Behavioral advertising and consumer welfare: An empirical investigation[1]

Federal Trade Commission PrivacyCon 2022
Draft. Please contact the authors for the most recent version.

Eduardo Abraham Schnadower Mustri, Idris Adjerid, and Alessandro Acquisti [2]

Abstract

The value that consumers derive from behavioral advertising has been more often posited than empirically demonstrated. The majority of empirical work in this area has focused on estimating the effectiveness of behaviorally targeted ads, measured in terms of click or conversion rates. We present the results of an online within-subject experiment that, instead, employs a counterfactual approach, designed to assess comparatively consumer welfare implications of targeted behavioral advertising. Participants are presented with alternative product offers: products associated with ads they were targeted with online (targeted ad condition); competing products obtained through online search (competitor condition); and random products (random condition). The alternatives are compared along a variety of metrics, including objective measures (such as product price and vendor quality) and participants' self-reported evaluations (such as purchase intention and product relevance). We find, first, that both targeted ads and organic search results are dominated by a minority of vendors; however, targeted ads are more likely to present participants with smaller and less familiar vendors. Second, we find that purchase intentions are higher in the targeted ad and the competitor conditions than in the random condition; the effect is driven by higher product relevance in the targeted ad and competitor conditions. However, in absolute terms, product relevance is low even in the targeted ad condition. Third, we find that targeted ads are more likely to be associated with lower quality vendors, and higher prices for identical products, compared to competing alternatives found in organic search results. We discuss limitations of the current analysis and the design of a replication experiment to test the robustness of the current findings.

1. Introduction

Online advertising has become an essential part of the Internet economy. It represents 54% of total advertising spending across all media, and amounted to $380 billion worldwide in 2020 (Cramer-Flood, 2021). Online

[1] The authors gratefully acknowledge support from the MacArthur Foundation and the Sloan Foundation. COI statement: https://www.heinz.cmu.edu/~acquisti/coi.htm.
[2] Eduardo Abraham Schnadower Mustri: eschnado@andrew.cmu.edu; Idris Adjerid: iadjerid@vt.edu; Alessandro Acquisti: acquisti@andrew.cmu.edu.

targeted advertising (OTA) allows ad exchanges to tailor ads to consumers. Some of the targeting occurs by analyzing a user's past behavior while browsing online, and this type of targeting is known as Online Behavioral Advertising (OBA). A significant proportion of online display ads today are behaviorally targeted (Fisher, 2019; Samuel et al., 2021). The value that consumers derive from behavioral advertising, however, has been more often posited than empirically demonstrated. Behaviorally targeted ads tend to receive higher click-through rates than non targeted ones (Bleier & Eisenbeiss, 2015; Yan et al., 2009), suggesting that the former can reduce consumer search costs. Other than through such cost reduction, however, little is known about the manner and extent to which targeted ads affect consumers' welfare. Thus the relationship between products associated with targeted ads and other factors that also affect the utility a consumer derives from a product - such as its price, quality, or novelty - is not fully understood. For instance: How does quality compare, on average, between products shown in targeted ads to others in the market? How often do targeted ads show new products and undiscovered brands to consumers? And what about prices of products that advertisers targeted to a consumer, versus the price of a competitor the consumer may have found through searches?

The answers to these questions are not obvious, on theoretical grounds, because different (even opposite) dynamics concerning the relationship between sellers' motivation to target ads and consumer welfare are plausible. For instance, sellers may have incentives to advertise their most profitable products - which may, or may not, align with the most preferred by consumers (Hagiu & Jullien, 2011; Zhang & Katona, 2012; Acquisti, 2014). Similarly, smaller vendors with higher margins may find it more profitable to use behavioral targeting than larger vendors that sell in bulk and pass on cost savings to consumers (Chen & Stellaert, 2014) - but larger vendors may, in theory, have access to more consumer information that allows them to target individuals' preferences and needs more efficiently. In fact, depending on whether vendors are able to target by price and preference, and how differentiated consumers are across those levels, targeting may either have a positive or negative effect on consumer welfare (Marotta et al., 2021).

The answers to the above questions are not obvious on empirical grounds, either, because empirical testing of those theories is lacking. While empirical attention has been devoted in recent years to online (and specifically behavioral) advertising, most studies have focused on advertising effectiveness - that is, the effectiveness of specific offers that consumers clicked or purchased, rather than on counterfactual alternatives that compare those offers to other opportunities consumers may face in the market (Section 2). For a consumer, clicking on an ad and purchasing the advertised product carries an implicit opportunity cost: the value of an alternative product the consumer could have bought with the amount she spent on the advertised product. Existing studies, however, have not explored the differences in product characteristics between products that are targeted and other products in the market that the recipient of the ad may also have access to. Therefore we lack *comparative* investigations of different dimensions of the possible consumer welfare implications of targeted ads.

We address this gap by conducting an online within-subjects experiment with 487 participants. The experiment consists of two stages. In the first stage, we ask participants to provide URLs from ads they see while browsing randomly selected websites from their own computers and browsers. In a second stage, which takes place some days after the first, participants are shown the products from ads that had been presented to them while they were browsing (targeted ad condition), as well as competing products obtained from the organic results of online searches (competitor condition), and randomly selected products (random condition). For each product, we capture objective metrics (such as prices and vendor quality), and participants are asked questions that capture self-reported metrics including purchase intentions, perceptions of product quality, price fairness, relevance, and novelty of product type, brand, and vendor. We evaluate the consumer welfare implications of behavioral advertising in comparative terms, by comparing both objective and subjective metrics across the three conditions.

We find three key sets of results. First, both targeted ads and organic search results are dominated by a minority of vendors; however, targeted ads are more likely to present participants with smaller and less familiar vendors. Second, purchase intentions are higher in the targeted ad and the competitor conditions than in the random condition. This effect is driven by higher product relevance in the targeted ad and competitor conditions. However, in absolute terms, average product relevance is low across all conditions: products offered in targeted display advertisements to participants in our experiment are *less irrelevant* than random products, but still not highly relevant to those participants. Third, we find that targeted ads are more likely to be associated with lower quality vendors, and higher prices for identical products, compared to competing alternatives found in organic search results.

The rest of this paper is organized as follows: In section (2) we discuss related literature and the theoretical background of the investigation; in section (3) we present our experimental design; section (4) describes our empirical approach; section (5) presents the results; section (6) discusses our limitations and the design of a replication experiment to test the robustness of the current findings. In section (7) we discuss the implications of our findings and conclude.

2. Background

Our work contributes to the large body of literature on online advertising and the economics of privacy. Targeted advertising has existed in some form since before the advent of the Internet through direct channels such as mail or telephone (Roscitt & Parkett, 1988), but was not applied as deeply and thoroughly until the emergence of modern tools for online tracking. Some targeting is contextual - for instance, ads may be related to a consumer's search query or to the content of a page she is visiting. An increasing share of display ads (those that appear along the content of websites) are behaviorally targeted - that is, ads are targeted to consumers based on interests inferred from their previous browsing behavior. These ads can be negotiated directly, through human negotiation, or with a process called programmatic advertisements, in which advertisers bid automatically for a placement in the ad based on the website's data and the user's profile. It is estimated that 87% of the ad placements are negotiated in this way (Fisher, 2019; Samuel et al., 2021).

Among theoretical papers, it is generally assumed that targeted advertising brings better matching of consumers to products. However, how this affects consumer welfare depends on the nuances of each model. Some have suggested that if consumers make the voluntary decision to provide personal information to advertisers, only those who benefit from it would do so, and therefore targeted advertising should be strictly beneficial to consumers (Chen & Stellaert, 2014; Picker; 2009). However, consumers are often not fully informed about the implications of targeted advertising and may not even be aware of what is happening (Goldfarb & Tucker, 2011a). Iyer et al. (2005) show that firms can increase their profits by targeting customers with strong preference for the product as opposed to comparison shoppers, to reduce price competition. However, this reduced price competition may still lead to consumer welfare improvements despite increases in prices, due to the improved matching (Esteban & Hernandez, 2007; Gal-Or & Gal-Or, 2005). It has also been suggested that either vendors or advertising platforms may have incentives to *not* target accurately and/or show consumers less preferred options if they are more profitable (Hagiu & Jullien, 2011; Zhang & Katona, 2012; Acquisti, 2014). Amaldoss & He (2010) indicate that when consumers' valuations are low, targeted advertising may lead to *higher* prices, as those few consumers who are less price sensitive have more relative importance for sellers' profits, while high valuations cause targeted advertising to have the opposite effect, as companies compete to attract a higher volume of consumers. Others, however, have argued that targeting benefits consumers as long as the targeting is based on product preferences and not on consumer valuations of the product, since revealing reservation prices may result in an overall negative impact on consumer welfare (Marotta et al., 2021; Varian, 1996). Kshetri (2014) proposes that, by allowing consumers to buy within their affordability range using price discrimination, targeted advertising can help eliminate deadweight losses. On the other hand, Kshetri also argues that highly customized offers can be "unpleasant, creepy and frightening." Some models have incorporated such annoyance and privacy concerns. For example, Johnson (2013) indicates that whether targeting benefits or annoys consumers depends on the increase in advertising volume it can produce, and whether such increased volume annoys consumers more than it brings them value. Similarly, Gal-Or et al. (2018) suggest that, depending on how users derive value from the ads, their level of privacy concerns, and the cost of improving targeting, the equilibrium level of targeting may be higher or lower than the consumer welfare maximizing level. All these models show that the impact that targeted advertising has on consumer welfare is highly nuanced. There is no simple answer as to whether it is beneficial or detrimental for consumer welfare.

Empirical studies have linked targeted display advertisements to increased click-through rates (Bleier, & Eisenbeiss, 2015; Yan et al., 2009), purchase intentions (Van Doorn & Hoekstra, 2013; Bart et al., 2012), and purchase probability (Manchanda et al., 2006; Lewis & Reily, 2009). These effects, too, are nuanced. For instance, the positive effects of targeted ads tend to be higher for more trusted vendors but lower for less trusted ones (Bleier, & Eisenbeiss, 2015), especially if the vendors disclose their data collection practices (Aguirre et al., 2015). Furthermore, making the user feel in control over their privacy increases ad performance (Tucker, 2014). On the other hand, when an ad is both targeted and obtrusive, ad performance may suffer (Goldfarb & Tucker, 2011b). Also, if targeting is too intense, it can be perceived as intrusive and negatively affect purchase intentions (Van Doorn & Hoekstra, 2013; Bart et al., 2012). Similarly, being interrupted (which occurs frequently with display advertisements) may negatively affect consumer's attitudes towards the ad (Acquisti & Spiekerman, 2011; Duff & Faber, 2011). Furthermore, highly targeted advertisements seem to work better when the customer is further along the purchase funnel (that is, how close is the consumer to making a purchase) (Lambrecht & Tucker, 2013; Hoban & Bucklin, 2015).

The experiments described above, however, tend to observe participants' responses under a specific scenario (for instance, the effects of a specific advertising campaign), and do not tend to capture the alternatives that may have been present at the time for each participant. Thus, they do not address the question we are asking in this paper - how do products shown in targeted ads compare to random products or those that can quickly be identified via search. Two studies showed that increased targeting to specific users can increase consumer welfare (Yao & Mela, 2011; Jeziorsky & Segal, 2015) by offering a better match. These conclusions, however, came from simulations. In addition, they were performed in a different environment (search advertising) from ours (display advertising). More recent developments have taken a different perspective: observing consumer's purchasing behavior with and without the presence of ad-blockers (Frik et al., 2020; Todri, 2020). Since ad-blockers prevent ads from being displayed, the changes in purchasing behavior can be considered as an indirect measure of how well ads perform. Frik et al., (2020) found mostly no impact of ad-blockers on consumers' economic outcomes. However, the study was based on a search environment and was performed on lab computers, with no behavioral targeting. In an observational study, Todri (2020) found that users' expenditures were reduced after installing ad-blockers, and that the decrease was higher for those brands that were new to consumers. The study, however, did not separate between behavioral and other types of advertisements. It does show, however, that advertisements are an important way for new vendors to reach consumers.

In short, the empirical body of work in this area shows that behaviorally targeted advertisements have more relevance than non-targeted advertisements, and therefore can help reduce search cost. However, this work by and large leaves unanswered the question of whether products in those advertisements tend to be associated, on average, with different prices, quality, or consumer satisfaction compared to search or random product alternatives. Thus, empirical research so far does not fully capture the welfare implications of targeted advertising. Our study contributes to this literature by using a counterfactual approach in which participants are presented with alternative offers, allowing us to comparatively assess the welfare implications of behavioral advertising.

3. Experimental design

The design consists of a within-subject experiment in which participants' preferences and product offers are compared across three conditions: a targeted ad condition, where products are associated with display ads; a competitor condition, where the products presented to participants are competitors of those that appeared in the ads they were served; and a random condition, which presents participants with randomly selected products. The study was pre-registered[3] and approved by Carnegie Mellon University's IRB.

The experiment has two stages, separated by an interstage process. In Stage 1, participants are assigned to browsing a random subset of preselected websites (Section 3.1), and are asked to obtain the URLS of product pages from ads served to them on those websites (Section 3.2). During the interstage process, we use a combination of scripts and manual searches to collect information about the products associated with the ads that had been served to participants, as well as competitor products and random products (Section 3.3). In Stage 2, each participant is presented with products from the three within-subject conditions (targeted ad products, competitor products, and random products), and is asked questions that capture various variables of interest (Section 3.4). Figure 1 summarizes the process.

[3] https://osf.io/vq58p.

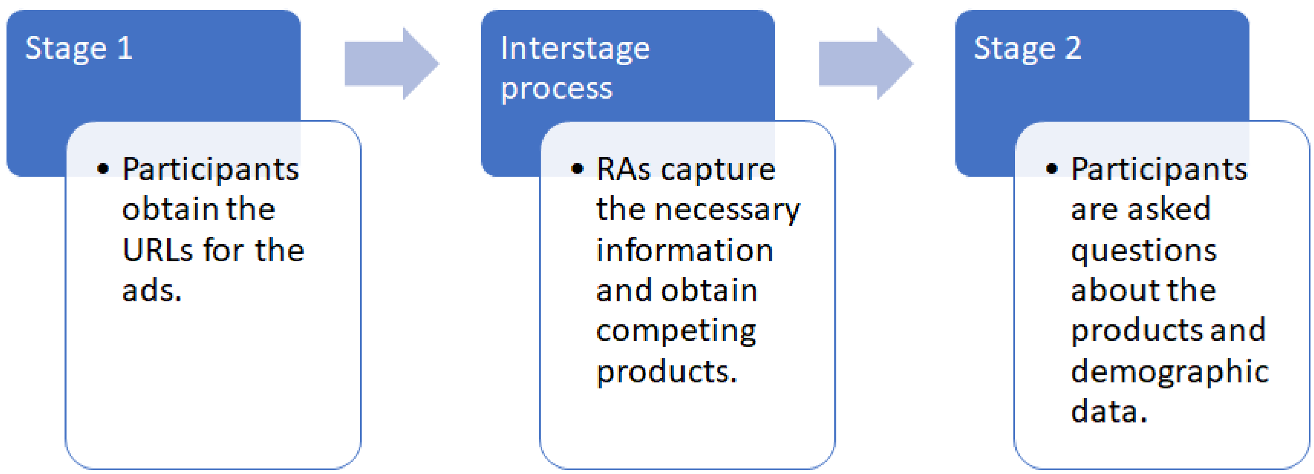

Figure 1: Summary of the experimental procedure.

3.1 Website selection process

Before we began the experiment, we identified websites that were appropriate for participants to visit and obtain ads from. The intent of the selection protocol was to expose participants during the experiment to ads with a high likelihood of being behaviorally targeted.

We focused on websites in English that serve text (as opposed to video or audio) content, cater to wide audiences, and serve behavioral ads. Our website selection process was modeled after Balebako et al. (2012). We seeded the set of sites with two popular news and entertainment websites,[4] as those websites cater to broad audiences and broad topical interests, making them more likely to serve behavioral advertisements (Balebako et al., 2012). While Balebako et al. (2012) used five news and entertainment websites, we expanded their approach to include a broader list of websites using Amazon Alexa "audience overlap" tool. The audience overlap tool provides a list of websites that have overlapping visitors with a given website. This generated two lists: one for news sites and one for entertainment sites. In addition, we included a set of smaller websites that we obtained from *Discuuver*, a random useful website generator that provides visitors with links to lesser known websites. We merged the various lists and went through each website individually. We kept in the final list only websites that showed, in our tests, to serve and display behavioral ads, provided visitors with text (as

[4] The seeding sites were www.cnn.com and www.tmz.com.

opposed to audio or video) content, and did not require login to see (at least some) content. This process produced a final list of 129 websites (see Appendix 1).

To determine whether a website frequently serves behavioral ads, each site was visited several times from different computers with different profiles. During each visit, the AdChoice icons associated with ads presented on those websites were clicked, in order to capture information about why the ad was served. Examples of the type of information obtained are shown in Figure 2. The AdChoice information is mostly available for ads served by Google. For ads that did not provide this information, we analyzed how closely the ad matched the website's content and the user's profile to determine whether the ad was behavioral. Smaller websites from *Discuuver* were added to account for the fact that some large vendors may engage in wide publicity campaigns in well-known websites, while smaller websites may have to rely more on purely behavioral targeting applied by the ad networks.

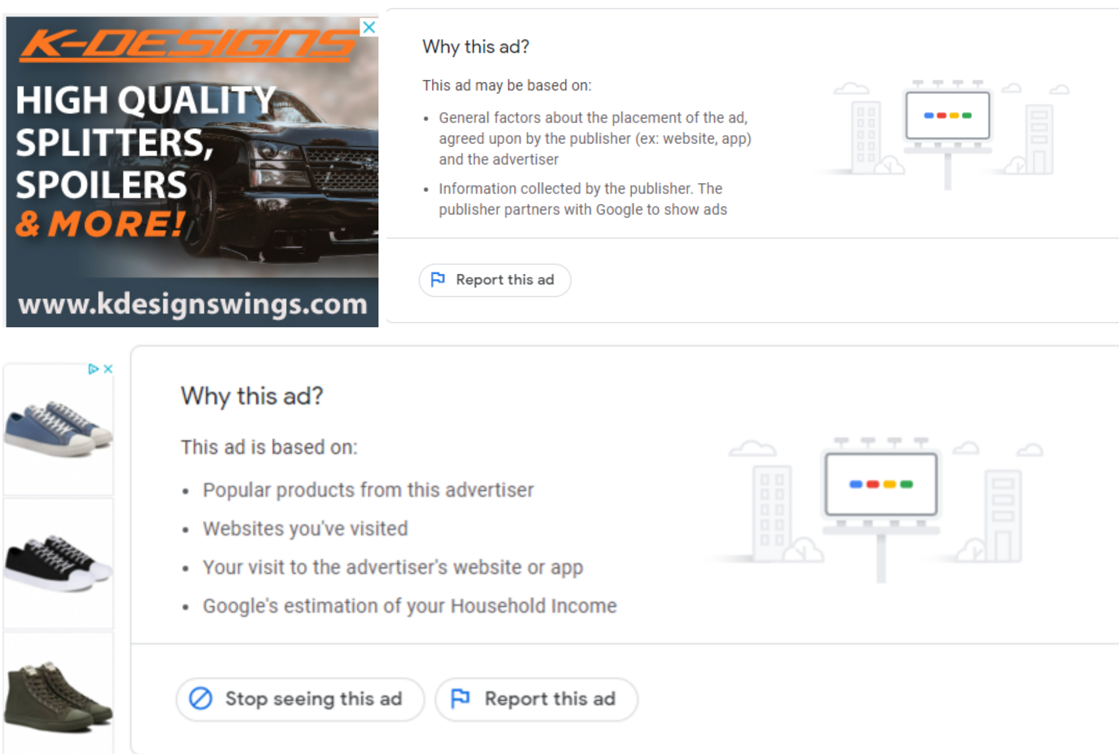

Figure 2: The ad on the top panel (on the left) appeared on an online trucking magazine. The profile used had never visited any website of the kind. The "Why this ad?" explanation suggests the ad is contextual. In the bottom panel, we see an ad for shoes in a computer that previously visited several shoes shopping sites. The information under "Why this ad?" clearly states that the ad was served based on behavioral data.

3.2 Stage 1

In Stage 1, participants visited a random subset of preselected websites and obtained the URLS of three product pages from ads served to them on those websites.

First, recruited participants were requested to fill out an online consent form. Next, participants were presented with instructions about the tasks they were expected to complete during the study, and were shown a short video tutorial that explained each step of the process. After the instructions, participants were asked attention check questions before being allowed to start performing the tasks.

The tasks consisted in participants visiting links to three randomly picked sites from the set of preselected websites. Participants were asked to visit each link with their own devices and browsers. On each website, participants were asked to obtain the product page URLs of the first ad they saw that met certain criteria that had been described to them in the instructions. The ad could take the form of a banner, a video or sponsored content within the site. However, participants were instructed to focus on physical products that could be purchased online, did not require subscriptions, preparation or customization, and could be delivered to participants' homes. (We used these criteria since those kinds of products are the ones which, in Stage 2, participants could more easily assess by looking at their image and a limited amount of additional textual information.) In addition, participants were asked to only use ads from the front page of the website, to reduce likelihood of contextual ads appearing in subsections (for example, ads for B2B services in the Business section). Participants were asked to paste those URLs into a survey form. In case they had issues obtaining a valid ad from the site, they were offered a button that would allow the request of a new randomly selected site. Figure 3 shows an example of an ad being shown on the CNN website.

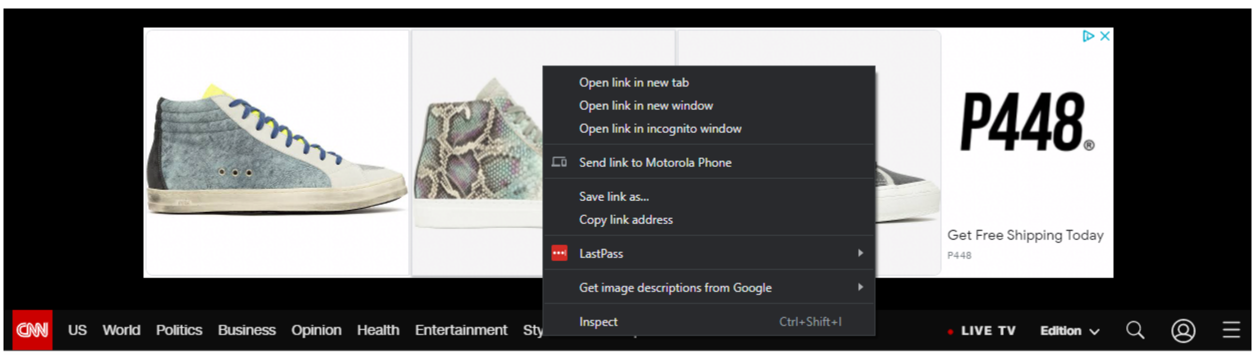

Figure 3: An advertisement on a website. The participant right-clicks on the advertisement for the menu to appear, and then must select "copy link address" to copy the URL. The participant then pastes the URL in our survey.

During Stage 1, we also automatically detected whether the participant was connecting through a VPN, Proxy or TOR, and whether they were logged in to Google.[5] Although we asked participants to disable ad blockers for the task (since otherwise they would not be able to perform it), the presence of ad blockers in the system may change the way participants are targeted when the ad blockers are paused. Similarly, the use of VPN, Proxy or TOR services may affect the quality of targeting. We captured these data to include them as controls in our analysis, since we are also interested in how these technologies change what participants observe.

3.3 Interstage process

During the interstage process, Research Assistants (RAs) collected objective information associated with the ads and with competitor products. This information forms the core of the objective measures used in the analysis.

[5] In addition, participants were asked about their previous usage of these and other technologies in Stage 2. See Section 3.4.

Immediately after each participant provided each of the three URLs, the html contents of the landing page were downloaded to a server by an automated script, and the interstage process of collecting information about the products associated with the ad landing URL commenced. Once the server downloaded the URLs and the html contents, RAs received a notification to start collecting the following information from the pages stored in the server: product image, price, description, brand and name of the website that sells the product. (If the server got blocked by the website, the RA downloaded the URL instead and uploaded a full page screenshot to the server.) If the landing page showed more than one product, the RAs obtained the information for the one in the top left. The position was picked based on the assumption that vendors sort their products to increase purchase likelihood, and therefore the first product should be the most likely to be relevant. If the landing page did not lead to an offer selling a physical product, then the link was discarded and excluded from the study. Figures 4 and 5 show examples of landing pages with a single and multiple products, respectively.

In order to ensure that this process would not introduce any bias in data collection (such as differences in the data collected by the RAs and the information that would have been shown to participants, had they visited the product page associated with the ad served to them) , we conducted a pilot test using computers in different locations. We obtained ads from a random sample of 50 of the preselected websites and captured ads from each site. Multiple testers shared in real time a link for each ad to the other testers, so that all testers could visit the link obtained from the displayed ad simultaneously but from multiple and different machines. The information that the testers could find for the product associated with the ad URL was then compared across all testers. In all cases, all testers were able to confirm that the shared URL, when visited from different locations, would lead to the same exact landing page, with the same products and the same, constant, prices. This suggests that any personalization of product and associated information (such as prices) takes place when the ad (and its URL) are *shown* to the visitor of a webpage. Once that unique URL link is created, we can be confident that what is downloaded by our server is what the user would have seen if they left-clicked on the ad instead of just copying the URL.

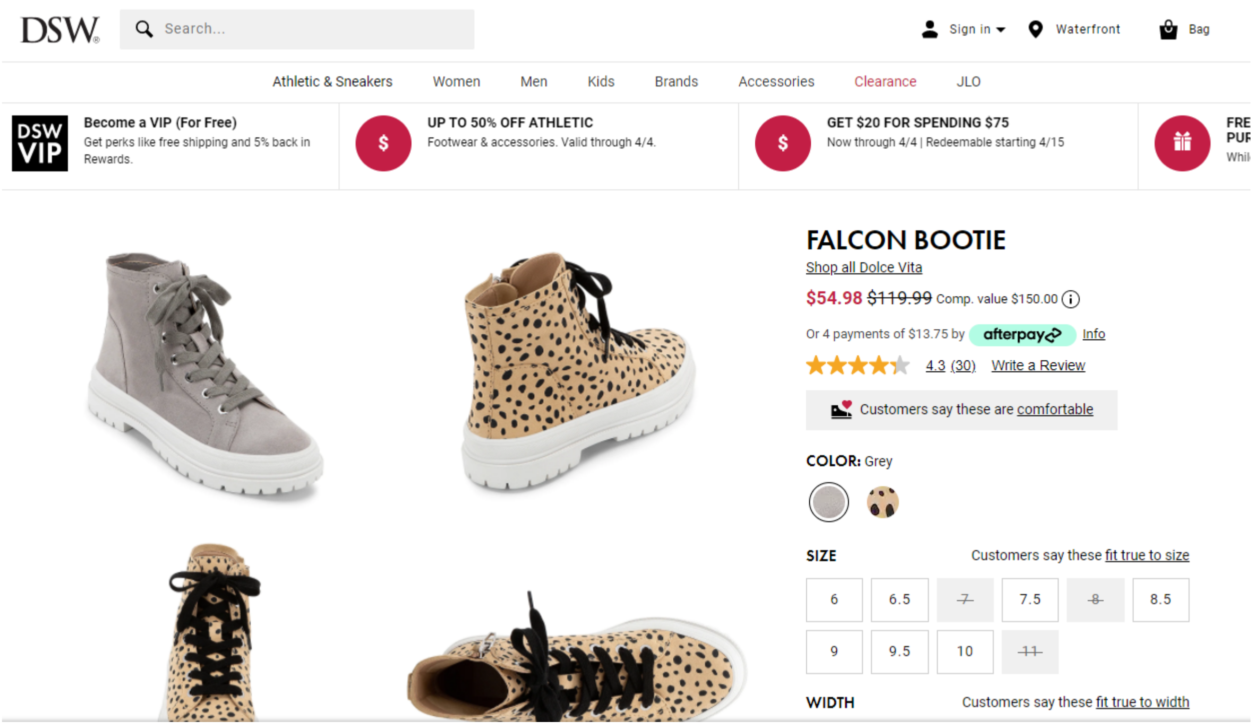

Figure 4: A landing page from an ad with a single product.

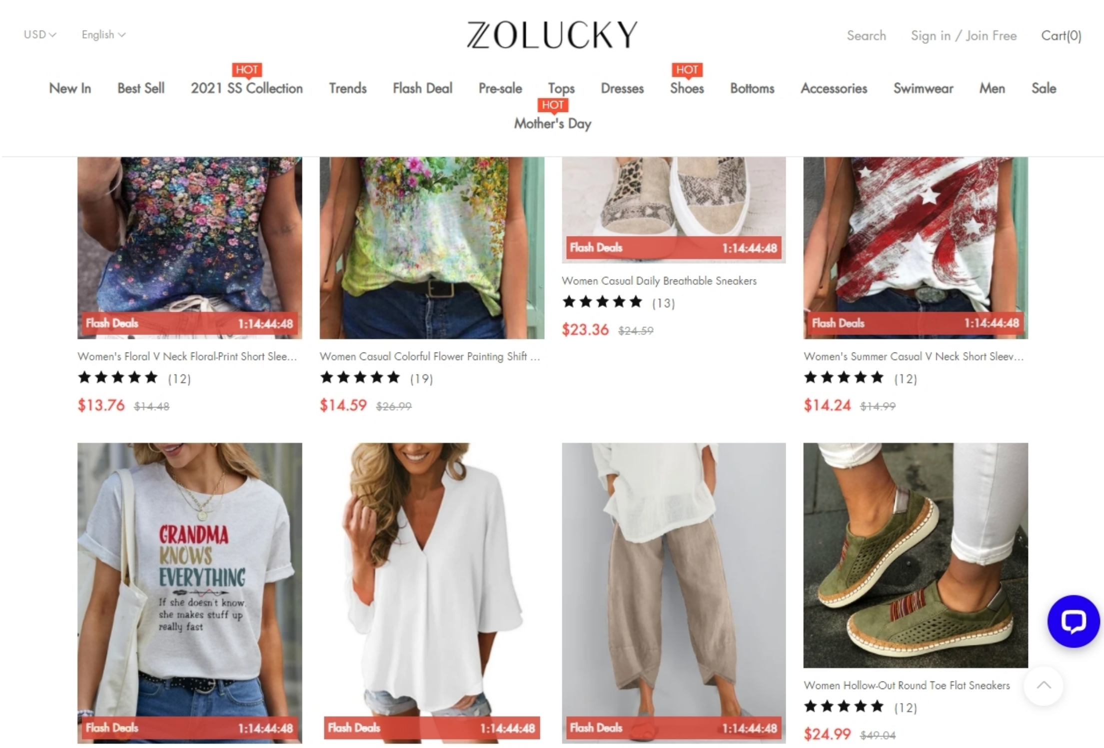

Figure 5: A landing page with multiple products. In these cases, the top left was selected to present to the participant.

Once the information for the product in the ad was obtained, the RAs obtained the competitor product by performing searches on a popular engine (Google). To find competing products, the RAs obtained the search term from the original product description. If they did not find competing products, they made incremental changes to the search terms to make them less specific and repeated the search. They did so until they obtained search results that showed offers for varied brands and vendors for competing products. Appendix 3 details the full algorithm that was followed. Once the RA landed on an appropriate page, they would collect the same information that was collected for products in the ads.

3.4 Stage 2

During Stage 2, experimental participants provided self-reported evaluations of products. This information forms the core of the subjective measures used in the analysis.

Participants who had taken part in the first stage of the study were invited to complete Stage 2 around a week later. The time lapse had two goals: adding time in-between the stages so as to reduce the likelihood of recall effects from Stage 1; but not adding so much time as to increase attrition. Each participant was presented with products from the three experimental conditions: the three products associated with the ads that had been served to him/her while he/she browsed randomly selected sites in Stage 1; three competing products found by RAs during the interstage process; and three random products - that is, products randomly chosen from the set associated with ads served to *other* participants in the experiment, and which therefore should be expected to be relevant to the preferences and interests of the latter participants but not the former (a script running on our server automatically assigned, to each participant, a random product for each valid link they provided). Participants were presented three such "triads" of products and were asked to evaluate each product independently and in random order.

For each product, the participant was shown its image, price, description, and brand - but (initially) no information about the vendor selling the product. Participants were asked a list of questions derived from the literature, and were asked to respond on 7 point likert scales. To measure purchase intention, we adapted a triad of questions from Shaouf et al. (2016) (all purchase intention questions used as endpoints *Strongly disagree - Strongly agree*):

1. "Based on the information above, I became interested in making a purchase." (PI1)
2. "Based on the information above, I am willing to purchase this product from this website." (PI2)
3. "Based on the information above, I will probably purchase this product in the next month." (PI3)

To measure other perceptions of each product, we used single likert scale questions:

4. Quality: "Based on the information above, what is the overall quality of this product likely to be?" (Extremely low - Extremely high) (Kirmani, 1990);
5. Price fairness: "How reasonable is the price for this product?" (Very unreasonable - Very reasonable) (Kwak et al., 2015);
6. Relevance: "This product might be relevant to my needs" (Strongly disagree - Strongly agree) (Laczniak & Muehling, 1993)
7. Familiarity with product type: "How familiar are you with [product type]?" (Not familiar at all - Extremely familiar) (Darley, & Smith, 1995)

8. Familiarity with brand: “How familiar are you with [brand]?” (Not familiar at all - Extremely familiar) (Darley, & Smith, 1995)

After capturing a participant’s answers to the above questions, we presented the same product information a second time, but added to it information about the vendor website. We then asked the participant to answer a modified purchase intention (PI4) question that makes reference to the website and a question about familiarity with the website:

9. “Based on the information above I will probably purchase this product from this website in the next month.” (PI4)
10. “How familiar are you with [vendor]?”

We did not include information on the website selling the product when we *first* presented it to the participant in order to isolate participants’ perceptions of the product from their perceptions of the vendor . Figure 6 shows an example of how we presented the product to the participants.

Next, we asked participants about their use of Privacy Enhancing Technologies (PET). A variety of PETs can be used by consumers to block advertisements, reduce the amount of information that can be collected about them, or disguise specific information about them such as location. These technologies affect the effectiveness of targeted advertisements, since they may prevent companies from collecting data about users. Although participants were asked to temporarily disable ad blockers for the experiment, the ads they see may be affected by the fact that they had been blocking ads before. We control for this both using our automated detection metrics (see Section 3.2) and by asking participants about their use of these technologies. The technologies we asked participants about included: 1) Browser extensions, which can be either ad blockers (prevents ads from loading) or anti trackers (prevents tracking consumers across websites) or both; 2) Networking based solutions, which can be used to disguise the location and IP Address of the user to reduce tracking. These include VPNs, Proxies and the TOR Browser; 3) Opt outs: These are provided by either individual companies or industry alliances for consumers to allow consumers to ask participating companies to not use tracking information to display advertisements, and can be effective in reducing the amount of behaviorally targeted ads (Belbako et al., 2012); specifically, we ask about opt outs for Google, Amazon, Facebook, and the DAA and NAI industry alliances; 4) Their current cookies settings: since settings within the browser related to cookies affect targeting, it is important to know if a participant has changed the cookie settings from the default settings.

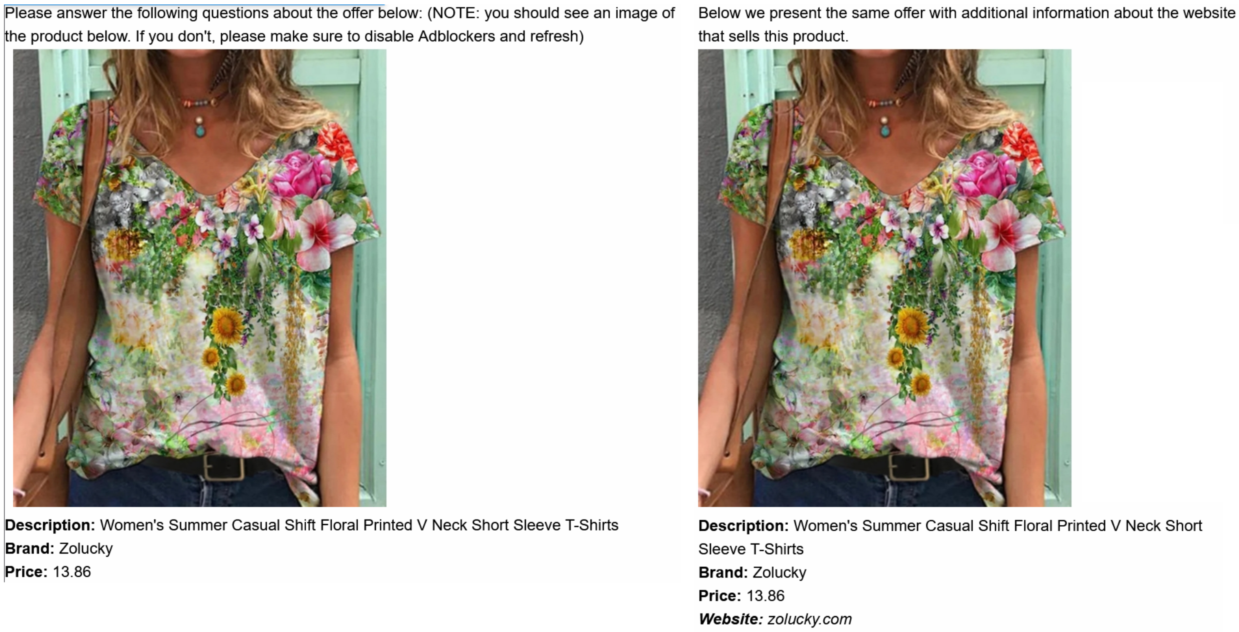

Figure 6: On the left side we observe how participants were presented with the product first without the name of the website and asked all the questions except PI4 and familiarity with the website. Then we presented the screen on the right side and asked the remaining two questions.

We also asked participants about the search engines and browsers they commonly use. Although, as further discussed below, we explicitly recruited participants that use Google Chrome as their main browser, we intended to control for potential usage of other browsers.[6]

Finally, we collected demographic information, such as age, gender, highest education level, employment and state of residence.

4. Empirical approach

Our empirical analysis focuses on two comparisons. First, we compare products associated with targeted advertisements served to a participant against competing alternatives present in the market, in order to assess whether the product presented in a targeted ad is offering a preferred option of a given good. Second, we compare products associated with targeted advertisements served to a participant against random products (products that appeared in the ad to products that appeared in ads viewed by *other* participants), so as to also assess the *relevance* of that offer comparatively. We compare products across objective and subjective metrics. Objective metrics include vendor popularity, vendor quality and prices. Subjective metrics include self-reported purchase intentions, price fairness, perceived quality of product, relevance and familiarity with brand, vendor and product type. In addition, in order to assess how the different subjective measures and their differences across conditions impact consumer welfare, we perform a latent utility analysis.

[6] It would be difficult to detect automatically whether a participant altered their browser settings, or had opted out of tracking. Hence survey responses complement our detection tools. In addition, survey questions allow us to determine whether participants regularly use those technologies even if they were not active during the time they fill out the survey.

Objective metrics. Objective metrics include vendor popularity, vendor quality and prices. Since the random products were sampled from the original advertisements seen by participants, objective metrics can only be meaningfully compared between the ads and the search results, as in expectation targeted ads and random ads should be equal. Also, these measures are uncorrelated with participants' characteristics. Hence, differences between conditions across objective metrics are assessed with simple differences in means.

Vendor popularity is obtained from SimilarWeb, from which we obtain websites rank and the number of monthly visits to each site. Vendor quality - an important element of consumer welfare - is assessed in two ways: with rankings from the Better Business Bureau (which ranks vendors using a letter system from A+ to F), and with customer rankings from SiteJabber (a popular online review platform, which is used by Google for measuring vendor quality: Google, n.d., and uses a star scale from 1 to 5 stars).

Prices are compared in two ways. First, we compare the mean prices between the ad and the search result using differences in mean of the log of price. Since the ad and the search results represent similar products, the prices are expected to be comparable. However, because we do not have an objective measure of product quality, this comparison may be affected by omitted variable bias. To address this bias we also collect, for products that were sold by more than one vendor, different prices available online for the same exact product. We then compare the price obtained from the ad with the prices found through search results. Since prices are heterogeneous, we use log transformation of prices.

Subjective metrics. We capture participants' self-reported purchase intentions for each product, as they are a frequently used measure of consumer preferences, and the empirical literature has found them to be correlated with actual purchasing behavior (Morwitz et al., 2007; Pavlou & Fygenson, 2006). Since several factors can influence purchase intentions, we also capture participants' perceptions of quality, price fairness, relevance and familiarity with product type, brand and vendor. Quality, price and relevance directly affect the utility a consumer derives from a product: quality by increasing the value to the consumer, relevance by reducing her search costs, and price as the main cost incurred.

To measure perceived quality, we ask participants to subjectively assess the quality of the product. Although participants most likely will not have any previous experience with many of the products, the expectation of quality is an essential part of the purchase decision. We do *not* use product ratings as a measure of quality, because we obtain products from several hundreds of unique websites, a majority of which do not offer these measures. In addition, they have been shown to have poor correlation with objective quality (Köcher & Köcher, 2018)[7]. Also, ratings are measured as to how well products meet or exceed expectations, rather than absolute measures of quality.[8]

[7] A simple example can illustrate this. Imagine two pairs of headphones, one from a luxury brand with very high sound quality and another from an obscure brand with average quality. However, because the expectations around the luxury brand are much higher, even while being objectively better, they may get a lower star rating than the cheap ones. At any given price, a consumer would objectively be better off with the luxury brand. However, the star rating quality would make it appear as if it is the opposite. This illustrates why online product ratings are not a good measure to use for welfare implications.

[8] This is better illustrated with an example. Suppose there is a budget laptop with low specifications with a star rating of 4.8 stars, and a high-end laptop with a rating of 4.5 stars. If we compare these two star ratings directly, one would assume that the quality of the low priced laptop is higher. In reality, any individual person, if both laptops were unexpectedly at

We measure price *fairness*, in addition to the price itself, because in many cases the prices cannot be directly compared. Consider, for example, that a product from a high prestige brand is sold five times higher than a similar product by a relatively unknown brand. Even if they are similar products, based on their expectations of quality, the consumer may consider that the price of the prestige brand is reasonable and the price of the unknown brand is not.

We capture different measures of familiarity (product type, brand, and vendor) for two purposes: first, since participants' previous experience will affect quality assessments, we need to have it as a control. Second, we want to gauge the informative value of behavioral advertising - that is, the extent to which behavioral ads bring new information to consumers. If a certain brand, vendor, or product type is new to consumers, it is obvious that they would not be familiar with them. Therefore we use familiarity also as an inverse measure of novelty.

One of the main goals of our study is to compare each of the variables we measured between our three experimental conditions. The simplest way to do that comparison is with a difference in means of each of our variables across conditions. We use Repeated Measures ANOVA for these comparisons, along with contrasts between the ad and each of the other two conditions. This allows us to control for unobserved heterogeneity across participants.

Latent utility analysis. While the comparison of means across experimental conditions may provide suggestive evidence of welfare implications of OBA, we seek to directly examine the welfare implications of targeted ads. This can be done by incorporating subjective measures into a latent utility model. As price fairness, quality, familiarity, and relevance have been shown to influence purchase intentions (PI) (Dursun et al., 2011; Laroche et al., 1996; Campbell, 1999; Alalwan, 2018), we expect the differences in purchase intentions across conditions to be impacted by differences in those variables. We expect those variables to be among the main drivers behind purchase intentions. We use those purchase intentions as a proxy for purchase behavior (Morwitz et al., 2007; Pavlou & Fygenson, 2006). We assume that our measure of purchase intentions is driven by a latent and unobserved utility. Utility takes the form:

$$U_{ij} = \beta_1 Log(Price_{ij}) + \beta_2 Quality + \beta_3 Relevance_{ij} + \beta_4 Fam_{ij} + \gamma X_i + \delta_j + Cat_{ij} + \epsilon_{ij} \quad (1)$$

Where *Fam* represents our familiarity measures, *X* is a vector if individual level controls, δ_j represents sequence level fixed effects (the sequence $j \in [1, 9]$ indicates the order in which products were presented) and *Cat* represents product category fixed effects. Since we measured purchase intentions independently for each product, this means that the choice made is between buying the product or buying nothing at all. We therefore assume that a purchase is made if the utility is greater than zero. In particular, we consider that a purchase would occur if the answer to the purchase intention question is greater than 4 (the neutral value in the likert scale). This allows us to use a logit model in which, for each observation, the participant chooses between buying or not buying the product. In this latent utility model it is assumed that the participant would buy the

---

the same price, would most likely be better off with the high-end laptop. Because of all these issues, product rating is not an appropriate measure to use in our models.

product if the utility is greater than zero. Train (2009) shows that expected consumer surplus for person *n* and choice set *J* can be estimated as:

$$E(CS_n) = \frac{1}{\alpha_n} ln\left(\sum_{j=1}^{j} e^{V_{nj}}\right) + C$$

Where *V* is the observable part of utility (that is, the fitted values of Model 1), α is the marginal utility of income and *C* is a constant that represents that absolute utility cannot be measured when a consumer chooses between alternatives. However, since in our case we observe purchase intentions for each product independent of each other, our alternative set consists of buying the product or not buying at all. By assuming that not buying has a utility of zero, we can then simplify the expected surplus for alternative j for participant i as:

$$\mathrm{E}(CS_{ij}) = \frac{1}{\alpha_{ij}} ln\left(e^{V_{ij}} + 1\right) + C \ (2)$$

Since we are not interested in the absolute value of surplus but rather in differences across conditions, *C* cancels out when we estimate such differences. Since income is not observed, we follow Train (2009) and obtain the marginal utility of income from the coefficient on price, since the price paid for a product can be considered a loss of income. As we are using log of prices, this means that:

$$\alpha_{ij} = -\partial U/\partial Price_{ij} = -\beta_1/Price_{ij} \ (3)$$

This allows us to estimate differences in consumer surplus across conditions.

5. Results

Participants were recruited using the Prolific Academic platform.[9] Participants were offered monetary compensation for their participation ($1.67 for stage 1, $2.83 for stage 2, $4.50 total). In order to participate, prospective participants were required to reside in the US, have completed at least 100 assignments on the platform, and have a 95% or higher approval rate. In addition, we used recruitment criteria to decrease the likelihood of confounds affecting the types of ads participants would be exposed to in the study. In particular, we did not want participants who would be served less behavioral ads than others because of their browser. Thus, participants were required to use Google Chrome - one of the browsers with the least restrictive default privacy settings - as their main browser. As Chrome is currently the most widely used browser (Statcounter, 2021), this restriction does not significantly limit the external validity of our results. Due to its default configuration, targeting will be the highest for Chrome users relative to users of other browsers, which should allow us to better identify effects between conditions. Participants were also requested not to use incognito mode. Incognito mode starts a temporary session without any cookies, and all cookies created during this mode are deleted after the session ends. Because of this, the information available for targeting during incognito sessions is reduced, which would limit the effects of targeting.

[9] Prolific Academic has been shown to be a reliable platform for research studies. See, for instance, Peer et al. (2017).

5.1 Participants' characteristics

Four hundred and eighty-seven participants completed the study. The mean completion time was 27 minutes. The target sample size was determined via power analysis through simulations of our empirical model. The analysis determined that 1,000 ads were required to be collected in order to obtain 95% power under several different scenarios. In Stage 1, each participant provided links to three ads, for a total of 1,461 ad links. However, 292 (20%) of links did not meet our specified criteria (physical products) and were not considered for Stage 2, producing a set of 1,169 usable advertised products, to which our experimental procedure added 1,169 competitor products and 1,169 random products, for a total of 3,507 data points.

The sample is quite diverse in terms of age (min: 18; max: 75; mean=35; sd=10.74); 41% percent of the participants are female and 58% male (5 participants chose not to answer). The sample tends to be highly educated: 84.61% of our sample has a degree and 14.37% only completed High School; 57% of the participants are full time employees and 14% are students. Two hundred and eighty participants use ad blockers or anti trackers; 82 report usage of TOR or VPN; and 247 report having used opt outs. Usage of these tools is relatively widespread: overall, 385 of our 487 participants used at least one of these technologies.

5.2 Descriptive statistics: objective measures

We classified each ad according to the category it belonged to. We identified 26 product categories, with more than half of the ads belonging to four categories: clothes, electronics, health and personal care, and footwear. This is consistent with current trends in online shopping categories (Kunst, 2021).

Vendor popularity. We detect very high market concentration for vendors in both ads and search results. We find 480 unique websites across all ads, and 458 unique websites across all search results. The top 20 websites represent 50% of the results in search and 35% of the results in ads, which is consistent with findings about increasing concentration trends in online markets (Autor, et al., 2010). However, as further discussed below (Section 5.3), targeted ads seem to serve an informational purpose, as they are more likely to present participants with offers from smaller and less familiar vendors. We obtained online ranking data and monthly activity for each website from SimilarWeb. As the number of monthly visits varies widely in orders of magnitude, we used the logarithm of monthly visits for our analyses to make them more comparable. We observe that the log monthly visits of the sites that appeared on ads ($M=15.82$, $SD=3.07$) is much lower than the search results ($M=17.77$, $SD= 3.29$): $t(2154) = -14.2068$, $p < 0.001$).[10]

Vendor quality. We collected Better Business Bureau Ratings for all vendors for which ratings were available. The Better Business Bureau publishes ratings in a scale that goes from A+ to F. They also publish the methodology for the scale (BBB, n.d.). They estimate a numeric score and then convert from such a score to the letter grade, using ranges (for example, a score from 97 to 100 is an A+). When comparing the Better Business Bureau grades between ads and search results, we found that ads have a surprisingly much higher number of incidences of the F rating as well as NR rating than search results (NR means that the Better Business Bureau does not have enough information to post a grade). This seems to indicate that websites with bad management of complaints are more prevalent in ads, while search ads have a significantly higher number of high rated websites.

[10] SimilarWeb lacked data on 96 of ads vendors and 88 search vendors.

In addition, we obtained online reviews for the vendors from SiteJabber, a popular online reviews site. From all the 1,169 ads, SiteJabber lacked data for 303. For the 1,169 corresponding search results (competitors), Sitejabber lacked data for 187. After excluding sites with missing reviews, reviews for ads (M=3.41, SD=1.09) were slightly but significantly lower than reviews for sites in search results (M=3.53, SD=0.97: (1846)=-2.62, p = 0.004).

Prices. The range of price for products in ads and searches was wide and skewed. The price range was [$0.11, $32,950.97], with a mean of 328 and a median of 44. Using paired differences in log prices, we find that on average, prices in search results were 10% lower than those in ads (t(1168)=3.58, p<0.001). This result holds even after removing outliers. In fact, 53.29% of competitor search results were priced lower than the product in the ad. Therefore, we would expect a price focused person to be able to find lower prices even more often.

Because the products are heterogenous, unobserved product characteristics may be biasing these results. We therefore repeated the analysis focusing on results for *identical* products - that is, products for which other vendors sell the exact same model shown in the offer from the ad. Of the 1,169 original ads, roughly half (609) were products that were sold by multiple vendors. For each product that was sold by more than one vendor, we found an average of 4.77 other vendors in the first page of search results. We then obtained the differences in log prices between the average price from search results and found that the price from the offer in the ad was, on average, 7.75% cheaper than the *mean* price from search results (t(608) = 6.70, p < 0.001); however, when compared against the *minimum* price available in the first page of search results, the price in the ad was, on average, 7.9% higher than that (t(608) = 5.65, p < 0.001). This means that, for products that are available from more than one vendor, there are usually lower-priced alternatives within the first page of results, although an effort may be required to find them.

5.3 Descriptive statistics: subjective measures

Table 1 shows summary results for our variables of interest. It also shows repeated measures Anova tests for differences across conditions and the t values of contrasts between the ad condition and the other conditions.

PI1-PI4 represents each of the individual purchase intention questions, PIAVG is the average of the first three purchase intention questions (those presented before the participant was provided information about the vendor websites). The purchase intention questions show high internal consistency (Chronbach's α = 0.95). We observe that the values for purchase intention are low. With 4 being the neutral value, all purchase intention measures were well below that value. Participants were, in general, not very interested in the products in any condition. Note that the phrasing of PI3 ("will probably purchase this product in the next month") and PI4 ("will purchase this product from this website in the next month") implied a more concrete and immediate intention to purchase than the first two. Accordingly, the means for PI3 and PI4 were considerably lower than those for PI1 and PI2. That noted, behavioral ads were associated in relative terms with higher purchase intentions than the random condition, but were only significantly different to the random condition in PI3 and PI4, and the average of the first three. We also see that, after revealing the website that sells the product, participants did not appear to change their purchase intentions much. We can observe this from the very similar values between PI3 and PI4.

Product relevance for the random condition is much lower than both search and behavioral ad. However, products are, on average, not very relevant to participants - even in the behavioral ads condition. In addition, the relevance between the ad and competitor conditions is very similar. This result suggests that behavioral ads are more relevant than random products, but that knowing only the category of products that might interest consumers is sufficient to achieve this lift in relevance.

Perceptions of price fairness and quality are in general above the neutral point in the respective scales, and thus positive. Both the ad and competitor condition have values for perceived price fairness significantly higher than 4 ($p < 0.01$ for both cases). Price fairness is significantly higher for the competitor condition. Quality perceptions are positive, and significantly above neutral ($p < 0.01$ for all conditions), and significantly higher for the ad than any of the other two conditions.

Participants were in general familiar with the product types (all averages well above 4, the neutral point on the Likert scale). As expected, they were less familiar with product types in the random condition. This is expected, as the presence of behavioral targeting makes it more likely for a participant to see familiar product types. Participants were only vaguely familiar with the brands, but we observed that ads offered more familiar brands than either of the other two conditions. Familiarity with vendors was similarly low (below 4) in the ad and random conditions, but was high for the competitor condition. This shows that, as expected, vendors in search results are in general better known to participants. All three familiarity measures were lower in the random condition than in either of the other two. This is also expected, as participants saw products that were targeted to others and therefore less likely to have previous experience with those products. Familiarity with the brand was higher for the behavioral ad than either of the other two conditions. This seems to suggest that participants are finding more products from brands they already know in the ads, while the search is offering more unknown brands.

In summary, purchase intentions are low, but the lowest for random products. Price fairness is higher for competitors, and perceived quality is higher for the ads (which may be due to higher brand familiarity). Familiarity in general was high for product types and low for brands and vendors. Participants were less familiar with product types in random ads than the other conditions, more familiar with the brand for the ad condition, and more familiar with the vendor in the competitor condition, suggesting that the ads present consumers with more novel and less familiar vendors.

The latent utility analysis is ongoing.

Variable/Type	Ad		Competitor		Random		RMANOVA	Contrast T	Contrast T
	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev	F	Ad vs Competitor	Ad vs Random
PI1 (interested in making a purchase)	3.28	1.96	3.20	1.92	2.96	1.90	7.23**	-0.08 (0.06)	-0.33** (0.06)
PI2 (willing to purchase this product)	3.30	1.97	3.24	1.95	2.99	1.92	12.54**	-0.06 (0.06)	-0.30** (0.06)
PI3 (will probably purchase this product in the next month)	2.65	1.77	2.54	1.73	2.38	1.68	13.78**	-0.10* (0.05)	-0.27** (0.05)
PI4 (will probably purchase this product from this website in the next month)	2.63	1.76	2.53	1.71	2.33	1.68	18.14**	-0.10* (0.05)	-0.30** (0.05)
Average of PI1, PI2, PI3	3.07	1.81	3.00	1.76	2.77	1.73	15.53**	-0.08 (0.06)	-030** (0.06)
Price Fairness	4.40	1.86	4.55	1.80	4.29	1.82	6.99**	0.15* (0.07)	-0.11 (0.0.07)
Perceived Quality	4.87	1.24	4.73	1.22	4.67	1.19	11.72**	-0.14** (0.04)	-0.20** (0.04)
Relevance	3.93	2.00	3.89	1.99	3.55	1.96	19.91**	-0.05 (0.06)	-0.43** (0.11)
Familiarity with product type	4.79	1.88	4.73	1.88	4.43	2.05	20.36**	-0.06 (0.06)	-0.37** (0.06)
Familiarity with brand	3.18	2.30	2.86	2.22	2.73	2.18	16.02**	-0.33** (0.08)	-0.45** (0.08)
Familiarity with vendor	3.53	2.49	4.37	2.51	2.97	2.39	107.54**	0.84** (0.10)	-0.55* (0.10)

*Table 1: Summary stats for all our variables of interest. 1 = Strongly Disagree/Extremely Unfamiliar/Very Low Quality/Very Unreasonable, 7 = Strongly Agree/Extremely Familiar/Very High Quality/Very Reasonable. The last three columns show the Repeated measures Anova for differences across conditions and differences in means for the contrasts. Standard errors in parenthesis. * p<0.05, **p<0.01*

5.5 Robustness checks

It is possible that our results are affected by a recall effect, that is, that right clicking on the ad can affect the perceptions of participants later on. Although most of our participants completed stage 2 at least a week after they did stage 1, we allowed 53 of them to perform stage 1 closer to stage 2, in order to be able to estimate how the time between both stages affects our models. In addition, several participants took more than one day to do stage 2 after they were invited. Overall, the mean number of days between stage 1 and stage 2 is 12.60 (14.16 standard deviation). This allows us to have a sample over a wide number of days. If there is a recall effect, the closer between the two stages, the higher the impact on the variable. In particular, we would expect this effect, if it exists at all, to be present in the ad condition and not in the other conditions, and could affect specifically purchase intentions, brand familiarity, vendor familiarity and relevance. We estimated the following model:

$$Y_i = \alpha + \beta_1 Competitor + \beta_2 Random + \gamma_1 Days \times Ad + \gamma_2 Days \times Competitor$$
$$+ \gamma_3 Days \times Random + \delta X_i + \xi_j + \epsilon_{ij}$$

Where Days represents the number of days between both stages, and we allow different slopes for each of our conditions. Table 2 shows the results. We tried in addition to the model above, adding squared terms to account for a potential non-linear relationship between time and recall, but as the results were similar, we did not include them for brevity. It is important to note that in this model we are only interested in the coefficients for the interaction between time and the ad. The main effects mainly represent the values of those variables when there are zero days between stage 1 and stage 2, which is not meaningful since we always waited at least one day. Because participants were not shown the competitor or the random product on stage 1, any apparent correlation between days and the other conditions would be spurious. We observe that in all four cases, the effect of days on the perception of the ad is small and not significant.

Coefficient\DV	(1) PIAVG	(2) Relevance	(3) Brand Familiarity	(4) Vendor Familiarity
Competitor	-0.161	-0.0882	-0.526***	0.805***
	(0.0841)	(0.0845)	(0.113)	(0.175)
Random	-0.445***	-0.495***	-0.470**	-0.489**
	(0.0976)	(0.108)	(0.145)	(0.150)
Ad*Days	-0.0111	-0.00503	-0.0120	-0.00249
	(0.00610)	(0.00702)	(0.00691)	(0.00732)
Competitor*Days	-0.00414	-0.00125	0.00513	0.00131
	(0.00576)	(0.00583)	(0.00575)	(0.0109)
Random*Days	0.00108	0.00518	-0.00793	-0.00440
	(0.00559)	(0.00671)	(0.00673)	(0.00758)
Constant	6.120***	5.322***	5.160***	5.275***
	(0.430)	0.646	(0.897)	(0.760)
N	3507	3507	3507	3507

Standard errors in parenthesis
* p < 0.05, ** p < 0.01, *** p < 0.001

Table 2: Regression models to measure recall effect. Standard errors in parenthesis. All models use robust standard errors clustered by participants.

6. Limitations

An important limitation in our experiment is that we use purchase intentions, which may not necessarily translate always into actual purchases. However, studies have shown that purchase intentions are in fact a good proxy (Morwitz et al., 2007; Pavlou & Fygenson, 2006).

Another limitation is that answers on the use of Privacy Enhancing Technologies are self reported, and therefore may be inaccurate. However, we were able to capture whether some of them were active during the study. Specifically, we were able to detect VPN, TOR and users being signed in to Google. We are unable to confirm automatically the use of incognito mode, opt outs or changes in the user's browser settings. We tried to obtain automatic measures for the presence of ad blockers, but they failed in the middle of the study and we were not able to incorporate them. For those that we detect, it was still necessary to ask participants since failure on our side to detect it would not mean that they have never used it, they could have disabled it temporarily before accessing the study. We therefore incorporate both self reported and automatically obtained measures into our dataset. Also, several tracking techniques may be used to track participants, such as cookies, tracking pixels, browser fingerprints and IP Addresses, each with different levels of accuracy. We are not able to assess which tracking technique was used, and therefore cannot identify the efficacy of these techniques.

Also, in order to reduce heterogeneity, our experiment was done with display advertisements that appear on desktop computers when visiting text content websites where log-in is not required. Our results may not extend to Social Networking websites or other types of websites where the user is required to be logged in, since those websites can collect more precise data about the consumer. They may also not extend to websites whose content is mainly audio or video. Since mobile platforms may use geolocation and other additional forms of data collection not available on Desktop, as well as using a different format, our results also may not extend to mobile platforms. These issues can be addressed in future experiments by focusing on those platforms.

Furthermore, our data is not able to measure search costs directly. However, since our searches involved randomly selecting from the first page of search results, we assume that the search costs for finding the products we showed in the "competitor" condition are relatively low, and therefore it does not significantly affect the validity of our results.

A pre-registered replication study is ongoing, to test the robustness of the current findings.

7. Discussion

Our results suggest that products offered in targeted display advertisements are, on average, not highly relevant to consumers, but *less* irrelevant than random products. We also observed that search results offer better prices, but ads in general offer more familiar brands. In terms of welfare, participants were significantly better off with targeted products than with random products, but they did not improve over randomly selecting from a search for a similar product. We also showed that there is a pervasive problem of low quality vendors in advertisements. These low quality vendors tend to offer lower prices, but less relevant products that also have lower quality perceptions.

We also found that ads tend to offer higher prices for the *same identical product* than what can be obtained from a search. However, since the *average* price from search results tends to be higher, this means a certain effort (that is, a search cost) is required in order to find the minimum price. For those consumers who are price focused and have a search cost that is low relative to the price of the product, a search will, on average, make them better off.

On the other hand, our results also show that competitors, which were obtained from organic search results, had better known vendors than in either of the other conditions, and although vendor familiarity significantly affects purchase intentions, the effect is relatively small. This implies that display advertisements can be a useful channel for new entrants to the market and smaller vendors and brands to be able to showcase their products to consumers, since it may be more difficult for them to appear in organic search results. In summary, higher relevance and the access to new vendors seems to be the main benefit that we observe from behavioral advertising.

The presence of low quality vendors, along with the recent increase in the use of ad blockers, makes it increasingly difficult for new, high quality vendors, to reach new clients. Consumers benefit from having access to new sellers that are able to meet their needs through behavioral ads, as long as they are good sellers. In the long run, the massive use of ad blockers may be detrimental to consumers as competition will be stifled by the difficulty that new entrants and smaller vendors have in showing up in organic search.

As some consumers are annoyed by the presence of targeted advertisements, their intrusiveness and all the privacy implications, it is increasingly important for advertising platforms to provide mechanisms that are less privacy invasive and for publishers to make ads less annoying, so that there are less incentives to install ad blockers. Improving filters to increase vendor quality in display ads is also essential, as having a negative experience with a vendor from this kind of ads may decrease the chances for clicks to ads in the future and also increase motivation to install an ad-blocker.

References

Acquisti, A. (2014). Inducing customers to try new goods. Review of Industrial Organization, 44(2), 131-146.

Acquisti, A., & Spiekermann, S. (2011). Do interruptions pay off? Effects of interruptive ads on consumers' willingness to pay. Journal of Interactive Marketing, 25(4), 226-240.

Aguirre, E., Mahr, D., Grewal, D., De Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. Journal of retailing, 91(1), 34-49.

Alalwan, A. A. (2018). Investigating the impact of social media advertising features on customer purchase intention. International Journal of Information Management, 42, 65-77.

Amaldoss, W., & He, C. (2010). Product variety, informative advertising, and price competition. Journal of Marketing Research, 47(1), 146-156.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. The Quarterly Journal of Economics, 135(2), 645-709.

Balebako, R., Leon, P., Shay, R., Ur, B., Wang, Y., & Cranor, L. (2012, May). Measuring the effectiveness of privacy tools for limiting behavioral advertising. Web.

Bart, Y., Stephen, A. T., & Sárváry, M. (2012). A Field Study of the Determinants of Mobile Advertising Effectiveness. INSEAD.

Better Business Bureau (BBB). (n.d.). Overview of BBB Ratings: Better Business Bureau®. Retrieved May 05, 2021, from https://www.bbb.org/overview-of-bbb-ratings

Bleier, A., & Eisenbeiss, M. (2015). Personalized online advertising effectiveness: The interplay of what, when, and where. Marketing Science, 34(5), 669-688.

Campbell, M. C. (1999). Perceptions of price unfairness: antecedents and consequences. Journal of marketing research, 36(2), 187-199.

Chen, J., & Stallaert, J. (2014). An economic analysis of online advertising using behavioral targeting. Mis Quarterly, 38(2), 429-A7.

Cramer-Flood, E. (2021, November 23). Worldwide Digital ad spending year-End update. Insider Intelligence. Retrieved December 9, 2021, from https://www.emarketer.com/content/worldwide-digital-ad-spending-year-end-update.

Darley, W. K., & Smith, R. E. (1995). Gender differences in information processing strategies: An empirical test of the selectivity model in advertising response. Journal of advertising, 24(1), 41-56.

Duff, B. R., & Faber, R. J. (2011). Missing the mark. Journal of Advertising, 40(2), 51-62.

Dursun, I., Kabadayı, E. T., Alan, A. K., & Sezen, B. (2011). Store brand purchase intention: Effects of risk, quality, familiarity and store brand shelf space. Procedia-Social and Behavioral Sciences, 24, 1190-1200.

Esteban, L., & Hernandez, J. M. (2007). Strategic targeted advertising and market fragmentation. Economics Bulletin, 12(10), 1-12.

Fisher, L. (2019, April 25). US programmatic ad SPENDING Forecast 2019. Retrieved April 01, 2021, from https://www.emarketer.com/content/us-programmatic-ad-spending-forecast-2019

Frik, A., Haviland, A., & Acquisti, A. (2020). The impact of ad-blockers on product search and purchase behavior: A lab experiment. In 29th {USENIX} Security Symposium ({USENIX} Security 20) (pp. 163-179).

Gal-Or, E., & Gal-Or, M. (2005). Customized advertising via a common media distributor. Marketing Science, 24(2), 241-253.

Gal-Or, E., Gal-Or, R., & Penmetsa, N. (2018). The role of user privacy concerns in shaping competition among platforms. Information Systems Research, 29(3), 698-722.

Goldfarb, A., & Tucker, C. E. (2011a). Online advertising, behavioral targeting, and privacy. Communications of the ACM, 54(5), 25-27.

Goldfarb, A., & Tucker, C. (2011b). Online display advertising: Targeting and obtrusiveness. Marketing Science, 30(3), 389-404.

Google. (n.d.). About seller ratings ads extensions. Retrieved May 08, 2021, from https://support.google.com/google-ads/answer/2375474?hl=en

Hagiu A, Jullien B (2011) Why do intermediaries divert search?RAND Journal of Economics42(2):337–362.

Hoban, P. R., & Bucklin, R. E. (2015). Effects of internet display advertising in the purchase funnel: Model-based insights from a randomized field experiment. Journal of Marketing Research, 52(3), 375-393.

Iyer, G., Soberman, D., & Villas-Boas, J. M. (2005). The targeting of advertising. Marketing Science, 24(3), 461-476.

Jeziorski, P., & Segal, I. (2015). What makes them click: Empirical analysis of consumer demand for search advertising. American Economic Journal: Microeconomics, 7(3), 24-53.

Johnson, J. P. (2013). Targeted advertising and advertising avoidance. The RAND Journal of Economics, 44(1), 128-144.

Kirmani, A. (1990). The effect of perceived advertising costs on brand perceptions. Journal of consumer research, 17(2), 160-171.

Köcher, S., & Köcher, S. (2018). Should we reach for the stars? Examining the convergence between online product ratings and objective product quality and their impacts on sales performance. Journal of Marketing Behavior, 3(2), 167-183.

Kshetri, N. (2014). Big data′ s impact on privacy, security and consumer welfare. Telecommunications Policy, 38(11), 1134-1145.

Kunst, A. (2021, January 25). Online purchases by category in the United STATES 2020. Retrieved May 03, 2021, from https://www.statista.com/forecasts/997093/online-purchases-by-category-in-the-us

Kwak, H., Puzakova, M., & Rocereto, J. F. (2015). Better not smile at the price: The differential role of brand anthropomorphization on perceived price fairness. Journal of Marketing, 79(4), 56-76.

Laczniak, R. N., & Muehling, D. D. (1993). The relationship between experimental manipulations and tests of theory in an advertising message involvement context. Journal of Advertising, 22(3), 59-74.

Lambrecht, A., & Tucker, C. (2013). When does retargeting work? Information specificity in online advertising. Journal of Marketing research, 50(5), 561-576.

Laroche, M., Kim, C., & Zhou, L. (1996). Brand familiarity and confidence as determinants of purchase intention: An empirical test in a multiple brand context. Journal of business Research, 37(2), 115-120.

Lewis, R. A., & Reiley, D. H. (2014). Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo!. Quantitative Marketing and Economics, 12(3), 235-266.

Manchanda, P., Dubé, J. P., Goh, K. Y., & Chintagunta, P. K. (2006). The effect of banner advertising on internet purchasing. Journal of Marketing Research, 43(1), 98-108.

Marotta, V., Zhang, K., & Acquisti, A. (2021). The welfare impact of targeted advertising. Information Systems Research. Forthcoming.

Morwitz, V. G., Steckel, J. H., & Gupta, A. (2007). When do purchase intentions predict sales?. International Journal of Forecasting, 23(3), 347-364.

Pavlou, P. A., & Fygenson, M. (2006). Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior. MIS quarterly, 115-143.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. Journal of Experimental Social Psychology, 70, 153-163.

Picker, R. C. (2009). Online advertising, identity and privacy. U of Chicago Law & Economics, Olin Working Paper, (475).

Roscitt, R., & Parket, I. R. (1988). Direct marketing to consumers. Journal of Consumer Marketing.

Samuel, A., White, G. R., Thomas, R., & Jones, P. (2021). Programmatic advertising: An exegesis of consumer concerns. Computers in Human Behavior, 116, 106657.

Shaouf, A., Lü, K., & Li, X. (2016). The effect of web advertising visual design on online purchase intention: An examination across gender. Computers in Human Behavior, 60, 622-634.

Statcounter. (2021, February). Browser market share worldwide. Retrieved March 31, 2021, from https://gs.statcounter.com/browser-market-share

Todri, V. (2020). The Impact of Ad-blockers on Online Consumer Behavior. Marketing Science (Forthcoming).

Train, K. (2009). Discrete Choice Methods with Simulation (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511805271

Tucker, C. E. (2014). Social networks, personalized advertising, and privacy controls. Journal of marketing research, 51(5), 546-562.

Varian HR (1996) Economic aspects of personal privacy. Privacy and Self-Regulation in the Information Age (National Telecommunications and Information Administration, Washington, DC)

Van Doorn, J., & Hoekstra, J. C. (2013). Customization of online advertising: The role of intrusiveness. Marketing Letters, 24(4), 339-351.

Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y., & Chen, Z. (2009, April). How much can behavioral targeting help online advertising?. In Proceedings of the 18th International Conference on the World wide Web (pp. 261-270).

Yao, S., & Mela, C. F. (2011). A dynamic model of sponsored search advertising. Marketing Science, 30(3), 447-468.

Zhang K, & Katona Z (2012) Contextual advertising.Marketing Science. 31(6):980–994