content Farm, f) Political Poll, g) Potentially Unwanted Software.

time, coding the results, and repeating until new themes appeared infrequently. The demographics of the participants are shown in Table 1. Our participant sample skewed younger, compared to the overall U.S. population, and contained more ad blocker users than some estimates [69].

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).

3.2 Survey 1 Results

Table 2 summarizes the reasons participants liked or disliked ads, based on the codes developed in our qualitative analysis.

3.2.1 *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:

> 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 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

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 2: The categories of reasons that participants gave for liking or disliking ads, in response to our qualitative Survey 1 (Section 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 4).

> 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 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 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 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 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 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.

3.2.2 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 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 5.

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?

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.

4.1 Survey 2 Methodology

4.1.1 Ads Dataset. We wanted to collect participant ratings on a large, diverse dataset of actual ads from the web. Thus we created a

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

new dataset by crawling the top 3000 most popular U.S.-registered websites, based on the Tranco top ranked sites list [60], matching the crawling methodology used to collect the ads in Survey 1 [106].

We crawled these sites using a custom-built web crawler based on Puppeteer, a browser automation library for the Chromium browser [42]. When the crawler visits a site, it identifies ads using the Easylist [25], 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. 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.

4.1.2 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 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 B for the full survey protocol.

4.1.3 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 [74].

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

4.1.4 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”.

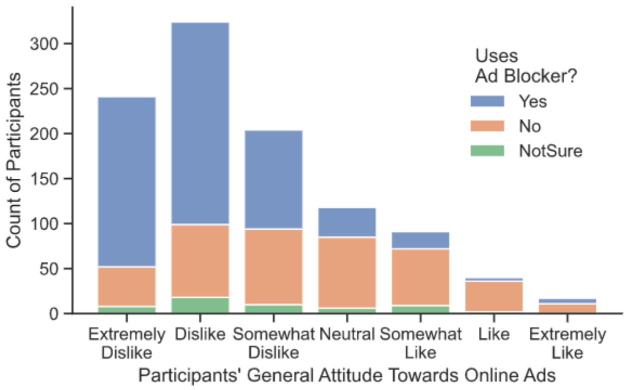

Figure 3: Histogram of how much participants reported liking/disliking seeing online ads in general. Overall, most participants disliked seeing ads; ad blocker users disliked ads more.

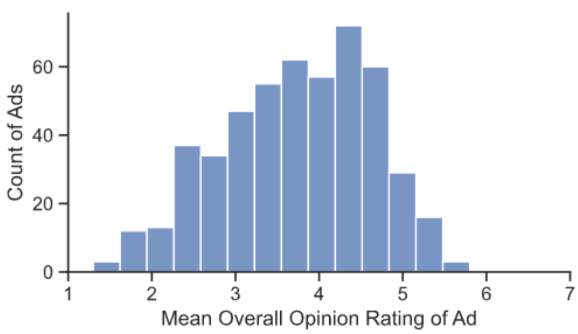

Figure 4: Histogram of average overall opinion rating for ads in our dataset, where the values 1-7 map to a Likert scale ranging from extremely negative to extremely positive. Ratings for ads skewed negative, with a median score of 3.8.

4.1.5 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 1. Our sample was younger than the overall U.S. population, and contained more ad blocker users than some estimates [69].

4.2 Survey 2 Results

4.2.1 General Attitudes Towards Ads. Our participants generally skewed towards disliking ads to begin with. Figure 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.

[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.

	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 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.

4.2.2 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 4 shows a histogram of the average opinion rating for each ad (on 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.

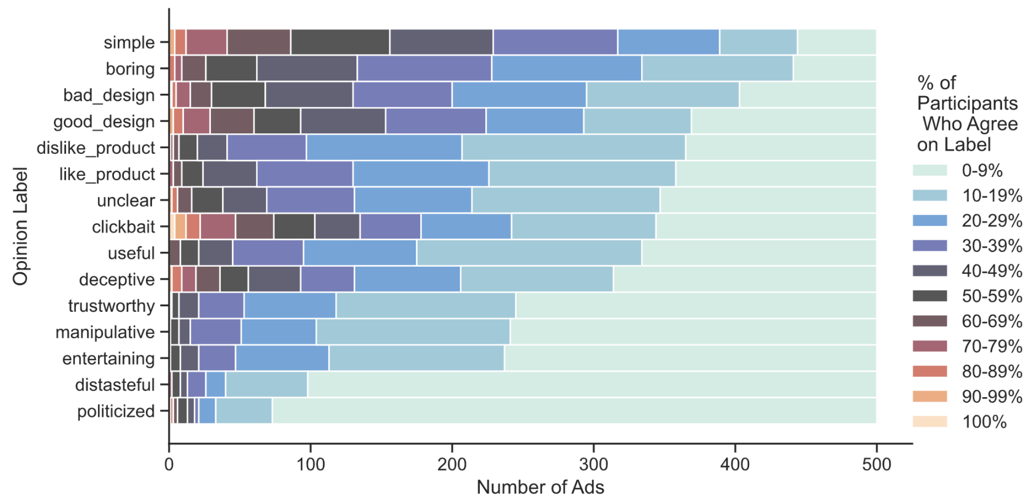

Figure 5: A stacked histogram representing the distribution of agreement values for each opinion labels. Each bar represents the number of ads annotated with an opinion label, subdivided into bars representing ranges of agreement values. For example, the width of the black sub-bar for the "Simple" code 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.

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 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 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 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.

4.2.3 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 [67].

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. [67] and Weerasooriya et al. [99] to have the best clustering performance on similar benchmark datasets, measured using Kullback–Leibler (KL) divergence, a measure of the difference between probability distributions [58]. The highest performing FMM model on our dataset was trained with 40 fixed clusters, and achieved a

[5] We used an implementation of FMM-vB in the bnpy library (https://github.com/bnpy/bnpy) [46].

KL-divergence of 0.227, similar to the performance measured by Liu et al. and Weerasooriya et al. for other PLDL datasets [67, 99].

Clustering Results Overview. Our model produced 16 thematically distinct clusters, which we summarize in Table 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 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" [68]. These ads imitate the design of links to news articles on the site, and have been considered borderline deceptive by the FTC and researchers [31, 47, 59, 102].

Numerous participants suspected that "clickbait" native ads, such as the ad in Figure 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 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 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 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 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 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 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 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.

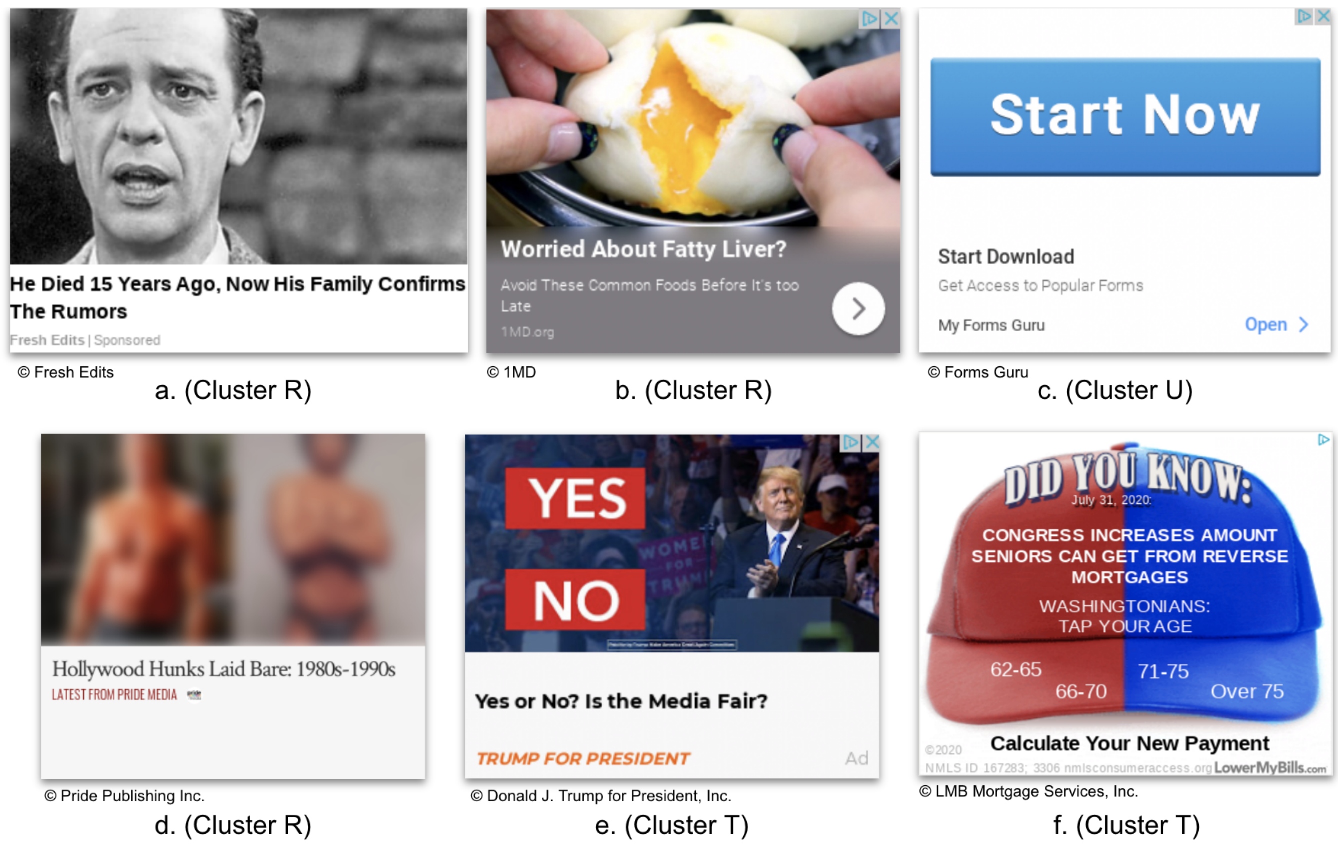

Figure 6: Examples of “bad” ads from our dataset, which appeared in clusters where participants frequently used labels such as “clickbait”, “deceptive/untrustworthy” “distasteful”, and “politicized”. (a) is a native ad, for a celebrity news “listicle”, sometimes called a “content farm”, an article designed 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.

- *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 7a), for popular products, like TV shows (Figure 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 7c) to lotions (Figure 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.

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 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.

- *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 7e). They also 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%).

4.2.4 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.

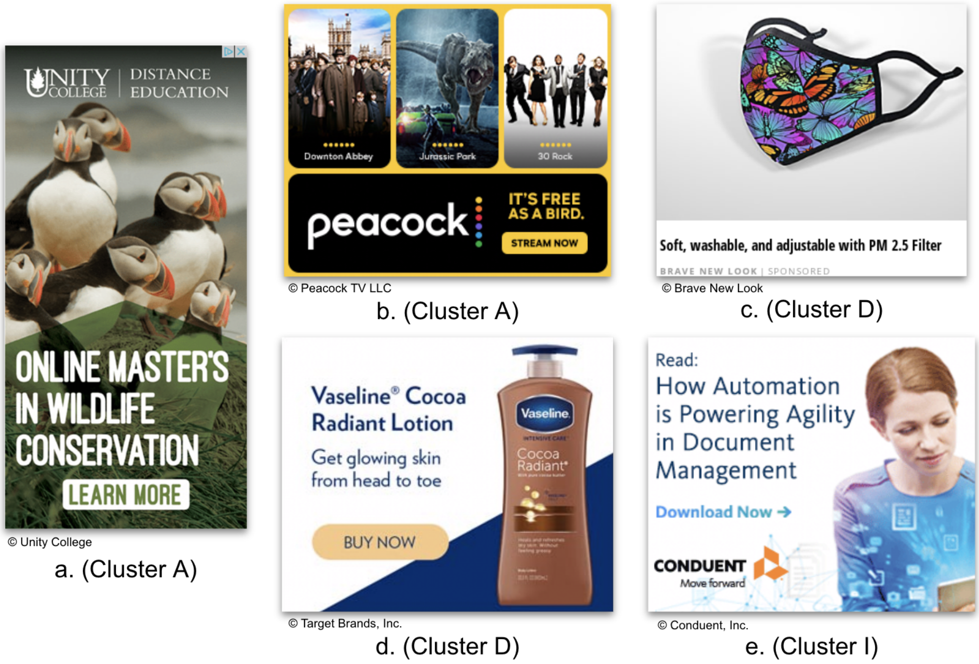

Figure 7: A sample of “good” ads from our dataset, that participants labeled positively, with labels like “good design”, “simple”, and “interested in product”. (a) and (b) are ads for consumer services, and were in the highest rated cluster because of attractive visuals and appealing products. (c) and (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.

```
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 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).

5 DISCUSSION

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

5.1 Broader Lessons on the Online Advertising Ecosystem

5.1.1 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 [41]), 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.

5.1.2 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).

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 6: Linear mixed effects regression model of participants' overall ratings for individual ads. 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.

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 [2]. 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. [13] 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" [66, 73, 91].

Moreover, even publishers themselves are incentivized to run "bad ads" at times, despite potential reputational risk [22], because "bad ads" can help generate revenue, especially when legitimate advertisers pull back on spending [10].

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.

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".

5.2.1 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 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 [20] and enforcement policy [21] 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 2g and 6c).

5.2.2 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 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 [35].

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 [88]. 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.

5.3 Future Research Directions

5.3.1 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" [21].

5.3.2 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 [3, 61, 62, 80].

5.3.3 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 [67], 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.

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 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 8 in Appendix B).

Our participant samples skewed towards younger people and adblocker users. This reflects the overall userbase of Prolific[7] and the tendency for younger people to use ad blockers [69]. 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 [106] 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 provide useful insight into clusters of ads that participants had unambiguously negative or positive views of.

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.

ACKNOWLEDGMENTS

We thank all of our study participants for making this work possible. We also thank Cyril Weerasooriya, Chris Homan, and Tong Liu for

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

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

sharing their codebase, and insights on population label distribution learning. We thank Kentrell Owens, Lucy Simko, Miranda Wei, and Sam Wolfson for testing our surveys and providing feedback on our paper. We thank all of the anonymous reviewers for their feedback and insight. This work was supported in part by the National Science Foundation under awards CNS-1565252 and CNS-1651230, and by 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.

REFERENCES

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

[2] AlgoAware. 2019. 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. Observatory on the Online Platform Economy.

[3] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW, Article 199 (Nov. 2019), 30 pages. https://doi.org/10.1145/3359301

[4] Michelle A. Amazeen and Bartosz W. Wojdynski. 2019. Reducing Native Advertising Deception: Revisiting the Antecedents and Consequences of Persuasion Knowledge in Digital News Contexts. *Mass Communication and Society* 22, 2 (2019), 222–247.

[5] Mimi An. 2016. 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. HubSpot Research.

[6] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebooks Explanations. In *25th Annual Network and Distributed System Security Symposium, NDSS 2018, San Diego, California, USA, February 18-21, 2018*. The Internet Society, Reston, VA, 15 pages. http://wp.internetsociety.org/ndss/wp-content/uploads/sites/25/2018/02/ndss2018_10-1_Andreou_paper.pdf

[7] Anocha Aribarg and Eric M. Schwartz. 2020. Native Advertising in Online News: Trade-Offs Among Clicks, Brand Recognition, and Website Trustworthiness. *Journal of Marketing Research* 57, 1 (2020), 20–34. https://doi.org/10.1177/0022243719879711

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

[9] Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson. 2016. Tracing Information Flows Between Ad Exchanges Using Retargeted Ads. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Austin, TX, 481–496. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/bashir

[10] Andrew Blustein. 2020. 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/. Adweek.

[11] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proceedings on Privacy Enhancing Technologies* 2016, 4 (July 2016), 237–254.

[12] Giorgio Brajnik and Silvia Gabrielli. 2010. A Review of Online Advertising Effects on the User Experience. *International Journal of Human–Computer Interaction* 26, 10 (2010), 971–997.

[13] Joshua A. Braun and Jessica L. Eklund. 2019. Fake News, Real Money: Ad Tech Platforms, Profit-Driven Hoaxes, and the Business of Journalism. *Digital Journalism* 7, 1 (2019), 1–21. https://doi.org/10.1080/21670811.2018.1556314

[14] Kathryn A. Braun and Elizabeth F. Loftus. 1998. Advertising's misinformation effect. *Applied Cognitive Psychology* 12, 6 (1998), 569–591.

[15] Harry Brignull. 2019. Dark Patterns. https://www.darkpatterns.org/.

[16] Moira Burke, Anthony Hornof, Erik Nilsen, and Nicholas Gorman. 2005. High-Cost Banner Blindness: Ads Increase Perceived Workload, Hinder Visual Search, and Are Forgotten. *ACM Trans. Comput.-Hum. Interact.* 12, 4 (Dec. 2005), 423–445. https://doi.org/10.1145/1121112.1121116

[17] Moira Burke, Anthony Hornof, Erik Nilsen, and Nicholas Gorman. 2005. High-Cost Banner Blindness: Ads Increase Perceived Workload, Hinder Visual Search, and Are Forgotten. *ACM Transactions on Compututer-Human Interaction* 12, 4 (Dec. 2005), 423–445.

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

[19] Margaret C. Campbell. 1995. 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), 225–254.

[20] Federal Trade Comission. 2015. Native Advertising: A Guide for Businesses. https://www.ftc.gov/tips-advice/business-center/guidance/native-advertising-guide-businesses.

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

[22] Henriette Cramer. 2015. Effects of Ad Quality & Content-Relevance on Perceived Content Quality. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2231–2234. https://doi.org/10.1145/2702123.2702360

[23] Peter R. Darke and Robin J. B. Ritchie. 2007. The Defensive Consumer: Advertising Deception, Defensive Processing, and Distrust. *Journal of Marketing Research* 44, 1 (2007), 114–127.

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

[25] Easylist Filter List Project. 2020. Easylist. https://easylist.to.

[26] L. Edelson, T. Lauinger, and D. McCoy. 2020. A Security Analysis of the Facebook Ad Library. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, New York, NY, 661–678. https://doi.org/10.1109/SP40000.2020.00084

[27] Steven M. Edwards, Hairong Li, and Joo-Hyun Lee. 2002. Forced Exposure and Psychological Reactance: Antecedents and Consequences of the Perceived Intrusiveness of Pop-Up Ads. *Journal of Advertising* 31, 3 (2002), 83–95.

[28] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-Million-Site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (Vienna, Austria) *(CCS '16)*. Association for Computing Machinery, New York, NY, USA, 1388–1401. https://doi.org/10.1145/2976749.2978313

[29] Motahhare Eslami, Sneha R. Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. 2018. Communicating Algorithmic Process in Online Behavioral Advertising. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174006

[30] Irfan Faizullabhoy and Aleksandra Korolova. 2018. Facebook's Advertising Platform: New Attack Vectors and the Need for Interventions. In *Workshop on Consumer Protection (ConPro)*. IEEE, New York, NY, 6 pages.

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

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

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

[34] Federal Trade Commission. 2019. The Truth Behind Weight Loss Ads. https://www.consumer.ftc.gov/articles/truth-behind-weight-loss-ads.

[35] Stacia Garlach and Daniel Suthers. 2018. 'I'm supposed to see that?'AdChoices Usability in the Mobile Environment. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. University of Hawaii at Manoa, Honolulu, HI, 10 pages. https://doi.org/10.24251/HICSS.2018.476

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

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

[38] Daniel G. Goldstein, R. Preston McAfee, and Siddharth Suri. 2013. The Cost of Annoying Ads. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13)*. Association for Computing Machinery, New York, NY, USA, 459–470. https://doi.org/10.1145/2488388.2488429

[39] Gustavo Gomez-Mejia. 2020. "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*. Springer International Publishing, Cham, CH, 309–324.

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

[41] Google. 2020. Google Ads policies. https://support.google.com/adspolicy/answer/6008942.

[42] Google. 2020. Puppeteer. https://developers.google.com/web/tools/puppeteer/.

[43] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI ’18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174108

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

[45] Weiyin Hong, James Y.L. Thong, and Kar Yan Tam. 2007. 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 (2007), 1467–1482. https://doi.org/10.1002/asi.20624

[46] Michael C Hughes and Erik B Sudderth. 2014. bnpy: Reliable and scalable variational inference for bayesian nonparametric models. In *Proceedings of the NIPS Probabilistic Programming Workshop, Montreal, QC, Canada*. MIT Press, Cambridge, MA, 8–13.

[47] David A. Hyman, David J. Franklyn, Calla Yee, and Mohammad Rahmati. 2017. Going Native: Can Consumers Recognize Native Advertising? Does it Matter? 19 Yale J.L. & Tech. 77.

[48] Gita Venkataramani Johar. 1995. Consumer Involvement and Deception from Implied Advertising Claims. *Journal of Marketing Research* 32, 3 (1995), 267–279.

[49] Gita Venkataramani Johar. 1996. Intended and Unintended Effects of Corrective Advertising on Beliefs and Evaluations: An Exploratory Analysis. *Journal of Consumer Psychology* 5, 3 (1996), 209–230.

[50] Gita Venkataramani Johar. 2016. Mistaken Inferences from Advertising Conversations: A Modest Research Agenda. *Journal of Advertising* 45, 3 (2016), 318–325.

[51] Gita Venkataramani Johar and Anne L. Roggeveen. 2007. Changing False Beliefs from Repeated Advertising: The Role of Claim-Refutation Alignment. *Journal of Consumer Psychology* 17, 2 (2007), 118–127.

[52] Joshua Gillin. 2017. 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/.

[53] A-Reum Jung and Jun Heo. 2019. Ad Disclosure vs. Ad Recognition: How Persuasion Knowledge Influences Native Advertising Evaluation. *Journal of Interactive Advertising* 19, 1 (2019), 1–14.

[54] Dan Keating, Kevin Schaul, and Leslie Shapiro. 2017. The Facebook ads Russians targeted at different groups. The Washington Post. https://www.washingtonpost.com/graphics/2017/business/russian-ads-facebook-targeting/.

[55] Young Mie Kim. 2018. Voter Suppression Has Gone Digital. Brennan Center for Justice. https://www.brennancenter.org/our-work/analysis-opinion/voter-suppression-has-gone-digital.

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

[57] Sara Kingsley, Clara Wang, Alex Mikhalenko, Proteeti Sinha, and Chinmay Kulkarni. 2020. Auditing Digital Platforms for Discrimination in Economic Opportunity Advertising. arXiv:2008.09656 [cs.CY]

[58] S. Kullback and R. A. Leibler. 1951. On Information and Sufficiency. *Ann. Math. Statist.* 22, 1 (03 1951), 79–86. https://doi.org/10.1214/aoms/1177729694

[59] Joe Lazauskas. 2016. Fixing Native Ads: What Consumers Want From Publishers, Brands, Facebook, and the FTC. Contently. https://the-content-strategist-13.docs.contently.com/v/fixing-sponsored-content-what-consumers-want-from-brands-publishers-and-the-ftc.

[60] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. 2019. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In *26th Network and Distributed System Security Symposium (NDSS)*. The Internet Society, Reston, VA, 15 pages.

[61] Mathias Lécuyer, Guillaume Ducoffe, Francis Lan, Andrei Papancea, Theofilos Petsios, Riley Spahn, Augustin Chaintreau, and Roxana Geambasu. 2014. XRay: Enhancing the Web’s Transparency with Differential Correlation. In *23rd USENIX Security Symposium (USENIX Security 14)*. USENIX Association, San Diego, CA, 49–64. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/lecuyer

[62] Mathias Lecuyer, Riley Spahn, Yannis Spiliopolous, Augustin Chaintreau, Roxana Geambasu, and Daniel Hsu. 2015. Sunlight: Fine-Grained Targeting Detection at Scale with Statistical Confidence. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (Denver, Colorado, USA) *(CCS ’15)*. Association for Computing Machinery, New York, NY, USA, 554–566. https://doi.org/10.1145/2810103.2813614

[63] Dami Lee. 2020. 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. The Verge.

[64] Adam Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. 2016. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Austin, TX, 997–1013. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/lerner

[65] Zhou Li, Kehuan Zhang, Yinglian Xie, Fang Yu, and XiaoFeng Wang. 2012. Knowing Your Enemy: Understanding and Detecting Malicious Web Advertising. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security* (Raleigh, North Carolina, USA) *(CCS ’12)*. Association for Computing Machinery, New York, NY, USA, 674–686. https://doi.org/10.1145/2382196.2382267

[66] Ad Lightning. 2020. The 2020 State of Ad Quality Report. Obtained at https://blog.adlightning.com/the-2020-state-of-ad-quality-report.

[67] Tong Liu, Akash Venkatachalam, Pratik Sanjay Bongale, and Christopher Homan. 2019. Learning to Predict Population-Level Label Distributions. In *Companion Proceedings of The 2019 World Wide Web Conference* (San Francisco, USA) *(WWW ’19)*. Association for Computing Machinery, New York, NY, USA, 1111–1120. https://doi.org/10.1145/3308560.3317082

[68] John Mahoney. 2015. A Complete Taxonomy of Internet Chum. The Awl. https://www.theawl.com/2015/06/a-complete-taxonomy-of-internet-chum/.

[69] Matthew Malloy, Mark McNamara, Aaron Cahn, and Paul Barford. 2016. Ad Blockers: Global Prevalence and Impact. In *Proceedings of the 2016 Internet Measurement Conference* (Santa Monica, California, USA) *(IMC ’16)*. Association for Computing Machinery, New York, NY, USA, 119–125. https://doi.org/10.1145/2987443.2987460

[70] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 81 (Nov. 2019), 32 pages. https://doi.org/10.1145/3359183

[71] Arunesh Mathur, Arvind Narayanan, and Marshini Chetty. 2018. Endorsements on Social Media: An Empirical Study of Affiliate Marketing Disclosures on YouTube and Pinterest. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 119 (Nov. 2018), 26 pages. https://doi.org/10.1145/3274388

[72] Cristina Miranda. 2017. Can you trust that ad for a dietary supplement? Federal Trade Commission. https://www.consumer.ftc.gov/blog/2017/11/can-you-trust-ad-dietary-supplement.

[73] John Murphy. 2020. 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. Confiant.

[74] Terry Nelms, Roberto Perdisci, Manos Antonakakis, and Mustaque Ahamad. 2016. Towards Measuring and Mitigating Social Engineering Software Download Attacks. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Austin, TX, 773–789. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/nelms

[75] Casey Newton. 2014. You might also like this story about weaponized clickbait. The Verge. https://www.theverge.com/2014/4/22/5639892/how-weaponized-clickbait-took-over-the-web.

[76] N. Nikiforakis, A. Kapravelos, W. Joosen, C. Kruegel, F. Piessens, and G. Vigna. 2013. Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting. In *2013 IEEE Symposium on Security and Privacy*. IEEE, New York, NY, 541–555. https://doi.org/10.1109/SP.2013.43

[77] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI ’20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376321

[78] Abby Ohlheiser. 2016. This is how Facebook’s fake-news writers make money. Washington Post. https://www.washingtonpost.com/news/the-intersect/wp/2016/11/18/this-is-how-the-internets-fake-news-writers-make-money/.

[79] Michel Tuan Pham and Gita Venkataramani Johar. 1997. Contingent Processes of Source Identification. *Journal of Consumer Research* 24, 3 (1997), 249–265.

[80] Angelisa C. Plane, Elissa M. Redmiles, Michelle L. Mazurek, and Michael Carl Tschantz. 2017. Exploring User Perceptions of Discrimination in Online Targeted Advertising. In *26th USENIX Security Symposium (USENIX Security 17)*. USENIX Association, Vancouver, BC, 935–951. https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/plane

[81] Enric Pujol, Oliver Hohlfeld, and Anja Feldmann. 2015. Annoyed Users: Ads and Ad-Block Usage in the Wild. In *Proceedings of the 2015 Internet Measurement Conference* (Tokyo, Japan) *(IMC ’15)*. Association for Computing Machinery, New York, NY, USA, 93–106. https://doi.org/10.1145/2815675.2815705

[82] Vaibhav Rastogi, Rui Shao, Yan Chen, Xiang Pan, Shihong Zou, and Ryan Riley. 2016. Are these Ads Safe: Detecting Hidden Attacks through the Mobile App-Web Interfaces. In *NDSS*. The Internet Society, Reston, VA, 15 pages.

[83] Filipe N. Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P. Gummadi, and Elissa M. Redmiles. 2019. 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* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 140–149. https://doi.org/10.1145/3287560.3287580
[84] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. 2012. Detecting and Defending against Third-Party Tracking on the Web. In *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation* (San Jose, CA) *(NSDI'12)*. USENIX Association, USA, 12.
[85] Anne L. Roggeveen and Gita Venkataramani Johar. 2002. Perceived Source Variability Versus Familiarity: Testing Competing Explanations for the Truth Effect. *Journal of Consumer Psychology* 12, 2 (2002), 81–91.
[86] Christian Rohrer and John Boyd. 2004. 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* (Vienna, Austria) *(CHI EA '04)*. Association for Computing Machinery, New York, NY, USA, 1085–1086. https://doi.org/10.1145/985921.985992
[87] Matthew Rosenberg. 2019. Ad Tool Facebook Built to Fight Disinformation Doesn't Work as Advertised. https://www.nytimes.com/2019/07/25/technology/facebook-ad-library.html.
[88] Rahul Roy-Chowdhury. 2017. Improving advertising on the web. https://blog.chromium.org/2017/06/improving-advertising-on-web.html. Chromium Blog.
[89] Navdeep S. Sahni and Harikesh Nair. 2018. Sponsorship Disclosure and Consumer Deception: Experimental Evidence from Native Advertising in Mobile Search. https://ssrn.com/abstract=2737035.
[90] Sapna Maheshwari and John Herrman. 2016. Publishers Are Rethinking Those 'Around the Web' Ads. https://www.nytimes.com/2016/10/31/business/media/publishers-rethink-outbrain-taboola-ads.html.
[91] Sergio Serra. 2019. Ad Quality: The Other Side Of The Fence. https://www.inmobi.com/blog/2019/05/10/ad-quality-the-other-side-of-the-fence. Inmobi.
[92] Alex Shellhammer. 2016. The need for mobile speed. https://web.archive.org/web/20200723132931/https://www.blog.google/products/admanager/the-need-for-mobile-speed/. Google Official Blog.
[93] Michael Swart, Ylana Lopez, Arunesh Mathur, and Marshini Chetty. 2020. 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* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376178
[94] James Temperton. 2017. We need to talk about the internet's fake ads problem. Wired. https://www.wired.co.uk/article/fake-news-outbrain-taboola-hillary-clinton.
[95] Kaitlyn Tiffany. 2019. A mysterious gut doctor is begging Americans to throw out "this vegetable" now. But, like, which? Vox. https://www.vox.com/the-goods/2019/5/8/18537279/chum-box-weird-sponsored-links-gut-doctor.
[96] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security* (Washington, D.C.) *(SOUPS '12)*. Association for Computing Machinery, New York, NY, USA, Article 4, 15 pages. https://doi.org/10.1145/2335356.2335362
[97] G. Venkatadri, A. Andreou, Y. Liu, A. Mislove, K. P. Gummadi, P. Loiseau, and O. Goga. 2018. Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, New York, NY, 89–107. https://doi.org/10.1109/SP.2018.00014
[98] Paul Vines, Franziska Roesner, and Tadayoshi Kohno. 2017. Exploring ADINT: Using Ad Targeting for Surveillance on a Budget - or - How Alice Can Buy Ads to Track Bob. In *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society* (Dallas, Texas, USA) *(WPES '17)*. Association for Computing Machinery, New York, NY, USA, 153–164. https://doi.org/10.1145/3139550.3139567
[99] Tharindu Cyril Weerasooriya, Tong Liu, and Christopher M. Homan. 2020. Neighborhood-based Pooling for Population-level Label Distribution Learning. arXiv:2003.07406 [cs.LG]
[100] Miranda Wei, Madison Stamos, Sophie Veys, Nathan Reitinger, Justin Goodman, Margot Herman, Dorota Filipczuk, Ben Weinshel, Michelle L. Mazurek, and Blase Ur. 2020. What Twitter Knows: Characterizing Ad Targeting Practices, User Perceptions, and Ad Explanations Through Users' Own Twitter Data. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, Austin, TX, 145–162. https://www.usenix.org/conference/usenixsecurity20/presentation/wei
[101] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L. Mazurek, and Blase Ur. 2019. Oh, the Places You've Been! User Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (London, United Kingdom) *(CCS '19)*. Association for Computing Machinery, New York, NY, USA, 149–166. https://doi.org/10.1145/3319535.3363200
[102] Bartosz W. Wojdynski. 2016. The Deceptiveness of Sponsored News Articles: How Readers Recognize and Perceive Native Advertising. *American Behavioral Scientist* 60, 12 (2016), 1475–1491.
[103] Bartosz W. Wojdynski and Nathaniel J. Evans. 2016. Going Native: Effects of Disclosure Position and Language on the Recognition and Evaluation of Online Native Advertising. *Journal of Advertising* 45, 2 (2016), 157–168.
[104] Xinyu Xing, Wei Meng, Byoungyoung Lee, Udi Weinsberg, Anmol Sheth, Roberto Perdisci, and Wenke Lee. 2015. Understanding Malvertising Through Ad-Injecting Browser Extensions. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy) *(WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1286–1295. https://doi.org/10.1145/2736277.2741630
[105] Apostolis Zarras, Alexandros Kapravelos, Gianluca Stringhini, Thorsten Holz, Christopher Kruegel, and Giovanni Vigna. 2014. The Dark Alleys of Madison Avenue: Understanding Malicious Advertisements. In *Proceedings of the 2014 Conference on Internet Measurement Conference* (Vancouver, BC, Canada) *(IMC '14)*. Association for Computing Machinery, New York, NY, USA, 373–380. https://doi.org/10.1145/2663716.2663719
[106] Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. 2020. Bad News: Clickbait and Deceptive Ads on News and Misinformation Websites. In *Workshop on Technology and Consumer Protection (ConPro)*. IEEE, New York, NY, 11 pages.
[107] Wei Zha and H. Denis Wu. 2014. The Impact of Online Disruptive Ads on Users' Comprehension, Evaluation of Site Credibility, and Sentiment of Intrusiveness. *American Communication Journal* 16 (2014), 15–28. Issue 2.
[108] Paolo Zialcita. 2019. FTC Issues Rules for Disclosure of Ads by Social Media Influencers. https://www.npr.org/2019/11/05/776488326/ftc-issues-rules-for-disclosure-of-ads-by-social-media-influencers. NPR.

A 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)

B 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 8.

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 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!

C AD CONTENT CODES

Tables 7 and 8 list the content codes used to describe the semantic content of ads in Survey 2.

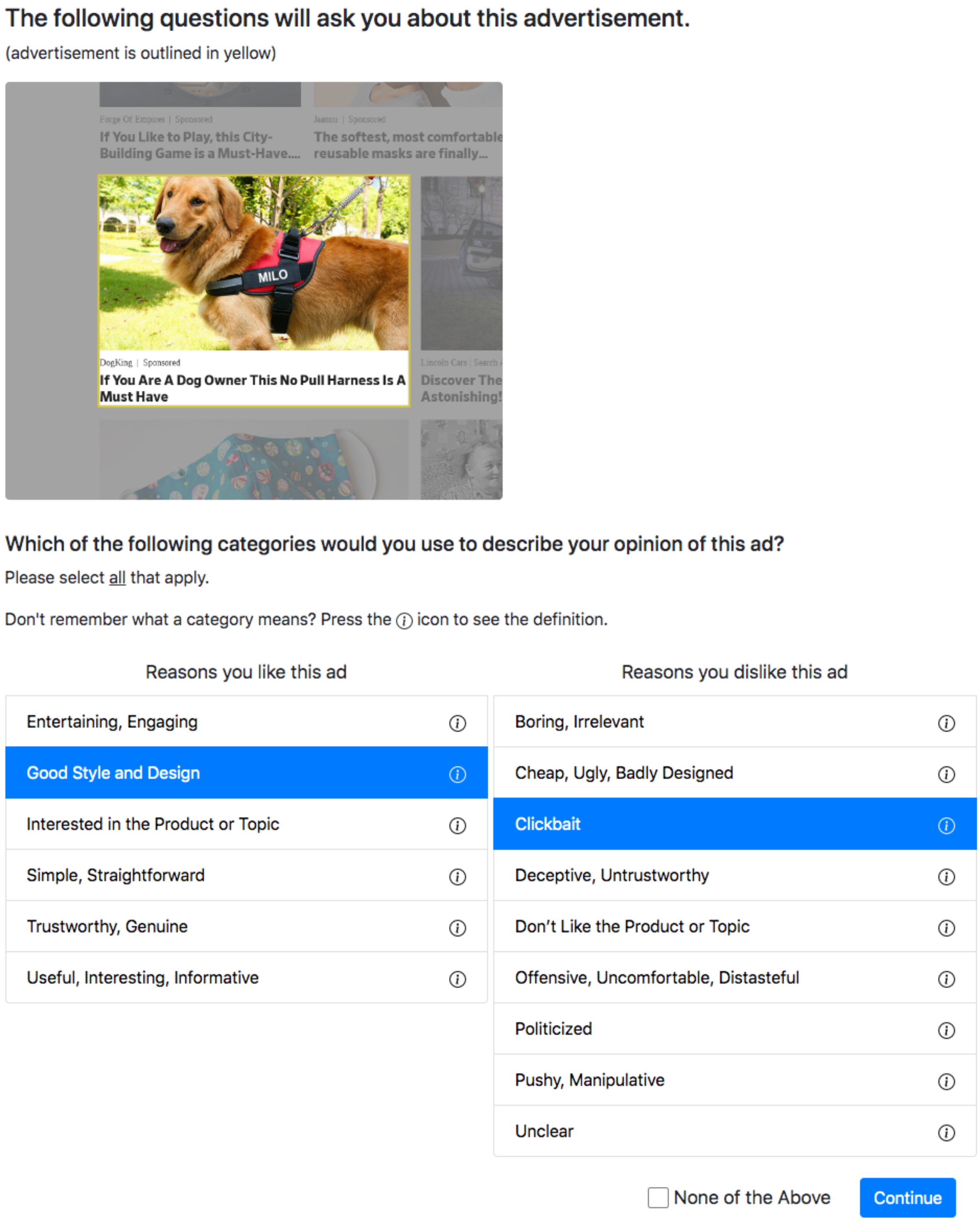

Figure 8: A screenshot of the survey interface in survey 2. For each ad, participants were able to pick multiple reasons for why they liked/disliked an ad. These responses were used as opinion labels in our analysis.

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 [40], 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 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 [74]
	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 [8].
	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 7: Content codes we (the researchers) used to label and describe our dataset of 500 ads. 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 8: Content codes we (the researchers) used to label and describe our dataset of 500 ads. Continued from previous page.