Attention Spillovers from News to Ads: Evidence from an Eye-Tracking Experiment*

Andrey Simonov† Tommaso Valletti‡ Andre Veiga§

February 27, 2023

Abstract

Does online news content facilitate display advertising effectiveness? We conduct an online experiment in which subjects read various articles and are shown (randomized) ads for brands next to these articles. Using non-intrusive eye-tracking technology, we measure the attention each individual pays to each article and ad. Then, respondents are asked which ads they recall seeing, and choose between cash or vouchers for the brands advertised. We show that articles that capture more of readers' attention increase the amount of attention readers pay to ads on the page. In turn, more attention to ads increases brand recall and purchase probability. Building on the experimental results, we formulate and estimate a stylized model of attention allocation, purchase and recall. The model features spillovers of attention from articles to ads. The type of news content ("hard" versus "soft" news) does not detectably impact ad effectiveness – evidence against the practice of "block lists" of sensitive news topics by advertisers. We discuss the implications of such attention spillovers for firms' investments in captivating news content.

JEL Classification Codes: M37, C91, L86

Keywords: Online Advertising, Online News, Experiments, Attention, E-commerce

---

*We thank Daniel Ershov, Garrett Johnson, Stephan Seiler, Anna Tuchman, Gokhan Yildirim, seminar participants at the Bank of Italy, Imperial College London, CESifo Conference on the Economics of Digitization, University of Milan, University of Bologna, Cergy Paris, and Yale-TSE Regulation of Digital Economy Workshop for helpful comments. Rosa Sanchis-Guerrer and Prashant Garg provided excellent research assistance. We thank the whole team of Lumen Research for giving us access to their eye-tracking technology. We gratefully acknowledge financial help from Imperial College COVID-19 Response Fund and from the NET Institute.

†Columbia University, Hoover Institution, and CEPR, as5443@gsb.columbia.edu

‡Imperial College London, NHH and CEPR, t.valletti@imperial.ac.uk

§Imperial College London, a.veiga@imperial.ac.uk

1 Introduction

Spring 2020, the beginning of the COVID pandemic in the U.S., was characterized by an unusual dynamic for digital advertising. While visits to online media sites and news consumption increased by almost 50% [ComScore, 2020], digital advertising – a rapidly growing area of spending for companies over the last two decades [Statista, 2023] – experienced a 25-35% decline [emarketer.com, 2020]. Part of this dip is explained by the overall uncertainty of companies due to the pandemic, but it was also largely driven by "block lists": advertising companies actively avoiding placing ads on pages with pandemic-related news content [digiday.com, 2020]. Such avoidance of "hard news," i.e., news that is thought to be potentially sensitive and upsetting to some readers, is driven by perception on the part of advertisers that placing ads with negative content could lead to negative associations with their brand, hurting the brand's image and dissuading readers from purchasing the advertised product. In turn, this practice discourages news publishers from investing into "hard news" stories, leading to a potential under-provision of content that could have high societal benefits [e.g. The Guardian, 2020, IAB UK, 2020].

In this paper, we examine the effect of news content on the effectiveness of online advertising via an online experiment that uses non-intrusive eye-tracking technology. By "ad effectiveness" we mean the incremental impact of an additional second of attention to ads on product recall and purchase probability of that product. In our study, subjects were exposed to a random sequence of online articles from well-known news outlets, and the ads next to each article were also randomized. News articles were selected to cover different news topics (e.g., "hard" and "soft" news). The outlets were chosen to be of well known political leanings, which could match or clash with the political ideology of the readers. Eye-tracking allows us to directly measure the attention paid by individuals to articles and to the ads placed next to them. After reading the articles, we asked individuals to recall the advertised brands and make incentivized purchase decisions (choose between a voucher for an advertised brand and cash). Since articles vary in how interesting they are and in the topics they cover, we use experimental variation in the attention paid to each ad to determine the impact of attention on recall and purchases.

We ground our investigation in a simple model of attention allocation to articles and ads, recall, and purchase decisions. Consumers optimally allocate their attention to articles and ads on the page, based on their preferences for both. Articles and ads are allowed to generate negative or positive spillovers of attention towards each other. The attention consumers pay to ads then potentially impacts ad recall and purchase probabilities of the advertised brand.

We then estimate the parameters of this model. Based on our estimates, we find that the overall attention that readers pay to articles has a positive spillover effect on the attention readers pay to ads displayed on the page. Moreover, we find that this incremental attention to ads increases ad recall and purchase probability (i.e., the probability to choose a brand-specific voucher over a cash reward). Thus, more captivating news content – e.g. one that attracts more attention from readers – increases the effectiveness of display ads shown on the same page.

Based on our preferred specification (OLS, using the entire sample), one additional second of attention to a brand's ad results in a 3.4 percentage points higher probability of recall and 0.7 percentage points higher probability of choosing that brand's gift card over cash. The latter estimate is confirmed by an IV specification where we only use the incremental attention to ads generated by spillovers from the attention to news content.

We further show that "hard news" content – in particular articles about the COVID-19 pandemic or the Black Lives Matter (BLM) movement in the summer of 2020 – does not detectably impact ad effectiveness.[1] We find that readers spend less time on articles covering "hard news" – and because of this, pay less attention to ads shown next to hard news articles – but ad effectiveness (the effect of incremental attention to ads on recall and purchases) is not statistically different for articles with "hard" versus "soft" news. If anything, ad effectiveness is 18-43% higher (albeit not significantly different) when article content is "hard news," which is confirmed throughout all OLS and IV specifications. On balance, this higher ad effectiveness compensates for the lower amounts of attention that readers pay to ads next to hard news articles.

Our results have important implications for both news producers and advertisers. Regarding news producers, we show that the key dimension to be optimized is how captivating news content is, whereas the exact content of articles is less important. Similarly, on the advertisers' side, we show that a key metric to keep in mind when allocating display advertising is the overall engagement of users with the webpage, not necessarily the specific content on the page. In fact,

[1] We validate our categorizing of an article as hard news via an independent survey on Amazon's Mechanical Turk (AMT). An important caveat is that the types of content that we have tested are limited.

our results suggest one should revisit the practice of blunt "block lists" of hard articles, providing an opportunity for optimizing ad allocation decisions for advertisers and marketing managers.

Apart from the substantive results that we put forward, we provide a novel empirical strategy to measure advertising effectiveness using non-intrusive eye-tracking tools that have recently become more widely available. These tools allow us to run eye-tracking studies through a standard laptop or smartphone web camera, greatly reducing the costs of eye-tracking studies that are typically done in lab settings. This approach also allows for the study of how users engage with online content in a much more realistic way.

The paper is organized as follows. Section 2 reviews the literature. Section 3 details the experiment. Section 4 describes the resulting dataset and presents some descriptive statistics. Section 5 presents a stylized model of attention allocation, recall, and purchase decisions. Section 6 presents estimates of our model, in particular with regards to the determinants of attention allocation to articles and ads. Section 7 estimates the parameters of the model that inform the impact of attention on ad recall and purchase, as well as the effect of news content and the reader/content ideological alignment on ad effectiveness. Section 8 validates our eye-tracking measurements and shows the robustness of our results to a battery of alternative specifications. Section 9 offers managerial implications of the results. Section 10 concludes.

2 Related Literature

This paper contributes to the vast literature that studies the effectiveness of online advertising. Relative to that literature, we make three key contributions.

Our first contribution is to show how more captivating news content creates attention spillovers towards ads and increases ad effectiveness. Two sets of papers are closest to ours. First, this paper builds on the sub-stream of the literature that has examined how the time spent on a webpage with an ad affects the memory and ad recall of users [e.g. Danaher and Mullarkey, 2003, Goldstein et al., 2011, 2015, Uhl et al., 2020].[2] Compared to these, we use eye-tracking to explicitly show the spillover from attention to webpage content towards the ads presented. Separately

[2]Other related papers include the literature that relates online engagement and advertising effectiveness. For instance, see Kilger and Romer [2007], Calder et al. [2009].

measuring the respondents' eye-sight dwell on article text and on ads allows us to rule out reverse causality as an alternative explanation [Becker and Murphy, 1993, Tuchman et al., 2018]. We are also able to link the incremental attention users pay to ads to user willingness to pay for brands, going beyond the more upstream metric of ad recall.

Our work is also related to the eye-tracking literature that examines advertising effectiveness. However, to the best of our knowledge, there is no evidence of the effect of news content on advertising effectiveness. A sub-stream of this literature leverages eye-tracking to study the psychological mechanisms behind advertising effectiveness [e.g. Wedel and Pieters, 2000, Wedel et al., 2008, Aribarg et al., 2010, Higgins et al., 2014]. Another sub-stream studies how different features and designs of advertisements increase viewers' attention [e.g. Nixon, 1924, MacKenzie, 1986, Pieters and Wedel, 2004, Pieters et al., 2007, 2010, Lee and Ahn, 2012, Scott et al., 2016, Zhang and Yuan, 2018]. A third sub-stream discusses how viewers' involvement and familiarity with the brand (effects typically grouped by the literature as "top-down") affect attention to advertising [e.g. Treistman and Gregg, 1979, Rayner et al., 2001, Pieters and Wedel, 2007].[3] Our contribution relative to this literature is that we employ eye-tracking data to examine how readers' attention to news content spills over to the advertising presented on the same page, allowing us to measure the causal effects of news content on attention to ads and thus assess the importance of investment in high-quality engaging content.[4] We build a stylized model of attention allocation to interpret this spillover effect, and to provide a framework to separate out this effect from ad avoidance of consumers. We also link this incremental attention to ads to subsequent ad recall and willingness to pay for the advertised brands, thereby providing a needed link between the incremental visual attention and a downstream brand choice measure, called for by Wedel and Pieters [2017].[5] Our analysis is further related to Brasel and Gips [2008], Teixeira et al.

[3] Apart from these areas of inquiry related to advertising effectiveness, eye-tracking has been used in the marketing literature to further our understanding of consideration sets formation [e.g. Chandon et al., 2009], how consumers search and choose products [e.g. Russo and Leclerc, 1994, Lohse, 1997, Janiszewski, 1998, Meißner et al., 2016, Shi and Trusov, 2021], and survey design [e.g. Redline and Lankford, 2001]. More broadly, eye-tracking has been used in many fields, including marketing, psychology, and economics, to study individual choices [e.g. Camerer et al., 1993, Armel et al., 2008, Brasel and Gips, 2008, Knoepfle et al., 2009, Reutskaja et al., 2011, Brocas et al., 2014, Pärnamets et al., 2015, Ghaffari and Fiedler, 2018]. See Wedel [2015] and Wedel and Pieters [2017] for reviews.

[4] One mechanism behind the spillover of attention can be a visual distraction (e.g. Navalpakkam et al. [2011]). Such distraction has a negative effect on the consumption of news content, as shown by Yan et al. [2020].

[5] See the discussion on page 144 of Wedel and Pieters [2017]. Treistman and Gregg [1979] is the closest paper that compares the designs of two commercials and links higher attention to more sales.

[2010] who use eye-tracking data to examine the determinants of attention to TV commercials.[6]

More broadly, our work is related to other papers that have shown links between user exposure to ads and later purchase choices. Several papers link exposure to users becoming aware of the ad [e.g. Danaher and Mullarkey, 2003, Wilson et al., 2015, Elsen et al., 2016]. Other articles explore the link between exposure, awareness and purchase [e.g. Hoyer and Brown, 1990, Macdonald and Sharp, 2000, Khurram et al., 2018, Martins et al., 2019]. Another literature examines the effectiveness of online advertising on product sales using natural experiments[e.g. Rutz et al., 2012, Narayanan and Kalyanam, 2015, Jeziorski and Moorthy, 2018, Simonov and Hill, 2021] and field experiments [e.g. Lewis and Reiley, 2014, Hoban and Bucklin, 2015, Sahni, 2015, Johnson et al., 2017b,a, Simonov et al., 2018, Gordon et al., 2021a].

Our second contribution is to examine the effect of the news content (in particular, "hard" versus "soft" news), and the ideological match between the newspaper and readers' opinions, on ad effectiveness. We find that more engaging news content increases the amount of attention the reader devotes to display advertising, adding to the results on the effect of page content on ad effectiveness [e.g. Goldfarb and Tucker, 2011]. Yet, beyond the effect of devoting more attention to the news page, news content does not have any detectable additional effect on ad effectiveness. In other words, once one statistically controls for attention to the article, whether the article is "hard news" or not has no impact on purchase. This result cautions against the practice of blank blacklisting certain news content for the purposes of targeted advertising [e.g. The Guardian, 2020]. We also find that users engage more with articles and ads presented in newspapers that are more aligned with the user's own political views. Our results on the drivers of attention to online news contribute to the broader literature understanding what makes people engage with news [e.g. Holmqvist et al., 2003, Pitler and Nenkova, 2008, Kazai et al., 2016, Lagun and Lalmas, 2016, Berger et al., 2019].

The third major contribution of this article is the use of scalable and non-intrusive eye-tracking technology.[7] This technology enables efficient collection of data on eye gaze, allowing us to gather a sizeable and reliable dataset at a low cost [e.g. see Wedel, 2015]. This approach

[6] More recent studies of attention to TV ads and programs include McGranaghan et al. [2022] and Liu et al. [2021].

[7] We use software provided by Lumen Research (see Section 3), one of several companies that have recently developed such a technology.

stands in contrast to classical eye-tracking studies that use lab equipment. Several recent papers have tested the precision of other web-based eye-tracking technologies – often based on the open source Javascript library Webgazer[8] – with mixed results [e.g. Semmelmann and Weigelt, 2018, Schneegans et al., 2021, Yang and Krajbich, 2021]. The eye-tracking technology we use was developed entirely by Lumen Research, using a proprietary machine learning algorithm to adjust and calibrate the eye-gaze location data and, in particular, is not based on the Webgazer library. We validate the precision of the eye-tracking technology we employ in a series of robustness checks and find that all our results are robust. We replicate our results on both desktop and mobile devices, showing the validity of the eye-tracking technology for our task on mobile phones and adding to the list of similarities in consumer behavior on desktop and mobile devices [e.g. Ghose et al., 2013].

Our other substantive contribution is to propose an empirical strategy that uses eye-tracking to measure ad effectiveness, based on a random assignment of ads to articles of different attractiveness. We also contribute a stylized model of attention allocation that can be used to microfound consumers' decisions of what to focus on the webpage.

3 Experimental Setting

We recruited 1,013 individuals, stratified evenly across two countries (United Kingdom and United States) and two types of devices (desktop and smartphone). Respondents were recruited to match the UK/US online population in terms of age, gender, income, and location. They were recruited via a specialist supplier of research and marketing panels, Panelbase.[9]

We began the experiment by confirming the viewer's consent. At the start of the experiment, participants were told only they were a part of "an academic study about media consumption," but were not given additional details.[10] Then, participants were asked to report their age, education, income, gender, and postcode.[11]

Participants were invited to read articles from two online newspapers: *The Guardian* and

[8] See https://webgazer.cs.brown.edu/

[9] See https://www.panelbase.net/.

[10] In particular, participants were not told that the goal was to measure ad effectiveness or that they would be making choices about brands later on in the experiment.

[11] Participants were allowed to omit this information, but none did.

Daily Mail for UK participants, *The New YorkTimes* and *USA Today* in the US. In each country, we chose outlets with a wide online readership.

We presented each individual with 9 articles. All articles had been published in the short time window prior to the experiment taking place, to ensure the articles were as relevant and interesting as possible. Within each newspaper, articles were split between soft and hard news. To select the latter, we followed the advice of industry experts and focused on articles about the COVID-19 pandemic and the Black Lives Matter (BLM) protests of the summer 2020. These two topics were frequently "blacklisted" by advertisers. We provide article titles and links to the articles we used in Appendix A. We validate our categorization of articles as "hard" or "soft" news using an independent survey on Amazon Mechanical Turk (AMT), described in Appendix B. The text of the articles shown on desktop and mobile was the same.[12]

We chose ads from well-known and widely available brands. In each country, we chose 8 prominent brands listed in Appendix A. Eight out of nine articles were accompanied by ads. All ads accompanying a given article were for the same brand, inserted at fixed points along the article's page. We included one horizontal "billboard" ad before the text of the article, and two smaller " side" ads, on the side of the article text (desktop) or in-between paragraphs of the text (mobile). This is illustrated in Figure 1 below. These ad locations were chosen to mirror typical online publications. For each participant, one of the articles was randomly chosen to be shown with blank spaces in the location where ads would be otherwise shown, to obtain a baseline level of interest in each article.

In each country, each participant was exposed to all 9 articles and all 8 brands. Each article and brand was shown only once. We randomized the order in which articles were presented to individuals and the pairing between articles and brands. Individuals were allowed to read the articles at their preferred pace.

For each individual, we obtained two measures of the attention devoted to each article and ad. First, we recorded the amount of time the article and the ad were *visible* on screen. This measurement does not require eye-tracking. Second, we recorded, via eye-tracking, the time that each individual's sight dwelled on each article and ad, referred to as *dwell* time in our anal-

[12]However, in our analysis, we consider these to be different articles, since the format of the text is quite different across devices.

ysis. These are our two measures of *attention*, as we discuss in Section 4 below.

After reading all the articles, individuals were asked if they could remember the brands whose ads had been shown to them. Individuals were presented with a list containing the eight brands shown, in addition to eight "decoy" brands, in a random order. The decoy brands were chosen to be well known in each country, and of the same industries as the shown brands. All shown brands and decoy brands were presented to the participant simultaneously, and participants were asked to select which of the 16 brands shown they remembered seeing.

After the recall task, participants were asked to make purchase decisions. For each of the brands whose ads were shown, individuals were asked to choose between (i) an e-voucher worth £10 (in the UK) or \$10 (in the US) specific to one of the brands shown, or (ii) a randomized amount of cash (£3-7 in the UK and \$3-7 in the US).[13] Vouchers were valid for at least a year but could only be used at one of the brands being advertised, and this was clearly explained to participants before they chose. In sum, each individual was asked to make 8 decisions, one for each of the brands shown. Individuals were informed they would be sent one outcome of their choices, selected uniformly at random.[14] That is, consumers made *incentivized* choices, rather than merely stating their preferences in a survey. The order in which these decisions were presented to each individual was randomized.[15]

In addition to the voucher/cash reward, participants were paid a fixed participation fee. Participants were anonymous to the research team, with all payments delivered via the recruiting firm.[16] The study protocol received ethical approval prior to the start of the experiment, which was conducted at the end of July 2020.

We do not use a standard between-subjects experimental design (i.e., there is no formal control group). This is because our goal is *not* to measure the effect of the presence of ads (as opposed to the absence of ads). Instead, our goal is to study how attention to articles results in

[13]This range was chosen so that a significant share of individuals chose the voucher, based on pilots of the experiment.

[14]Participants received either a cash transaction via the recruiting firm Panelbase or received the code that activated an e-voucher.

[15]Another common incentive-compatible approach to elicit willingness to pay (WTP) is a second-price auction (i.e., the Becker-DeGroot-Marschak (BDM) mechanism). Since the experiment took place online (so researchers were not available to provide clarifications), we took the simpler approach of estimating the distribution of WTP by presenting individuals with take-it-or-leave-it (TIOLI) offers. Berry et al. [2020] show that TIOLI performs well in practice and is simpler to implement than BDM.

[16]This was done to facilitate compliance with GDPR.

attention spillovers to ads. With this aim, exogenous variation in attention to ads was induced by the random pairing of articles and ads. As we discuss further in Section 6, some articles are more interesting than others, which in turn leads participants to devote more attention to those articles, which then influences the attention devoted to the ads placed next to them. It is this exogenous variation in attention that we use to discuss the causal effect of attention on recall and purchase. This method for identifying the causal effect of attention to ads closely tracks our research question – the possible complementarity of the news content and ads – and is, to our knowledge, a novel way to measure ad effectiveness.[17]

For the purposes of this study, an online experiment provided several advantages. First, it allows for a larger data collection effort, across multiple countries and devices, at a relatively low cost. Second, we were able to show recently published articles to a large number of individuals, whereas this would have been challenging in a lab setting since, typically, few subjects can participate in lab experiments each day. Third, our experimental setting is closer to the conditions under which individuals normally engage with online content, which increases external validity.

The eye-tracking technology used was supplied by Lumen Research, a specialist advertising research agency.[18] The technology employs software that uses the camera of a desktop/mobile phone and measures where on the screen the retina of the participant is focused. No additional hardware is needed. To ensure the accuracy of tracking participants' eye gazes, in the beginning of the study, the user is taken through calibration and validation procedures, and there are two additional validation procedures throughout the study (after the third and sixth articles). If its quality is deemed to be low, the eye-tracking data is not used in the analysis. More details on the eye-tracking technology and validation steps are provided in Appendix A.

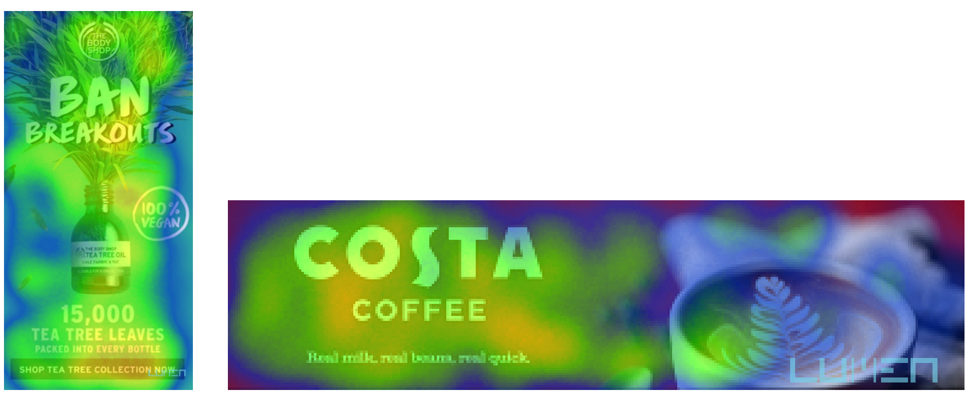
The heat map provided in Figure 1 Panel (a) is an example of how these metrics are constructed. The figure shows an article, as well as the ads (a "billboard" ad on top and two "side" ads) for one brand. The map highlights the regions on the screen that were actively dwelled upon by the participant. In Figure 1 Panel (b) we present examples of heatmaps for the ads of two different brands.

[17] Goldstein et al. [2011] also randomize pairings of articles and ads in their first study, but they force ads to always be visible on the page and do not measure attention to ads via eye-tracking.

[18] See https://lumen-research.com/.

Figure 1: Example of Heat Maps

(a) Heat Map of a Page



(b) Heat Maps of Two Ads

4 Data

4.1 Variables

Table 1 presents summary statistics of our sample. The data is at the individual×article level. About half of the observations occur on desktops (56%), correspond to female participants (55%), are from the US (48%), and correspond to articles considered to be "hard" news (55%).

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Desktop	6,431	0.563	0.496	0	1
Female	6,431	0.556	0.497	0	1
U.S.	6,431	0.483	0.500	0	1
Hard News	6,431	0.550	0.498	0	1
Article Visible (s)	6,431	143.301	169.341	20.130	1,894.635
Ad Visible (s)	5,707	19.027	17.371	0.000	291.905
Price (USD/GBP)	5,707	5.017	1.436	3.000	7.000
Recall	5,707	0.484	0.500	0.000	1.000
Buy	5,707	0.347	0.476	0.000	1.000
Article Dwell (s)	4,426	74.813	97.918	0.112	966.945
Ad Dwell (s)	3,925	2.755	3.161	0.000	40.214

Each observation is at the individual x article level.

Visibility Measures The variable *Article Visible* reports the number of seconds any part of the article was visible on screen (the sample mean is about 2 minutes and 23 seconds). The variable *Ad Visible* reports the total number of seconds that any ad on the page was visible, according to Media Rating Council standards.[19] An ad is considered visible if at least 50% of the pixels of the ad are displayed on the screen for 1 continuous second or more. The sample mean is approximately 19 seconds per article. These measures do not use eye-tracking.

Eye-tracking Measures The variable *Article Dwell* is the total time an article was actively being read, which is recorded via eye-tracking. The sample mean is about 1 minute and 15 seconds per article. Similarly, *Ad Dwell* reports the total time that all ads associated with an article were

[19]MRC is the nonprofit organization that manages accreditation for media research and rating purposes (for the industry standard definition of visibility, see http://mediaratingcouncil.org/Standards.htm).

actively looked at (that is, we sum the dwell time of the 3 ads shown on each page). The sample mean is just short of 3 seconds.

Purchase Participants were offered choices between vouchers worth $10 (£10) for each of the brands advertised and random amounts of cash. The amount of cash offered to the individuals is captured by the variable *Price*. An individual who chooses to obtain the voucher forsakes the cash, therefore the cash is the price of choosing the voucher. For about 35% of observations, individuals chose the voucher (as measured by the dummy variable *Buy*), while the rest opted for cash.

Recall For about 48% of observations, the individual correctly recalled seeing the brand (as measured by the dummy variable *Recall*). In contrast to product choices, recall was not incentivized. However, this measure is commonly used in marketing literature [e.g., Danaher and Mullarkey, 2003, Elsen et al., 2016] so it provides a useful robustness check.

Individual Demographics The sample characteristics are similar when the data is split by device type (mobile phone vs. desktop computers) and country (UK vs. US). In Appendix Tables 15 and 16, we replicate Table 1 for mobile and desktop devices separately and find the demographic composition, news types, prices, purchasing, and brand recall summaries to be very consistent. The only notable difference is that ads are more visible on desktop computers (average of 23.4 seconds) than on mobile phones (13.4 seconds). Still, ad dwell is around 2.7 seconds on average on both types of devices. Consumers also spend a bit more time reading articles on desktops (on average 83 seconds) than on mobile phones (65 seconds).

Article Characteristics The main article characteristic we consider is whether the article constitutes "hard news". As described above, these were articles focusing on the COVID-19 pandemic and the BLM protests during the summer of 2020. For some robustness checks, we also use the article's word count.

Our final dataset comprises of 6,431 observations at the individual×article level. This is less than the 9 observations per person we have originally targeted. During the study, some individuals experienced connectivity issues, and no data was recorded for around 30% of individual-article pairs. These individuals experienced the study in the same way as others, but no data was recorded as to which article they were shown at each stage in the experiment. These missing observations are slightly more prominent on mobile phones (43%) than on desktop computers (13.5%) and in later steps of the study (e.g., about 90% during the readership of the first three articles was recorded, but only around 60% in the later 6 articles). Importantly, this does not introduce a bias in our analysis, since the order in which the articles and ads were shown was randomized. We confirm that there is no selection in terms of which brands' and articles' observations have experienced the connectivity issues in Appendix Figures 9 and 10 and provide further balance checks in Appendix C.1.[20] Only 89% of observations in our main sample (5,707 out of 6,431) have data related to the advertised brands (e.g., *Price*) simply because 1 out of 9 articles did not have a branded ad shown on the page.

For a subset of participants, eye-tracking was of poor quality. High-quality eye-tracking relies on the respondent not moving their head too much, so that the software can maintain a continuous tracking of the individual's retina. If an individual moves too much, the measurements of *Article Dwell* and *Ad Dwell* for that individual-article may be deemed invalid (but *Article Visible* and *Ad Visible* are still recorded). We only include participants for whom the eye-tracking data is deemed to be high quality. This explains why we have fewer observations with eye-tracking than visibility data – around 69% both for articles (4,426 out of 6,431) and ads (3,925 out of 5,707). To identify observations with low-quality eye-tracking data, we rely on the firm that provided the eye-tracking technology (Lumen). However, we also confirm the robustness of our results by considering several measures of eye-tracking data quality. We confirm that eye-tracking data quality is higher for retained individuals and that our results are robust to alternative criteria. In Section 8, Appendices A.3 and C we describe these quality measures and the validation procedure in detail, show there is no selection in terms of which brands' and articles' observations have low-quality eye-tracking datas, and provide robustness checks on the basis of the quality of the eye-tracking data.

[20]One notable exception are three articles on mobile phones in the UK that experienced systematically more connectivity issues than other UK mobile articles (see Figure 10 in Appendix C.). We confirm that all our results hold if we exclude observations corresponding to these articles.

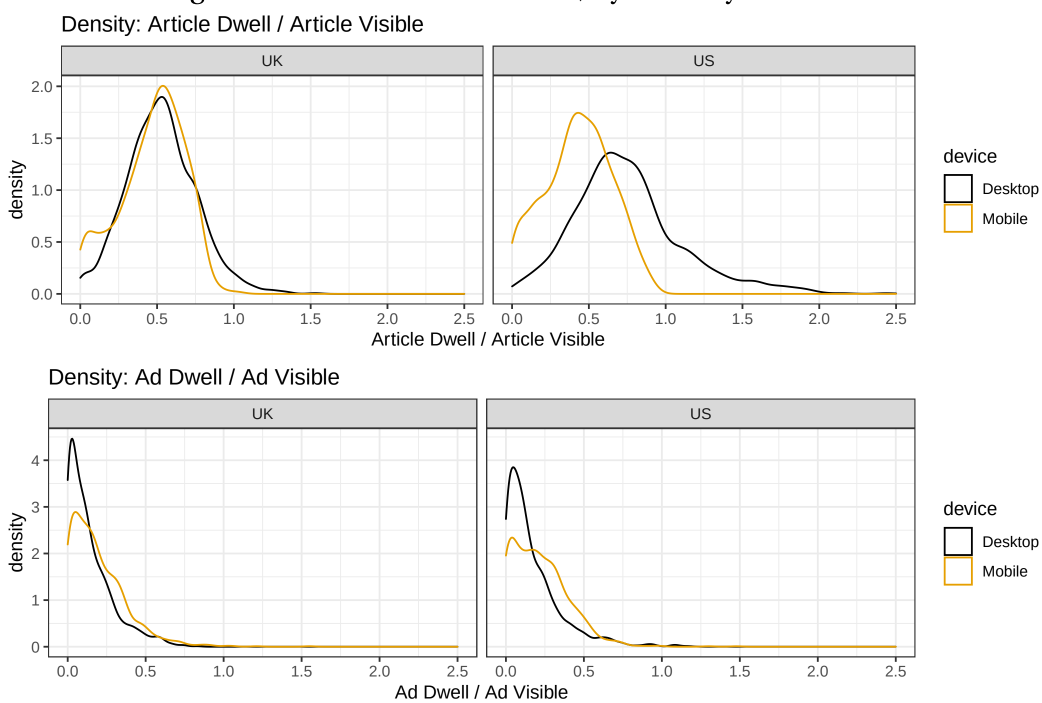
4.2 Some Descriptive Statistics

Before turning to the main analysis, we present some descriptive statistics of the data.

Distributions of Attention Measures

Figure 2 compares dwell times to whether the object is visible on the page.[21] In the top section of the figure, we present ratios of *Article Dwell* to *Article Visible* for each country and device type. On average, *Article Dwell* is around 50% of *Article Visible* , indicating that an average reader spends around 50% of their time looking at the article when the page is loaded.[22]

Figure 2: Dwell to Visible Ratios, by Country and Device



The plots show the ratio between time spend dwelling and time visible, for both articles and ads, computed across all observations in the data.

The lower part of Figure 2 presents ratios of *Ad Dwell* to *Ad Visible* . The average ratio is much lower compared to the analogous ratio for articles (around 18% instead of 50%). This agrees

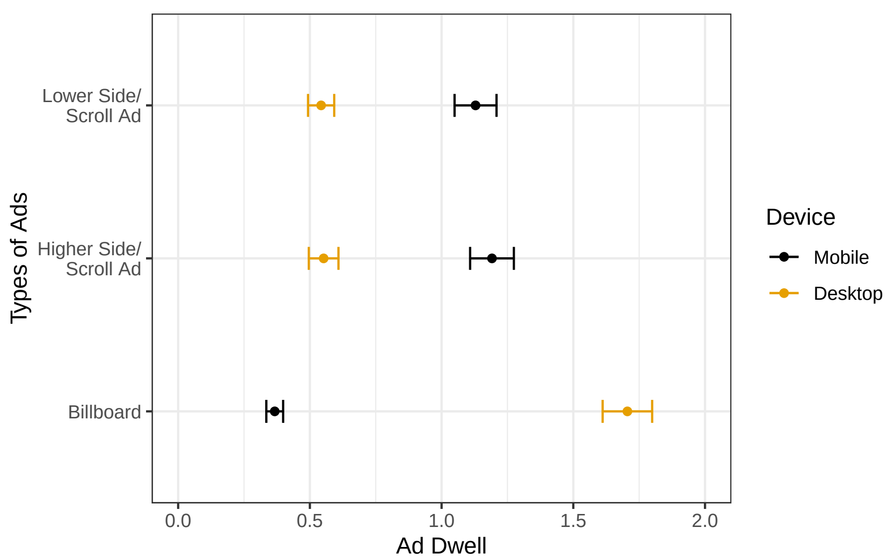
[21]Appendix Figures 15 and 16 present marginal distributions of dwell measures (*Article Dwell* and *Ad Dwell*) and visibility measures (*Article Visible* and *Ad Visible*), across all observations and on average per consumer.

[22]The average is slightly lower for mobile devices (45%) as compared to desktops (62%). This is largely explained by the desktop page design of *USA Today*, that shows only a small fraction of the article at first and therefore undercounts *Article Visible* . If we exclude USA Today articles, the average ratio of dwell-to-visible measures is 48% for mobile and 54% for desktops.

with past work that has found that TV ads can be visible for around 55% of viewers – meaning that viewers stay in the room for commercials – but only 7.7% of viewers actually pay visual attention to TV commercials [McGranaghan et al., 2022]. The ratio of *Ad Dwell* to *Ad Visible* is slightly higher for mobile devices (21%) than desktops (15%). This reflects different prominence of display ads on desktop and mobile devices, and, in particular, the difference in prominence of "side" ads: on desktops, side ads are on the right side of the page, visible but easy not to pay attention to, whereas on mobile phones they occupy blocks between the text in the center of the screen.

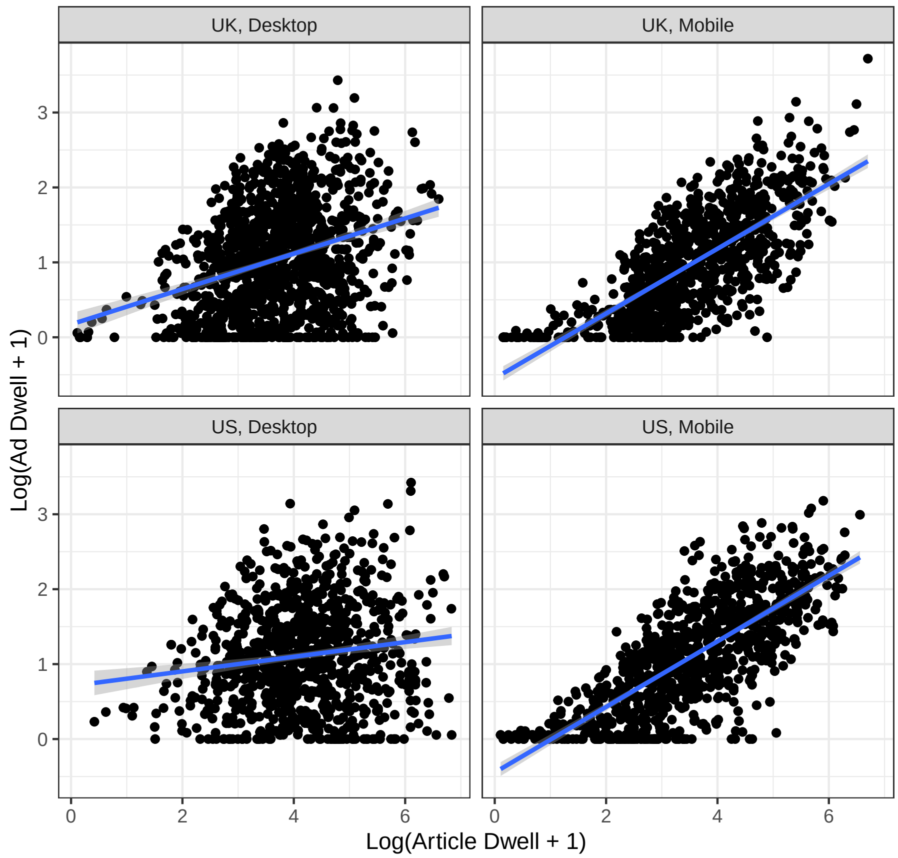
Figure 3 splits the attention paid to ads by their types – the "billboard" ad on top of the screen, the higher "side" located closer to the top of the page, and the lower " side" ad located further down the page. As discussed above, for mobile phones, "side" ads are shown in the center of the screen between paragraphs of text and therefore capture more of the consumer's attention than side ads on desktops. Consumers pay slightly less attention to the lower side ad on both types of devices. In contrast, billboard ads on desktop computers capture much more attention, because the wide format of desktop computers allows for a longer presence and share of the screen occupied by billboard ads.

Figure 3: Attention by Types of Ads



Attention to billboard, higher side, and lower side ads, by device type. Billboard ads receive more attention on desktop devices, but less attention on mobile phones. Bars correspond to 95% confidence intervals.

Figure 4: Positive Correlation in Article and Ad Dwell



Correlation between attention to article and attention to ad, for each article shown to UK and US individuals. Each panel corresponds to a set of articles in this country on this device type. Ad and article dwell times are transformed into the logarithmic scale to make the visualization easier to read. The blue line corresponds to the best linear prediction of the variable on the vertical axis by the variable on the horizontal axis.

Attention devoted to articles and ads declines as the experiment progresses, both on mobile and desktop devices. Appendix Figure 17 considers *Article Dwell* and *Ad Dwell* for the 9 "steps" of the study (e.g., the third article shown corresponds to step 3, etc.). On average, *Article Dwell* is 117 seconds in the first step of the study on desktop computers (99 seconds on mobile), but only 64 seconds in the last step of the study (42 seconds on mobile). Similarly, *Ad Dwell* in the first step is around 4 seconds on both types of devices, but only about 2.2 seconds in the last step of the study.

The similarity of patterns of attention for articles and ads suggests a positive relationship between the two. Figure 4 plots the distribution of *Ad Dwell* and *Article Dwell* . Across both countries and device types, there is a strong positive correlation (0.36) in the amount of attention consumers pay to articles and ads on the same page. The correlation is stronger for mobile devices (0.65) than for desktops (0.16). Appendix Figure 18 further breaks down this relationship by article and shows that this positive correlation holds also within each article. The positive correlation between *Article Dwell* and *Ad Dwell* holds even after we control for country, device, step-order, and demographics fixed effects (henceforth "FE").[23] In Section 8 we show that this positive correlation holds in a battery of robustness checks where we account for potential measurement error in attention, suggesting that attention to articles generates positive attention spillovers towards ads on the page.

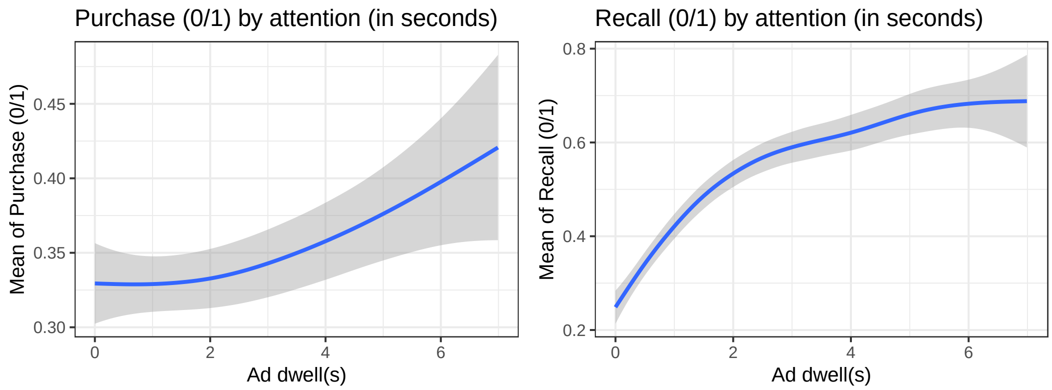
Purchase and Recall Measures

Our main outcome metrics are the share of consumers who recall the advertised brand and who make an incentivized "purchase" (i.e., choose a \$10/£10 voucher for the brand over a lower amount in cash). We validate that purchase choices were meaningful for consumers (i.e., that they were incentivized and paid attention to the presented options) by estimating the demand curves implied by the randomly assigned amounts of cash, or "prices". Appendix Figure 20 plots the average share of consumers in the US and UK who have chosen the brand voucher over cash. On average, 52% of consumers in the US and 40% of consumers in the UK chose the brand voucher over \$3/£3. This share gradually decreases as the cash amount increases, with only 34% of consumers in the US and 20% of consumers in the UK choosing the brand voucher over \$7/£7. In Appendix Figure 21, we estimate demand curves separately for each brand and confirm that this relationship is not driven by outlier brands.

In Figure 5, we examine how recall and purchase correlate with *Ad Dwell* . The figure shows that the percentage of individuals who chose the voucher and recalled seeing the brand increases with the amount of attention devoted to the ad.

[23] Appendix Figure 19 presents a version of Figure 4 that uses only residual variation in attention measures.

Figure 5: Purchase and Recall Increase in Ad Dwell Time



The panels show non-parametric regressions of purchase / recall on *Ad Dwell*, together with 95% confidence intervals. The automatic optimal bandwidth is used. The range of the x-axis is capped at 7 seconds, which is approximately the 90th percentile of the distribution.

5 A Simple Model of Attention Allocation, Recall, and Purchase

To ground our analysis, we begin by presenting a stylized model of how individuals allocate their attention to articles and ads. Then, we discuss how this attention impacts recall and purchase.

5.1 Attention Allocation

Consider reader i deciding how much attention to pay to article j (x_{art}) and display ads of brand k shown next to this article (x_{ad}). The reader chooses x_{art} and x_{ad} to maximize utility

$$U_{ijk}(x_{\text{art}}, x_{\text{ad}}) = \alpha_{ij} x_{\text{art}} - \frac{x_{\text{art}}^2}{2} + \mathbb{1}_{ijk}(-\beta x_{\text{art}} + \delta_{ik} x_{\text{ad}} + \gamma x_{\text{art}} x_{\text{ad}} - \frac{x_{\text{ad}}^2}{2}). \tag{1}$$

Here, α_{ij} captures reader i's interest in article j. The indicator $\mathbb{1}_{ijk}$ describes whether the ad of brand k was shown next to article j for participant i. The coefficient β is the reader's disutility from paying attention to the article when any ad is shown next to it (or utility if $-\beta > 0$). The coefficient δ_{ik} is the reader's preference for devoting attention to the ad of brand k. Finally, γ is the parameter that determines whether the reader prefers to spend more attention on the ad if they spend more attention on the article, and vice versa (i.e., it measures the complementarity or substitutability between article and ad). We include negative quadratic terms $x_{\text{art}}^2/2$ and $x_{\text{ad}}^2/2$

to ensure an interior solution while keeping the setting simple.[24]

Maximizing utility with respect to $x_{\text{art}}, x_{\text{ad}}$, and denoting the solutions $x^*_{\text{art},ijk}, x^*_{\text{ad},ijk}$, yields the First Order Conditions

$$x^*_{\text{art},ijk} = \alpha_{ij} + \mathbb{1}_{ijk}(-\beta + \gamma x^*_{\text{ad},ijk}) \tag{2}$$

$$x^*_{\text{ad},ijk} = \mathbb{1}_{ijk}(\delta_{ik} + \gamma x^*_{\text{art},ijk}). \tag{3}$$

There are two coefficients of particular interest: β and γ. The sign of β reflects whether the reader is an "ad avoider" or "ad lover". It is possible that individuals can be "ad lovers" (e.g., this might be particularly likely in the context of car or beauty magazines, [Kaiser and Wright, 2006]). However, past literature has tended to find that consumers are more likely to be ad avoiders [e.g. Wilbur, 2008, 2016, Huang et al., 2018], so we expect $\beta > 0$.

The coefficient γ determines whether articles and ads are substitutes or complements. A priori, both could happen: a more interesting article could grab the reader's attention more effectively due to voluntary ("top-down") attention and increase ad avoidance [Drèze and Hussherr, 2003, Stenfors et al., 2003, Simola et al., 2011], but more time spent on the page reading the article also provides more opportunities for the ad to distract the reader with its visual stimuli, working through a model of "bottom-up" attention [Koch and Ullman, 1987, Itti et al., 1998, Pieters and Wedel, 2007, Cerf et al., 2007, Milosavljevic and Cerf, 2008].

5.2 Recall and Purchase

We now consider how attention to ads determines recall and purchase. We first specify a recall model. We assume the consumer's recall is determined by

$$r_{ijk} = f_{ik}(\cdot) + \rho x^*_{\text{ad},ijk} + \epsilon^r_{ijk}, \tag{4}$$

where $r_{ijk} = \{0, 1\}$ is an indicator for whether individual i recalls brand k. The function $f_{ik}(\cdot)$ captures the effect of various characteristics of the consumer and brand on recall, and is described in more detail below. The parameter ρ is recall ad effectiveness: the effect of additional attention to the ad of brand k shown next to article j ($x^*_{\text{ad},ijk}$) on recall. Finally, ϵ^r_{ijk} captures other

[24] The model can be easily extended to allow for an overall time constraint such that $x_{\text{art}} + x_{\text{ad}} \leq \bar{x}$, and for differential costs of attention for ads and articles, but without additional insight.

idiosyncratic shocks and match values that determine consumer i's recall of brand k, when its ad is shown next to article j.

Similarly, we assume that consumer i's utility from purchasing the voucher for brand k at a price p_{ik} after devoting attention $x^*_{\text{ad},ijk}$ to the ad for brand k is

$$v_{ijk} = g_{ik}(\cdot, p_{ik}) + \lambda x^*_{\text{ad},ijk} + \epsilon^v_{ijk}, \tag{5}$$

where $v_{ijk} = \{0, 1\}$ is an indicator describing whether individual i chooses the voucher for brand k (as opposed to choosing the cash p_{ik}). The function $g_{ik}(\cdot, p_{ik})$ captures the baseline (indirect) utility the consumer i gets from brand k and the effect of price p_{ik}, and is described in more detail below. The parameter λ is the purchase ad effectiveness: the effect of additional attention to the ad of brand k on the decision to purchase that brand. Finally, ϵ^v_{ijk} captures other idiosyncratic utility shocks.

6 Results: Determinants of Attention Allocation

We start our empirical analysis by estimating the attention allocation parameters. Consumers decide how much attention to allocate to articles and ads on the page according to Equations (2) and (3). The ad preference and spillover parameters (β, γ) determine the interdependence of attention to articles and ads.

To estimate these parameters, we proceed sequentially. First, we use Equation (3) to estimate γ and δ_{ik} by an OLS regression of $x^*_{\text{ad},ijk}$ on $x^*_{\text{art},ijk}$, FE, and controls. We use only observations when the ad is present on the page, since otherwise $x^*_{\text{ad},ijk}$ is zero.

There is a potential bias associated with this estimation strategy, since $x^*_{\text{ad},ijk}$ is jointly determined by the two FOCs (Equations 2 and 3). However, the bias associated with this estimation is likely to be small, since an average consumer allocates around 30 times more attention to the article (75 seconds) over the ad (2.7 seconds). We also show below that an IV specification yields essentially the same estimates.

With these estimates in hand, we next use Equation (2) to estimate α_{ij} and β via an OLS regression of $x^*_{\text{art},ijk} - \mathbb{1}_{ijk}\hat{\gamma}x^*_{\text{ad},ijk}$ on $\mathbb{1}_{ijk}$, FE, and controls, where $\hat{\gamma}$ is the estimate of γ from the first step. In this case, we use the entire sample, including the articles shown without ads. Our identification argument relies on the random assignment of ads to articles, including a random chance that an article is shown with no ads on the page.

Table 2: Estimates of attention spillovers and ad avoidence

	Ad Dwell					
	OLS			IV		
Panel I	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\delta}_1$	2.715***			3.083***		
	(0.197)			(0.306)		
$\hat{\gamma}$	0.011***	0.012***	0.012***	0.008***	0.009***	0.009***
	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)
1st Stage Incr. F-Stat				65.86	119.78	124.93
Observations	3,925	3,925	3,925	3,925	3,925	3,925
R^2	0.145	0.156	0.549	0.135	0.150	0.547
	Article Dwell - $\hat{\gamma}$ Ad Dwell					
Panel II	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\alpha}_1$	105.894***			105.907***		
	(4.521)			(4.521)		
$\hat{\beta}$	7.024*	6.845*	8.798**	7.015*	6.837*	8.791**
	(3.919)	(3.741)	(3.509)	(3.919)	(3.741)	(3.509)
Observations	4,426	4,426	4,426	4,426	4,426	4,426
R^2	0.030	0.112	0.640	0.030	0.112	0.640
FE:						
Step Order	Y	Y	Y	Y	Y	Y
Article	N	Y	Y	N	Y	Y
Brand	N	Y	Y	N	Y	Y
Individual	N	N	Y	N	N	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include step order fixed effects, with step order = 1 normalized to zero. Estimates in Panel I represent coefficients from a regression of Ad Dwell on Article Dwell. In the IV specification, Article Dwell is instrumented for by the average amount of attention devoted to that article by all but this individual (Leave One Out IV). Estimates in Panel II represent coefficients from a regression of Article Dwell on an indicator of whether the ad is present on the news article. We subtract $\hat{\gamma}$ Ad Dwell from Article Dwell in Panel II to control for the attention spillover from ad to article. Standard errors clustered at the individual level.

Table 2 presents the estimates under alternative assumptions about α_{ij} and δ_{ik}. In Column (1), we include only step-order FE (i.e., we assume $\alpha_{ij} = \alpha_{\text{step}_{ij}} + \epsilon^{\alpha}_{ij}$ and $\delta_{ik} = \delta_{\text{step}_{ij}} + \epsilon^{\delta}_{ik}$), where we denote step_{ij} as the step order at which article j was shown to individual i. In the first article read (i.e., when $\text{step}_{ij} = 1$), an average consumer decides to allocate $\hat{\alpha}_1 = 106$ seconds to the article and $\hat{\delta}_1 = 2.7$ seconds to the ad. An extra second spent looking at the article increases the amount of time a consumer looks at the ad by $\hat{\gamma} = 0.011$ seconds. Thus, an average 106 seconds spent looking at the first article creates $106 \cdot 0.011 = 1.166$ seconds of positive spillover attention to the ad. Likewise, the baseline of 2.7 seconds spent looking at the ad creates $2.7 \cdot 0.011 = 0.03$ seconds extra attention to the article – a positive but arguably economically negligible amount of attention. Having no ad next to the article increases the amount of attention readers spend looking at the article by $\widehat{\beta} = 7$ seconds – a much more meaningful effect, showing that the average consumer is indeed an ad avoider.[25]

In Column (2) of Table 2, we allow these terms to vary across articles and ads, i.e., $\alpha_{ij} = \alpha_j + \alpha_{\text{step}_{ij}} + \epsilon^{\alpha}_{ij}$ and $\delta_{ik} = \delta_k + \delta_{\text{step}_{ij}} + \epsilon^{\delta}_{ik}$, by including article and ad FE. While the baseline levels of $\hat{\alpha}$ and $\hat{\delta}$ are now subsumed by the FE, the estimates of the ad avoidance and attention spillover parameters $(\hat{\beta}, \hat{\gamma})$, are nearly identical to the estimates in Column (1).

In Column (3), we further include individual FE, i.e., we assume $\alpha_{ij} = \alpha_i + \alpha_j + \alpha_{\text{step}_{ij}} + \epsilon^{\alpha}_{ij}$ and $\delta_{ik} = \delta_i + \delta_k + \delta_{\text{step}_{ij}} + \epsilon^{\delta}_{ik}$, thus allowing different readers to have different overall preferences for articles and ads. Again, the estimates of $\hat{\beta}$ and $\hat{\gamma}$ are statistically indistinguishable from those in Column (1). This similarity in the estimates is expected since the key source of identification is the random assignment of ads to articles. Still, the results validate that the randomization was done correctly.

In Columns (4-6) of Table 2, we test the assumption that the spillover effect of attention from ads to articles is negligible. The concern is that there is a feedback loop from $x^*_{\text{ad},ijk}$ to $x^*_{\text{art},ijk}$ as determined by Equations (2) and (3), creating a simultaneity problem. We instrument $x^*_{\text{art},ijk}$ with the average amount of attention devoted to that article by all *other* individuals in the sample. Recall that articles are randomly coupled with ads, and thus other individuals are equally likely to see any ads on a given article. We refer to this instrument as the "Leave One Out" (L1O) mean of article attention.[26]

[25] We omit the step-order FE estimates from Table 2 to improve readability. Articles and ads shown in later steps of the experiment obtained less attention from participants, as illustrated in Appendix Figure 17.

[26] This "jackknife" instrument is similar to the use of article FE as an instrument, but eliminates the bias associated with including the current respondent when computing the FE, as discussed by Angrist et al. [1999] and Kolesar [2013]. This IV approach is identical to the "random judges" instruments used in Dahl et al. [2014] and Dobbie

Column (4) presents the estimates using the L1O instrument and assumes that the average preference for allocating attention to articles and ads is affected only by step-order FE. The first stage relationship is highly significant with an incremental F-statistic of 65.9. Again, the estimates are statistically indistinguishable from the estimates in Column (1). For the first article shown (i.e., when $\text{step}_{ij} = 1$), an average consumer decides to allocate $\hat{\delta}_1 = 3.1$ (versus 2.7 in Column 1) seconds to the ad, and $\hat{\alpha}_1 = 106$ (same as Column 1) seconds to the article. An extra second spent looking at the article increases the amount of time a consumer looks at the ad by $\hat{\gamma} = 0.008$ seconds, statistically similar to the 0.011 seconds estimate in Column (1) given the standard error of 0.003. Similarly, the ad avoidance parameter $\hat{\beta}$ is almost identical to Column (1), implying that removing ads increases the amount of attention consumers pay to an article by 7 seconds.

Columns (5-6) confirm the similarity in the attention spillover and ad avoidance estimates under different assumptions about α_{ij} and δ_{ik}. Most importantly, we obtain very similar estimates of the attention spillover parameter ($\hat{\gamma}$) – around 0.008-0.011 – using OLS or IV. This suggests that endogeneity between *Article Dwell* and *Ad Dwell* is not a big concern in this setting. In turn, this validates that *Article Dwell* serve as a valid instrument for *Ad Dwell* , the cornerstone of our empirical approach when estimating the effect of ad attention on recall and purchases, below.

Appendix Tables 20-21 replicate the attention allocation model estimates separately for mobile and desktop devices. For both device types, results are qualitatively consistent: attention to articles generates positive spillovers towards attention to ads, and consumers tend to pay more attention to news content in the absence of ads. However, there are some interesting differences. First, the attention spillover is stronger on mobile devices than on desktops, although the difference shows up mainly in the OLS specifications. This is presumably because mobile ads – and especially "side" or "scroll" ads – are located much closer to the main text, making it more likely that consumers pay attention to ads while reading the article. Second, ad avoidance seems to be stronger on desktops than on mobile devices – throughout all specifications. Removing ads increases the amount of attention to the article by 8.5-11.5 seconds on desktops and only by 3.3-

et al. [2018]. In this case, each article is a "judge", and ads are randomly assigned to articles. Articles of different attractiveness serve as judges who vary in leniency.

5.5 seconds on mobile devices. However, standard errors on these estimates are relatively large (5.5-6.2 seconds) making the difference not statistically significant.

7 Results: Determinants of Recall and Purchase

We now estimate the effect of ad attention on consumers' recall and purchase decisions. Our experiment mimics the setting described in the stylized model above. We have randomly assigned ads to articles, have shown these articles to consumers in a random order, have measured attention to articles and ads, and then asked consumers if they recall the brands and to make a purchase choice (between a voucher for a product and cash) of each brand they have seen.

7.1 OLS

Our baseline analysis uses OLS. We observe $x^*_{\text{ad},ijk}, r_{ijk}$, and v_{ijk} in the data. Given the experimental variation, we proceed by estimating Equations (4) and (5) directly. We parameterize $f_{ik}(\cdot)$ and $g_{ik}(\cdot, p_{ik})$ and estimate the following linear probability models [Heckman and Snyder Jr, 1997]:

$$r_{ijk} = \theta^r_{o_{ik}} + \eta^r_k + \mu^r X_i + \rho x^*_{\text{ad},ijk} + \epsilon^r_{ijk} \tag{6}$$

$$v_{ijk} = \theta^v_{o_{ik}} + \eta^v_{k,p_{ik}} + \mu^v X_i + \lambda x^*_{\text{ad},ijk} + \epsilon^v_{ijk}. \tag{7}$$

The coefficients $\theta_{o_{ik}}$ are the FEs for the "step-order" variable, $o_{ik} = \{1, 2, ..., 9\}$, in which ad k was shown to individual i. These parameters capture the effects of seeing an ad later or earlier in the experiment. The effect is a priori ambiguous. Ads shown later may receive less attention due to fatigue; however, an ad shown at the end of the experiment might be more vivid in the participant's memory when they are asked to make a decision.

The coefficients η^r_k and $\eta^v_{k,p_{ik}}$ are, respectively, brand and brand×price FE. Brand FE, η^r_k, control for the fact that some brands might be more popular than others. We assume that price does not affect consumer recall but potentially affects purchase decisions. By including brand×price FE ($\eta^v_{k,p_{ik}}$) we allow the price elasticity to vary flexibly along the demand curve for each product, and allow these demand curves to differ across products.

We use X_i to denote a vector of controls at the individual level. These include country×device

FE (e.g., participants on desktops in the UK) and various socio-demographic characteristics (also included as dummy variables).[27] We also include in X_i a fourth-order degree polynomial of the average time spent on each article by each individual. We include this as an individual characteristic, to account for the fact that some individuals read more slowly or are intrinsically more engaged by articles than other individuals.

Table 3 shows the OLS estimates of the effect of ad attention – measured both with *Ad Visible* and *Ad Dwell* – on recall and purchase. Column (1) reports the estimates of recall ad effectiveness ($\hat{\rho}$) based on all observations in the sample. Panel I considers attention measured by *Ad Visible* . If a brand's ad is visible for 1 extra second, this increases the probability of the individual remembering that brand by about 0.32 percentage points. An increase in *Ad Visible* of one standard deviation (17.37 seconds, from Table 1) is associated with an increase in recall of $0.32 \cdot 17.37 = 5.55$ percentage points.

Panel II considers attention measured by *Ad Dwell* . If an ad is viewed for 1 extra second, this increases the probability of recall by 3.43 percentage points. An increase in one standard deviation of *Ad Dwell* (3.16 seconds) increases recall probability by 10.83 percentage points. In line with intuition, the magnitudes of the estimates are larger when attention is measured using the time individuals actually spend engaging with the ad (which we measure using eye-tracking). *Ad Dwell* explains an additional 5.2% of the variation in recall than *Ad Visible* – R^2 is 0.143 in Column (2) and 0.091 in Column (1). This highlights the value of eye-tracking as a more direct measure of attention.

Column (6) of Table 3 reports our estimates of purchase ad effectiveness, the effect of ad attention on the incentivized purchase behavior ($\hat{\lambda}$). Panel I shows that, if an article is visible for an extra second, this increases the probability of purchase by 0.13 percentage points. An increase of one standard deviation of *Ad Visible* increases the probability of purchase by 2.26 percentage points. If an ad is actually looked at for an extra second, the probability of purchase increases by 0.73 percentage points. An increase of one standard deviation in *Ad Dwell* leads to a 2.31 percentage points higher purchase probability. As in the case of recall, *Ad Dwell* has a better predictive power of the outcome measure (purchases) than *Ad Visible* – R^2 increases by

[27] For instance, individuals reported their age in bins of 10 years, so we include an indicator for each such bin. This specification non-parametrically controls for the level of each covariate.

Table 3: Estimates of advertising effects on recall and purchase: OLS

	Recall ($\hat{\rho}$)					Purchase ($\hat{\lambda}$)				
	All	Device		News Type		All	Device		News Type	
		Mobile	Desktop	Hard	Soft		Mobile	Desktop	Hard	Soft
Panel I	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ad Visible	0.003***	0.003***	0.004***	0.004***	0.002***	0.001**	0.002**	0.001*	0.001**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	5,707	2,495	3,212	3,154	2,553	5,707	2,495	3,212	3,154	2,553
R^2	0.091	0.103	0.120	0.102	0.096	0.130	0.164	0.147	0.147	0.147
Panel II	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ad Dwell	0.034***	0.028***	0.036***	0.041***	0.029***	0.007**	0.009**	0.008**	0.009**	0.005
	(0.004)	(0.006)	(0.005)	(0.004)	(0.005)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	3,925	1,824	2,101	2,165	1,760	3,925	1,824	2,101	2,165	1,760
R^2	0.143	0.133	0.188	0.167	0.139	0.136	0.200	0.153	0.168	0.159
FE:										
Brand	Y	Y	Y	Y	Y					
Price x Brand						Y	Y	Y	Y	Y
Step Order	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Country x Device	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Dem. Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include a quartic polynomial in log of average time that an average article was visible for by each individual. All specifications include step order and device x country fixed effects. All specifications include fixed effects for individual covariates (income, gender, education, age, and self-reported political leaning). Regressions with recall as an outcome include brand fixed effects, and regressions with purchase as an outcome include brand by price fixed effects. Standard errors clustered at the individual level.

0.6%. Again, this highlights the importance of our eye-tracking measure.

In Columns (2-5) and (7-10) of Table 3, we examine the effect of ad exposure on recall and purchase separately for different devices and types of news. Columns (2-3) and (7-8) report $\hat{\rho}, \hat{\lambda}$ separately for consumers participating in the experiment from mobile devices and desktop computers. Across all subsamples, estimates of the purchase and recall ad effectiveness are nearly identical, validating the importance of advertising both on mobile and desktop devices.

Second, we separately examine the effect of advertising on recall and purchase for "hard" and "soft" news. As discussed above, industry practitioners are wary of advertising next to "hard news" articles because of the perceived negative effect on their brand. This should imply that for "hard" news we should see smaller or even negative estimates of ρ, λ.[28] Columns (4-5) and (9-10) report the estimates of ρ, λ separately for ads that were randomly matched to "hard" and "soft" articles. Across all specifications, estimates of the recall and purchase ad effectiveness are qualitatively similar on both types of news. If anything, we find that the magnitudes of the estimates are slightly *higher* for ads shown next hard news. For instance, a one-second increase in *Ad Dwell* for ads next to hard news articles increases purchase probability by 0.9 percentage points, whereas a similar estimate for ads on soft articles is 0.5 percentage points.

In order to better understand the interaction between article content, ad effectiveness, and the total amount of ad attention, we regress our four attention variables (*Ad Dwell* , *Ad Visible* , *Article Dwell*, *Article Visible*) on an indicator of whether the article is classified as hard news. To keep the estimates consistent, we include the same controls as in Table 3. Further, to keep articles comparable, we control for their length by including the number of words as a control. This addresses the concern that "hard news" articles might systematically be longer or shorter than other articles, which would mechanically affect attention. Results are shown in Table 4.

Hard news articles, and the ads randomly shown next to these articles, receive less attention than other ads and articles. Individuals spend less time looking at the ad (Columns 1 and 2), and also less time looking at the article itself (Columns 3 and 4). In terms of *Article Dwell* , there is a reduction of almost 7 seconds (about 10% of the median), and a reduction of 0.49 seconds for *Ad Dwell* (about 15% of the median). However, these results should be interpreted

[28] Recall that, for hard news, and based on discussions with industry experts, we included articles about the COVID-19 pandemic and BLM protests, since these were often blocked by advertising intermediaries. The experiment took place in late July 2020.

Table 4: Attention and Hard News

	Measure of attention:			
	Ad Visible	Ad Dwell	Article Visible	Article Dwell
	(1)	(2)	(3)	(4)
Hard News	-0.9049^{***}	-0.4902^{***}	-10.1627^{***}	-6.7510^{***}
	(0.3168)	(0.0828)	(2.6201)	(2.0077)
Observations	5,707	3,925	5,707	3,925
R^2	0.4187	0.1461	0.6355	0.4705

Note: *p<0.1; **p<0.05; ***p<0.01. Fixed Effects: Individual covariates (income, gender, education, age, politics), Step Order, Brand, Country by Device. Includes a quartic polynomial in total time an average page is visible for each individual. Includes a linear control for number of words in article. Standard errors clustered at the individual level.

with caution. It is possible that, since there were many hard news stories at the time of the experiment, individuals could already be informed about those topics (the experiment did not allow testing for pre-experiment knowledge), or possibly individuals had wearied of such stories. We cannot say whether our finding is due to participants disliking hard news articles or because they were shown articles on topics they were already aware of, and so they skimmed through them quickly.

Even if we interpret the overall lower attention that consumers pay to hard news as causal, this need not imply that advertising next to hard news ultimately brings less attention to the brand than advertising next to soft news. For this, we combine the effects of hard news on the amount of attention readers pay to ads (Table 4) and the effect of this attention on purchases (Columns (9-10) in Part II of Table 3). The average attention to ads is 2.76 seconds. Paying this amount of attention to ads next to "soft news" articles increases the purchase probability by $0.5 \cdot 2.76 = 1.38$ percentage points. For hard articles, the same effect is $0.9 \cdot (2.76 - 0.49) = 2.04$ percentage points, since hard news on average induce 0.49 seconds less attention to ads but have higher (0.9 instead of 0.5) ad effectiveness. On balance, this implies that the benefits of advertising next to hard news articles are similar, if not higher, compared to advertising next to soft news articles.

Previous work has found that attention fatigue and decay can take place [Goldstein et al., 2011, Ahn et al., 2018]. With this in mind, we allow outcomes to be function of a quadratic polynomial in attention in Equations (6) and (7). We find that indeed returns to attention are diminishing; for all specifications, the quadratic term on the advertising attention is negative

(Appendix Table 17). However, within our sample, the non-linear effects are small and, for our main outcome variable (incentivized purchase) statistically insignificant.[29]

Our estimates are similar (although less precise) in a much more demanding specification that includes individual FE (Appendix Table 18). Our results are also robust to using a logit specification instead of the linear probability model (Appendix Table 19).

7.2 IV

In Section 7.1, we estimated the effects of ad attention on recall and purchases (ρ, λ) by OLS, following Equations (6)-(7). Consistency of the estimates $\hat{\rho}, \hat{\lambda}$ requires that the amount of time consumers allocate to ads, $x^*_{\text{ad},ijk}$, is uncorrelated with the idiosyncratic shocks that influence consumers' recall and purchase outcomes, $\epsilon^r_{ijk}, \epsilon^v_{ijk}$. However, these assumptions are violated if $\epsilon^r_{ijk}, \epsilon^v_{ijk}$ are correlated with the idiosyncratic shock that enters the decision of consumer to pay attention to the ad, ϵ^δ_{ik}.

One concern is omitted variable bias. For instance, perhaps individuals who enjoy shopping tend to pay more attention to ads and are also more likely to purchase. Since we cannot observe individual enjoyment of shopping, this could generate correlations between ϵ^δ_{ik} and $\epsilon^r_{ijk}, \epsilon^v_{ijk}$. While this is a concern in a model without individual FE, it is addressed by the robustness check where FE are included (Appendix Table 18).

Another concern is reverse causality. For instance, if individual i enjoys hamburgers and dislikes coffee, she may devote significant attention to hamburger ads and purchase hamburgers (high $\epsilon^\delta_{ik'}$ and $\epsilon^r_{ijk'}$ for k' = hamburgers, even controlling for δ_i), and little attention to coffee ads and not purchase coffee (low $\epsilon^\delta_{ik''}$ and $\epsilon^r_{ijk''}$ for k'' = coffee). This intuition was formalized by Stigler and Becker [1977], Becker and Murphy [1993], who model advertising as having consumption value and entering viewers' utilities, and empirically tested by Tuchman et al. [2018].

We address these concerns by leveraging the results from our attention allocation model. Our goal is to rule out the effect of consumer-specific tastes for ads, ϵ^δ_{ik}, which both influence $x^*_{\text{ad},ijk}$

[29]The median value of *Ad Dwell* is approximately 3 seconds. At this level of attention, ignoring the non-linear term, attention raises the purchase probability by $3 \cdot 0.0104 = 0.0312$. The non-linear term lowers this figure by $-(3^2) \cdot 0.0002 = -0.0018$. Also, the quadratic specification implies that, for large values of attention, purchase probability would eventually decrease with attention. However, this never occurs in the sample. For instance, looking at the effect of *Ad Dwell* on purchase (column 4) of Table 17, purchase probability would be declining in attention only for attention greater than $0.0108/(0.0002 \cdot 2) \approx 54$ seconds, whereas the median value of *Ad Dwell* is less than 3 seconds.

and are potentially correlated with preference shocks ϵ^r_{ijk} and ϵ^v_{ijk}. To do so, we instrument the amount of attention a reader pays to an ad (x^*_{ad}) with the amount of attention she has devoted to the article randomly paired with that ad (x^*_{art}). Our estimates in Table 2 have shown that there is a strong positive spillover in the consumer's attention from article to ads, making x^*_{art} a relevant instrument. We have also shown that the reverse effect (of x^*_{ad} on x^*_{art}) is minuscule and that estimates of the spillover effect are robust to instrumenting for x^*_{art} using a L1O strategy, removing concerns of the reverse effect and validating the exogeneity of x^*_{art} as an instrument for x^*_{ad}. Intuitively, we exploit the fact that ads are randomly paired with more or less interesting articles, and use only the incremental exposure to ads due to positive spillovers of attention to measure its effect on recall and purchase.

Table 5 presents the estimates of ρ and λ from the IV regressions of Equations (6)-(7), where we instrument for *Ad Visible* and *Ad Dwell* with *Article Dwell* . Since *Article Dwell* is measured using eye-tracking, this measure of attention to articles is not contaminated by attention to ads. Panel I presents the results with *Ad Visible* as the measure of attention to ads. The first stage results are presented in the bottom part of Panel I. For all specifications, we have strong instruments – incremental F-statistics vary from 31.9 to 79. The first stage regressions confirm strong positive attention spillovers between the article and ads, described in Table 2. The second stage IV estimates are presented at the top part of Panel I. For the outcome of recall (Columns 1-5), the estimates $\hat{\rho}$ are too imprecise to conclude that they are different from the OLS results. When we include all observations (Column 1), $\hat{\rho} = 0.0003$ is smaller than the OLS estimate of 0.003 – but the standard error of the estimate is 0.002, making the difference statistically insignificant.

For the purchase outcome (Columns 6-10), the estimates $\hat{\lambda}$ are positive and statistically significant – the spillover attention to ads due to an interesting article leads to a higher purchase probability of the product that was advertised. If anything, the estimated magnitudes of $\hat{\lambda}$ are larger for the IV case – although differences between the IV and OLS estimates are only marginally significant, due to larger standard errors of the IV estimates. For instance, results in Column (6) show that increasing *Ad Visible* by one second increases the purchase probability by 0.6 percentage points (a standard error of 0.2), compared to the OLS estimate of 0.1 percentage points.

Interestingly, the IV estimate of $\hat{\lambda}$ is larger than the OLS estimate. This suggests that reverse causality is not a big concern in this case, since consumer behavior *à la* Stigler and Becker [1977],

Table 5: Estimates of advertising effects on recall and purchase: Article Dwell IV

	Recall ($\hat{\rho}$)					Purchase ($\hat{\lambda}$)				
	All	Device		News Type		All	Device		News Type	
		Mobile	Desktop	Hard	Soft		Mobile	Desktop	Hard	Soft
Panel I	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ad Visible	0.0003	0.008	-0.001	0.001	-0.0004	0.006***	0.013**	0.004*	0.007**	0.005*
	(0.002)	(0.007)	(0.002)	(0.003)	(0.003)	(0.002)	(0.006)	(0.002)	(0.003)	(0.002)
Observations	3,925	1,824	2,101	2,165	1,760	3,925	1,824	2,101	2,165	1,760
R^2	0.105	0.097	0.138	0.122	0.103	0.123	0.128	0.146	0.148	0.149
						First Stage				
Article Dwell	0.058***	0.024***	0.076***	0.057***	0.059***	0.058***	0.026***	0.075***	0.057***	0.059***
	(0.007)	(0.007)	(0.009)	(0.010)	(0.007)	(0.007)	(0.007)	(0.008)	(0.010)	(0.007)
Observations	3,925	1,824	2,101	2,165	1,760	3,925	1,824	2,101	2,165	1,760
R^2	0.459	0.282	0.566	0.430	0.516	0.467	0.307	0.582	0.447	0.531
1st Stage Incr. F-Stat	75.85	12.83	76.71	31.92	71.4	77.74	14.53	79.01	33.5	70.79
Panel II	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ad Dwell	0.001	0.008	-0.009	0.005	-0.002	0.028**	0.015**	0.052	0.036**	0.025*
	(0.010)	(0.007)	(0.032)	(0.014)	(0.013)	(0.011)	(0.007)	(0.037)	(0.016)	(0.013)
Observations	3,925	1,824	2,101	2,165	1,760	3,925	1,824	2,101	2,165	1,760
R^2	0.107	0.122	0.117	0.130	0.100	0.119	0.199	0.079	0.145	0.143
						First Stage				
Article Dwell	0.011***	0.022***	0.005***	0.011***	0.011***	0.011***	0.023***	0.005***	0.011***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	3,925	1,824	2,101	2,165	1,760	3,925	1,824	2,101	2,165	1,760
R^2	0.205	0.476	0.129	0.208	0.232	0.220	0.500	0.156	0.235	0.262
1st Stage Incr. F-Stat	48.23	173.37	11.96	34.63	34.09	48.52	183.4	10.41	36.78	34.52
FE:										
Brand	Y	Y	Y	Y	Y					
Price x Brand						Y	Y	Y	Y	Y
Step Order	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Country x Device	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Dem. Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include a quartic polynomial in log of average time that an average article was visible for by each individual. All specifications include step order and device x country fixed effects. All specifications include fixed effects for individual covariates (income, gender, education, age, and self-reported political leaning). Regressions with recall as an outcome include brand fixed effects, and regressions with purchase as an outcome include brand by price fixed effects. Standard errors clustered at the individual level.

Becker and Murphy [1993] would upward bias the OLS estimates.

Panel II of Table 5 presents the results with *Ad Dwell* as the measure of attention to ads. All conclusions are the same as in the case of *Ad Visible* . The first stage results confirm a strong complementarity between attention devoted to articles and ads. Incremental F-statistics are between 10 and 183 across specifications, with the strongest relationship for mobile and the weakest for desktop devices. The estimates $\hat{\rho}$ are imprecise across the specifications (Columns 1-5), while the estimates of $\hat{\lambda}$ are positive and statistically significant (Columns 6-10). The IV estimates of $\hat{\lambda}$ are larger than the OLS estimates, but the difference is not statistically significant.[30]

Columns (7-10) of Table 5 report the estimates of λ separately for different devices and types of news. The estimates confirm the conclusions from the OLS analysis. For mobile and desktop devices, the estimates of λ are not statistically different when using *Ad Visible* or *Ad Dwell* as the attention metrics. Similarly, the effect of advertising on recall and purchase for "hard" and "soft" news are qualitatively similar. Again, the magnitudes of the estimates are slightly *higher* for ads shown next to articles with hard news.

As an additional robustness check, we use the L1O article attention (the average amount of attention devoted to the article by all *other* individuals in the sample, as defined in Section 6) to instrument for the amount of attention to ads. From Table 2, the L1O attention significantly shifts the amount of time a consumer pays to the article, which in turn has a positive spillover effect on the attention to ads on this page. We present these results in Appendix Table 22. Once again, the first stage estimates confirm the positive spillover of attention from articles to ads, although the strength of the instrument is weaker (e.g. for *Ad Dwell* as a measure of ad attention, F-stats vary from 1.8 to 13.3). Because of lower statistical power of this instrument, the second stage (IV) coefficients are also estimated imprecisely although, reassuringly, they have the same magnitude as previous OLS and IV results.

Finally, we check the robustness of our estimates using an alternative shifter of consumers' attention to articles: the (mis)alignment between consumers' and newspapers' political views. Past work has shown that news readers prefer articles with an ideological slant aligned with their

[30]For instance, pooling the data across all devices and types of news (Column 6) we find that increasing *Ad Dwell* by one second increases the purchase probability by 2.8 percentage points (a standard error of 1.1). This estimate is larger in magnitude but statistically similar (the difference significant only at 10% level) to the OLS estimate of 0.7 percentage points (with a standard error of 0.3).

own views [Schmuck et al., 2019]. The outlets we chose in each country are widely perceived as differing in their political leaning. For instance, in the UK, *The Guardian* is perceived as left-leaning and *The Daily Mail* as right-leaning. We construct a measure of political alignment of consumers and news outlets by asking experimental participants about their political views as well as classifying news outlets as left, center, or right.[31] Appendix I provides more details on the classification of news outlets and the construction of a political "mismatch" variable. In short, a large mismatch occurs when a person is presented with an article from an outlet at the opposite end of the political spectrum, while there is no mismatch when there is full alignment. We show that our political mismatch variable strongly predicts consumers' attention to articles – going from fully aligned views to completely misaligned views decreases the time people read the article by around 18 seconds. This, in turn, decreases the attention people pay to ads on the page, with ads becoming visible for 1.64 seconds (a standard error of 0.57 seconds) less and attracting 0.24 seconds (a standard error of 0.16) less active attention (dwell) time. The magnitudes align with the previous results on attention spillovers well – e.g., the first stage results in Table 5 (Column 1) implies that extra 18 seconds of article dwell increase *Ad Visible* and *Ad Dwell* by 1.1 and 0.2 seconds, respectively. However, the effect of political mismatch on the attention to ads is too imprecise to produce conclusive estimates of ρ and λ. We report the estimates and provide their discussion in Appendix Tables 23-25.

8 Robustness of Eye-Tracking Measurements

In this section we address the issue of potential measurement error in eye-tracking. Throughout the study, consumers perform multiple validation procedures (initially, and after the third and sixth articles) where the quality of the eye-tracking data is determined. At each validation, individuals are asked to look at a moving point on the screen, and their eye gaze is measured (see Appendix A.3 for details).

We consider primarily three measures of the quality of eye-tracking data. These measures are used extensively by the provider of the eye-tracking software (Lumen Research) and by the lit-

[31] This was done at the end of the experiment so that this question would not bias the behavior of participants. We validate the news outlets' classifications in an independent survey on AMT (see Appendix B for details).

erature [e.g. Semmelmann and Weigelt, 2018, Schneegans et al., 2021, Yang and Krajbich, 2021]. First, "accuracy": the average Euclidean distance between the dot that individuals are instructed to look at and the recorded gaze points. Second, "precision": the re-scaled standard deviation of gaze points around the dot that individuals are instructed to look at. Third, "gaze duration": a measure of how frequently the camera is able to record how eyes are moving. Notice that these are *inverse* measures of eye-tracking quality (higher values imply lower quality).

For our main analysis, we determine the quality of eye-tracking data using the three measures described above. If, for a given respondent$\times$article, the eye-tracking data is deemed to be "invalid", the eye-tracking data for that observation is dropped as we discuss in Section 4 and Appendix C.2. For our main analysis, we rely on Lumen Research's proprietary algorithm to flag invalid observations. However, we also validate that the participants selected by this procedure indeed have a lower quality of eye-tracking measures in Appendix A.3 and show robustness checks with respect to all these measures below.

For the individuals deemed to have valid eye-tracking data, we confirm that the data is indeed of high quality. These individuals have an average accuracy of 201 and 137 CSS pixels (on desktop and mobile, respectively) and an average precision of 115 and 65 CSS pixels (on desktop and mobile devices, respectively).[32] The accuracy and precision are low compared to the size of ads in our study: 970×250 CSS pixels for desktop billboard ad, 300×600 CSS pixels for side (desktop) and scroll (mobile) ads, and 320×80 CSS pixels for mobile billboard ad. The eye-tracking data is good enough to capture gazes of respondents in the interior of ads for all types of ads, except for billboard ads on mobile devices (which have a very low contribution to the ad dwell anyway, as illustrated by Figure 3 in Section 4.2).

To confirm that our results are not driven by imprecision in our measure of *Ad Dwell* , in Tables 6 and 7 below we repeat our analysis with observations re-weighted by the (inverse of the) three measures of eye-tracking data quality described above, and two other measures of quality that we describe below. We also repeat the analysis using two alternative *Ad Dwell* measures that reduce potential measurement error.

First, we confirm our OLS and IV results when observations are re-weighted by five different

[32]A CSS pixel is a metric used in web browsers that ensures that regardless of the device's physical pixel density, web objects always occupy the same proportion of the screen. See `https://www.w3.org/Style/Examples/007/units.en.html` for details.

metrics of eye-tracking data precision. Three of these metrics are the inverses of the quality metrics we introduced above (i.e., 1/precision, 1/accuracy, and 1/gaze duration). We also use two additional measures of eye-tracking quality. First, the "hit rate", the share of gaze points falling within 200 CSS pixels distance from the calibration point individuals are instructed to look at during the validation steps. Second, the inverse of the time since the beginning of the study, since the quality of eye-tracking metrics deteriorates as the study progresses (see Appendix Table 9). To each (individual$\times$article) observation, we assign the quality metrics obtained from the last validation step before the article is shown (e.g. an article shown in step 5 is assigned metrics from the second validation, which occurred after article 3).

Second, we use raw eye-tracking data to re-create the *Ad Dwell* measure in two ways that minimize potential measurement error. First, we consider only gazes that fall into the "interior" of the ad, i.e., we exclude borders of 25% of the ad's height and width. For instance, for a 300×600 CSS pixel ad, we count only gazes falling on the 150×300 pixel surface in the center. For the second adjusted *Ad Dwell* measure, we exclude gazes that fall on the 50% of the surface of the ad that is closest to the main text. For instance, for a 970×250 pixel billboard ad on the top of the screen, we consider only gazes falling on the top 970×125 pixel surface, since the article text is below the ad. We re-scale both of these measures to make the mean of the adjusted *Ad Dwell* similar to the mean of the original *Ad Dwell*, so that the interpretations of the coefficients' magnitudes are comparable.

Tables 6 and 7 replicate our main results with re-weighted samples and alternative measures of *Ad Dwell*. In Table 6 we replicate the main OLS estimates for *Ad Dwell* from Table 3, Columns (1) and (6) in Panel II. We report these estimates in Column (1) of Table 6 to make comparisons of estimates easier. We find that the magnitude of coefficients is the same across all seven alternative specifications. For the recall outcome (Panel I), the estimates vary from 0.017 to 0.036, but none are significantly different from the main estimate of 0.034 (considering the standard errors). For the purchase outcome (Panel II), the estimates vary from 0.005 to 0.008, but again none are significantly different from the main estimate of 0.007.

In Table 7 we replicate the main IV estimates for *Ad Dwell* for the purchase outcome from Table 5 (since we obtained statistically significant estimates only on the purchase outcome). Once again, we report the main estimates in Column (1) of Table 7 to make comparisons easier.

Table 6: Estimates of advertising effects on recall and purchase: OLS, Robustness in Ad Dwell Measurements

	Recall ($\hat{\rho}$)							
	Main	Re-weighted observations by					Adjusted Ad Dwell	
		1/precision	1/accuracy	1/gaze duration	Hit Rate	1/time since calibr.	Interior	Away from Text
Panel I	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ad Dwell	0.034***	0.017*	0.036***	0.033***	0.034***	0.029***	0.026***	0.026***
	(0.004)	(0.010)	(0.004)	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)
Observations	3,925	3,875	3,875	3,925	3,925	3,925	3,925	3,925
R^2	0.143	0.176	0.153	0.147	0.146	0.161	0.142	0.141
	Purchase ($\hat{\lambda}$)							
Panel II	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ad Dwell	0.007**	0.007*	0.007**	0.006*	0.008**	0.005	0.005**	0.005**
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)
Observations	3,925	3,875	3,875	3,925	3,925	3,925	3,925	3,925
R^2	0.136	0.281	0.174	0.137	0.168	0.174	0.136	0.136
FE:								
Price x Brand	Y	Y	Y	Y	Y	Y	Y	
Step Order	Y	Y	Y	Y	Y	Y	Y	Y
Country x Device	Y	Y	Y	Y	Y	Y	Y	Y
Dem. Controls	Y	Y	Y	Y	Y	Y	Y	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include a quartic polynomial in log of average time that an average article was visible for by each individual, step order and device x country fixed effects, fixed effects for individual covariates (income, gender, education, age, and self-reported political leaning), and brand by price fixed effects. A few observations have missing values of precision and accuracy variables; we drop those observations from the analysis in specifications with 1/precision and 1/accuracy weights. Standard errors clustered at the individual level.

Table 7: Estimates of advertising effects on recall and purchase: Article Dwell IV, Robustness in Ad Dwell Measurements

	Purchase ($\hat{\lambda}$)							
	Main	Re-weighted observations by					Adjusted Ad Dwell	
		1/precision	1/accuracy	1/gaze duration	Hit Rate	1/time since calibr.	Interior	Away from Text
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ad Dwell	0.028**	0.017*	0.023**	0.028**	0.019*	0.022	0.030**	0.038**
	(0.011)	(0.010)	(0.011)	(0.012)	(0.011)	(0.014)	(0.012)	(0.015)
Observations	3,925	3,875	3,875	3,925	3,925	3,925	3,925	3,925
R^2	0.119	0.277	0.165	0.118	0.164	0.164	0.096	0.073
	First Stage							
Article Dwell	0.011***	0.018***	0.011***	0.011***	0.012***	0.011***	0.011***	0.009***
	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
Observations	3,925	3,875	3,875	3,925	3,925	3,925	3,925	3,925
R^2	0.220	0.691	0.240	0.217	0.234	0.272	0.144	0.128
1st Stage Incr. F-Stat	48.52	34.82	72.81	44.07	100.4	61.32	33.06	36.5
FE:								
Price x Brand	Y	Y	Y	Y	Y	Y	Y	Y
Step Order	Y	Y	Y	Y	Y	Y	Y	Y
Country x Device	Y	Y	Y	Y	Y	Y	Y	Y
Dem. Controls	Y	Y	Y	Y	Y	Y	Y	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include a quartic polynomial in log of average time that an average article was visible for by each individual, step order and device x country fixed effects, fixed effects for individual covariates (income, gender, education, age, and self-reported political leaning), and brand by price fixed effects. A few observations have missing values of precision and accuracy variables; we drop those observations from the analysis in specifications with 1/precision and 1/accuracy weights. Standard errors clustered at the individual level.

Both the IV and the first stage estimates are not statistically different from our main results. IV estimates vary from 0.017 to 0.038 across specifications, whereas in Section 7.2 we obtained an estimate of 0.028. Similarly, the first-stage estimates vary from 0.009 to 0.018, whereas in Section 7.2 the estimate was 0.011. The stability of the first stage coefficients is especially re-assuring, since these coefficients correspond to the $\hat{\gamma}$ estimate, the attention spillover coefficient from Table 2. Thus, we show that the complementarity in consumers' attention to articles and ads is not an artifact of the measurement error, and instead captures true spillover in attention.

Appendix Tables 26–29 replicate the robustness checks separately for desktop and mobile devices.

9 Managerial Implications

Our paper has three sets of findings that lend themselves to managerial implications. First, attention to ads can be measured and leads to higher ad recall and product purchases, providing a way to measure ad effectiveness and price display advertising. Second, attention to articles has a positive spillover to ads placed next to them, highlighting the value of high-quality content. Third, "hard news" article content does not make ads less effective, cautioning against the practice of blunt " block lists" of advertisers. We consider each of these in turn.

We can use our results to calculate rough estimates of the costs and benefits of online display ads. First, we discuss the benefits. In our experiment, the ads on each page had an average dwell time of about 2.76 seconds per individual (i.e., the time individuals are attentive to the ad). At the mean, this attention increases the probability of purchase by $2.76 \times 0.007 \approx 0.02$, or about 2%.[33] In the US, for instance, the opportunity cost to individuals of acquiring the voucher (the amount of cash individuals had to forgo, or the "price" of the voucher) was on average \$5. Therefore, we take the revenue for the brand from purchase to \$5. This implies that an ad is worth $5 \times 0.02 = 10$ cents of revenue per person exposed to the ad, or \$100 for 1,000 people.[34]

On the cost side, the advertising industry typically uses the metric of a "cost per mille" (CPM, or cost per 1,000 impressions). For a digital inventory, this is difficult to assess because it is

[33] Here we use the OLS estimates from Table 3. The IV estimates would imply an even higher value of advertising.

[34] We are considering only revenue, not profit, since we have no estimate of the cost to the brand of producing and supplying the goods.

the result of an auction every time an ad is available rather than the setting of a price in general. Things are further complicated because advertisers tend to pay for targeting information (e.g., to ensure that a particular ad is shown to individuals who, based on their known characteristics, are likely to be interested in the product), which further influences the cost. Still, Lumen Research shared with us their estimate of the cost per *attentive* 1000 views (aCPM), which is £21.88 ($\approx$ \$30) on desktops and £13.54 ($\approx$ \$19) on mobile devices. On top of this, we would have to include technology and agency fees – that is, the cost of creating the ads and employing marketers to manage the advertising agencies. However, on the whole, these figures suggest that advertising is likely worth its cost.

Our second set of results shows that there is a positive attention spillover from articles to ads. These results emphasize the value of good, captivating news content – not only does such content drive more visitors to news outlets and increase their reputation, it also increases the effectiveness of advertising on news outlets' webpages. Thus, by investing in the quality of news content, publishers can charge higher CPM rates to advertisers. These findings provide business justifications against the practice of "clickbait" (using catchy titles or images to entice users to visit low-quality articles). Instead, the result suggests that publishers should be incentivized to invest in more captivating and high-quality news content, even when only considering ad revenue. Our results on the "political mismatch" between outlet and readers further corroborate this idea: newspapers that cater to their audiences attract valuable attention to the article that spills over to the ad.

The third managerial implication that arises from our results is a word of caution when it comes to block lists that often do not allow ads to be placed next to "hard news." In our experiment, these were articles associated with the COVID-19 pandemic and the BLM protests. Our results reject the hypothesis of a negative effect of hard news *per se* on either ad recall or brand purchase. There may still be other reasons, such as brand safety [marketingweek.com, 2017], or preferences and career concerns of brand managers [Gordon et al., 2021b], to limit exposure of ads to certain types of content. However, our results suggest that the current system might be too blunt or exhibit excessive risk aversion. Limiting the practice of block lists is particularly important at times of major societal events – e.g. pandemics, wars, and the fight against climate change – since block lists penalize news outlets for providing detailed coverage of these important issues and informing citizens.

Our results validate the importance of the attention of website visitors for display advertising effectiveness, which can be "priced in" by the publishers and platforms. This view is aligned with the current thinking in the media industry. For instance, Mail Metro Media, which represents several UK's media brands (such as The Daily Mail, i newspaper, and The Telegraph) created a "high attention" package of advertising, for which they charge a price premium to brands [dmgmedia.co.uk, 2021]. A similar program is run by Ozone, the aggregated selling house used by The Guardian. Again, they charge a premium on their advertising inventory which is justified in part by the higher attention their ads receive because of the intensity of the engagement with the content [ozoneproject.com, 2021]. This does not seem to be only a sell-side or online news phenomenon. Havas, one of the biggest media buying networks in the world, has adopted an explicit position that it will pay more for the quality of attention an ad receives [The Media Leader, 2022]. Like our method, these pricing strategies and measures of ad effectiveness benefit from a novel approach of leveraging the intensive margin of attention to ads, rather than an extensive margin of showing or not showing an ad on the page. McGranaghan et al. [2022] discuss similar strategies for incorporating attention metrics into the measures of ad effectiveness and pricing for TV ads.

10 Conclusions

This paper has used measures of attention obtained with eye-tracking to estimate advertising effectiveness in online markets. We run an experiment that focuses on display advertising online, in which ads are shown next to articles. We showed that more engaging news articles generate positive spillovers of attention from the news to the ads. This incremental ad attention increases the probability that the advertised brands are correctly recalled and subsequently purchased.

There are several important caveats to keep in mind regarding the external validity of our results, typical for similar experimental settings. First, we asked individuals to make an *immediate* purchase decision, so we are likely overestimating the effect a real ad would have on purchases. We note, however, that the brand-specific vouchers that individuals could obtain were valid for 1 year or more, so *consumption* does not need to be immediate, and hence possibly mitigates this bias.

Second, we may be underestimating the impact of ads, since our ads are not targeted to specific individuals but are instead shown at random. We relied on the representativeness of the panel selected by a specialist supplier of research and marketing panels, and we chose brands that are of sufficient appeal to large audiences. We cannot estimate the effectiveness of targeted ads (as this was not the goal of our experiment), since this would require access to one of the algorithms that assign ads to readers online, which we do not possess.

Notwithstanding these limitations, we hope that our experimental design will prompt more research on the drivers and effects of digital attention. Tools such as eye-tracking software are now increasingly precise and available at scale in realistic settings to measure this.

References

Jae-Hyeon Ahn, Yoon-Soo Bae, Jaehyeon Ju, and Wonseok Oh. Attention adjustment, renewal, and equilibrium seeking in online search: An eye-tracking approach. *Journal of Management Information Systems*, 35(4):1218–1250, 2018.

Joshua David Angrist, Guido W Imbens, and Alan B Krueger. Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14(1):57–67, 1999.

Anocha Aribarg, Rik Pieters, and Michel Wedel. Raising the bar: Bias adjustment of recognition tests in advertising. *Journal of Marketing Research*, 47(3):387–400, 2010.

K Carrie Armel, Aurelie Beaumel, and Antonio Rangel. Biasing simple choices by manipulating relative visual attention. *Judgment and Decision making*, 3(5):396–403, 2008.

Gary S Becker and Kevin M Murphy. A simple theory of advertising as a good or bad. *The Quarterly Journal of Economics*, 108(4):941–964, 1993.

Jonah Berger, Wendy W Moe, and David Schweidel. What leads to longer reads? psychological drivers of reading online content. *ACR North American Advances*, 2019.

James Berry, Greg Fischer, and Raymond Guiteras. Eliciting and utilizing willingness to pay: Evidence from field trials in northern ghana. *Journal of Political Economy*, 128(4):1436–1473, 2020.

S Adam Brasel and James Gips. Breaking through fast-forwarding: Brand information and visual attention. *Journal of Marketing*, 72(6):31–48, 2008.

Isabelle Brocas, Juan D Carrillo, Stephanie W Wang, and Colin F Camerer. Imperfect choice or imperfect attention? understanding strategic thinking in private information games. *Review of Economic Studies*, 81(3):944–970, 2014.

Bobby J Calder, Edward C Malthouse, and Ute Schaedel. An experimental study of the relationship between online engagement and advertising effectiveness. *Journal of interactive marketing*, 23(4):321–331, 2009.

Colin F Camerer, Eric Johnson, Talia Rymon, and Sankar Sen. Cognition and framing in sequential bargaining for gains and losses. *Frontiers of Game Theory*, 104:27–47, 1993.

Moran Cerf, Jonathan Harel, Wolfgang Einhäuser, and Christof Koch. Predicting human gaze using low-level saliency combined with face detection. *Advances in neural information processing systems*, 20, 2007.

Pierre Chandon, J Wesley Hutchinson, Eric T Bradlow, and Scott H Young. Does in-store marketing work? effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *Journal of marketing*, 73(6):1–17, 2009.

ComScore. Revisited: Media consumption during the coronavirus pandemic. https://www.comscore.com/Insights/Blog/Revisited-Media-Consumption-during-the-Coronavirus-Pandemic [Acccessed: January 17, 2023], 2020.

Gordon B Dahl, Andreas Ravndal Kostøl, and Magne Mogstad. Family welfare cultures. *The Quarterly Journal of Economics*, 129(4):1711–1752, 2014.

Peter J Danaher and Guy W Mullarkey. Factors affecting online advertising recall: A study of students. *Journal of advertising research*, 43(3):252–267, 2003.

digiday.com. Coronavirus climbs up keyword block lists, squeezing news publishers' programmatic revenues. https://digiday.com/media/coronavirus-climbs-keyword-block-lists-squeezing-news-publishers-programmatic-revenues/ [Acccessed: January 17, 2023], 2020.

dmgmedia.co.uk. Mail metro media unveils centre for attention. https://www.dmgmedia.co.uk/news/mail-metro-media-unveils-centre-for-attention/ [Acccessed: January 31, 2023], 2021.

Will Dobbie, Jacob Goldin, and Crystal S Yang. The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108(2):201–40, 2018.

Xavier Drèze and François-Xavier Hussherr. Internet advertising: Is anybody watching? *Journal of interactive marketing*, 17(4):8–23, 2003.

Millie Elsen, Rik Pieters, and Michel Wedel. Thin slice impressions: how advertising evaluation depends on exposure duration. *Journal of Marketing Research*, 53(4):563–579, 2016.

emarketer.com. Estimated short-term change in average digital ad spending due to the coronavirus pandemic. https://www.emarketer.com/chart/235710/estimated-short-term-change-average-digital-ad-spending-due-coronavirus-pandemic-according-us-agency-brand-ad-buyers-by-format-march-2020-change [Acccessed: January 17, 2023], 2020.

Minou Ghaffari and Susann Fiedler. The power of attention: Using eye gaze to predict other-regarding and moral choices. *Psychological science*, 29(11):1878–1889, 2018.

Anindya Ghose, Avi Goldfarb, and Sang Pil Han. How is the mobile internet different? search costs and local activities. *Information Systems Research*, 24(3):613–631, 2013.

Avi Goldfarb and Catherine Tucker. Online display advertising: Targeting and obtrusiveness. *Marketing Science*, 30(3):389–404, 2011.

Daniel G Goldstein, R Preston McAfee, and Siddharth Suri. The effects of exposure time on memory of display advertisements. *Proceedings of the 12th ACM conference on Electronic commerce*, pages 49–58, 2011.

Daniel G Goldstein, R Preston McAfee, and Siddharth Suri. Improving the effectiveness of time-based display advertising. *ACM Transactions on Economics and Computation (TEAC)*, 3(2): 1–20, 2015.

Brett Gordon, Robert Moakler, and Florian Zettelmeyer. Close enough? a large-scale exploration of non-experimental approaches to advertising measurement. *Working paper*, 2021a.

Brett R Gordon, Kinshuk Jerath, Zsolt Katona, Sridhar Narayanan, Jiwoong Shin, and Kenneth C Wilbur. Inefficiencies in digital advertising markets. *Journal of Marketing*, page 0022242920913236, 2021b.

James J Heckman and James M Snyder Jr. Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *The Rand Journal of Economics*, 28:S142, 1997.

Emily Higgins, Mallorie Leinenger, and Keith Rayner. Eye movements when viewing advertisements. *Frontiers in psychology*, 5:210, 2014.

Paul R Hoban and Randolph E Bucklin. 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, 2015.

Kenneth Holmqvist, Jana Holsanova, Mari Barthelson, and Daniel Lundqvist. Reading or scanning? a study of newspaper and net paper reading. In *The Mind's Eye*, pages 657–670. Elsevier, 2003.

Wayne D Hoyer and Steven P Brown. Effects of brand awareness on choice for a common, repeat-purchase product. *Journal of consumer research*, 17(2):141–148, 1990.

Jason Huang, David Reiley, and Nick Riabov. Measuring consumer sensitivity to audio advertising: A field experiment on pandora internet radio. *Available at SSRN 3166676*, 2018.

IAB UK. Covid-19 x keyword blocking: 9 tips for advertisers. https://www.iabuk.com/news-article/covid-19-x-keyword-blocking-9-tips-advertisers [Acccessed: January 17, 2023], 2020.

Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11): 1254–1259, 1998.

Chris Janiszewski. The influence of display characteristics on visual exploratory search behavior. *Journal of consumer research*, 25(3):290–301, 1998.

Przemysław Jeziorski and Sridhar Moorthy. Advertiser prominence effects in search advertising. *Management Science*, 64(3):1365–1383, 2018.

Garrett Johnson, Randall A Lewis, and Elmar Nubbemeyer. The online display ad effectiveness funnel & carryover: Lessons from 432 field experiments. *Available at SSRN 2701578*, 2017a.

Garrett A Johnson, Randall A Lewis, and Elmar I Nubbemeyer. Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research*, 54(6):867–884, 2017b.

Ulrich Kaiser and Julian Wright. Price structure in two-sided markets: Evidence from the magazine industry. *International Journal of Industrial Organization*, 24(1):1–28, 2006.

Gabriella Kazai, Iskander Yusof, and Daoud Clarke. Personalised news and blog recommendations based on user location, facebook and twitter user profiling. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 1129–1132, 2016.

Mehreen Khurram, Faisal Qadeer, and Muhammad Sheeraz. The role of brand recall, brand recognition and price consciousness in understanding actual purchase. *Journal of Research in Social Sciences*, 6(2):219–241, 2018.

Max Kilger and Ellen Romer. Do measures of media engagement correlate with product purchase likelihood? *Journal of Advertising Research*, 47(3):313–325, 2007.

Daniel T Knoepfle, Joseph Tao-yi Wang, and Colin F Camerer. Studying learning in games using eye-tracking. *Journal of the European Economic Association*, 7(2-3):388–398, 2009.

Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*, pages 115–141. Springer, 1987.

Michal Kolesar. Estimation in an instrumental variables model with treatment effect heterogeneity. *Unpublished Working Paper*, 2013.

Dmitry Lagun and Mounia Lalmas. Understanding user attention and engagement in online news reading. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 113–122, 2016.

JooWon Lee and Jae-Hyeon Ahn. Attention to banner ads and their effectiveness: An eye-tracking approach. *International Journal of Electronic Commerce*, 17(1):119–137, 2012.

Randall A Lewis and David H Reiley. 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, 2014.

Xiau Liu, Matthew Shum, and Kosuke Ueteke. Passive and active attention to baseball telecasts: Implications for content (re-)design. *Available on SSRN (https://ssrn.com/abstract=3717894)*, 2021.

Gerald L Lohse. Consumer eye movement patterns on yellow pages advertising. *Journal of Advertising*, 26(1):61–73, 1997.

Emma K Macdonald and Byron M Sharp. Brand awareness effects on consumer decision making for a common, repeat purchase product: A replication. *Journal of Business Research*, 48(1):5–15, 2000.

Scott B MacKenzie. The role of attention in mediating the effect of advertising on attribute importance. *Journal of Consumer Research*, 13(2):174–195, 1986.

marketingweek.com. Should brands be pulling digital ad spend over “funding terror” claims? https://www.marketingweek.com/brands-pulling-digital-ad-spend-terror-funding-claims/ [Acccessed: January 17, 2023], 2017.

José Martins, Catarina Costa, Tiago Oliveira, Ramiro Gonçalves, and Frederico Branco. How smartphone advertising influences consumers’ purchase intention. *Journal of Business Research*, 94:378–387, 2019.

Matthew McGranaghan, Jura Liaukonyte, and Kenneth C Wilbur. How viewer tuning, presence, and attention respond to ad content and predict brand search lift. *Marketing Science*, 2022.

Martin Meißner, Andres Musalem, and Joel Huber. Eye tracking reveals processes that enable conjoint choices to become increasingly efficient with practice. *Journal of Marketing Research*, 53(1):1–17, 2016.

Milica Milosavljevic and Moran Cerf. First attention then intention: Insights from computational neuroscience of vision. *International Journal of advertising*, 27(3):381–398, 2008.

Sridhar Narayanan and Kirthi Kalyanam. Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Science*, 34(3):388–407, 2015.

Vidhya Navalpakkam, Justin Rao, and Malcolm Slaney. Using gaze patterns to study and predict reading struggles due to distraction. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 1705–1710. 2011.

Howard Kenneth Nixon. *Attention and interest in advertising.* Number 72. 1924.

ozoneproject.com. Ozone study shows clear link between attention paid to quality content and accompanying ads. https://www.ozoneproject.com/blog/ozone-study-shows-clear-link-between-attention-paid-to-quality-content-and-accompanying-ads [Acccessed: January 31, 2023], 2021.

Philip Pärnamets, Petter Johansson, Lars Hall, Christian Balkenius, Michael J Spivey, and Daniel C Richardson. Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences*, 112(13):4170–4175, 2015.

Rik Pieters and Michel Wedel. Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *Journal of marketing*, 68(2):36–50, 2004.

Rik Pieters and Michel Wedel. Goal control of attention to advertising: The yarbus implication. *Journal of consumer research*, 34(2):224–233, 2007.

Rik Pieters, Michel Wedel, and Jie Zhang. Optimal feature advertising design under competitive clutter. *Management Science*, 53(11):1815–1828, 2007.

Rik Pieters, Michel Wedel, and Rajeev Batra. The stopping power of advertising: Measures and effects of visual complexity. *Journal of Marketing*, 74(5):48–60, 2010.

Emily Pitler and Ani Nenkova. Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the 2008 conference on empirical methods in natural language processing*, pages 186–195, 2008.

Keith Rayner, Caren M Rotello, Andrew J Stewart, Jessica Keir, and Susan A Duffy. Integrating text and pictorial information: eye movements when looking at print advertisements. *Journal of experimental psychology: Applied*, 7(3):219, 2001.

Cleo D Redline and Christopher P Lankford. Eye-movement analysis: a new tool for evaluating the design of visually administered instruments (paper and web). 2001.

Elena Reutskaja, Rosemarie Nagel, Colin F Camerer, and Antonio Rangel. Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, 101(2):900–926, 2011.

J Edward Russo and France Leclerc. An eye-fixation analysis of choice processes for consumer nondurables. *Journal of consumer research*, 21(2):274–290, 1994.

Oliver J Rutz, Randolph E Bucklin, and Garrett P Sonnier. A latent instrumental variables approach to modeling keyword conversion in paid search advertising. *Journal of Marketing Research*, 49(3):306–319, 2012.

Navdeep S Sahni. Effect of temporal spacing between advertising exposures: Evidence from online field experiments. *Quantitative Marketing and Economics*, 13(3):203–247, 2015.

Desirée Schmuck, Miriam Tribastone, Jörg Matthes, Franziska Marquart, and Eva Maria Bergel. Avoiding the other side? an eye-tracking study of selective exposure and selective avoidance effects in response to political advertising. *Journal of Media Psychology: Theories, Methods, and Applications*, 2019.

Tim Schneegans, Matthew D Bachman, Scott A Huettel, and Hauke Heekeren. Exploring the potential of online webcam-based eye tracking in decision-making research and influence factors on data quality. 2021.

Noel Scott, Christine Green, and Sheranne Fairley. Investigation of the use of eye tracking to examine tourism advertising effectiveness. *Current Issues in Tourism*, 19(7):634–642, 2016.

Kilian Semmelmann and Sarah Weigelt. Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, 50(2):451–465, 2018.

Savannah Wei Shi and Michael Trusov. The path to click: Are you on it? *Marketing Science*, 40 (2):344–365, 2021.

Jaana Simola, Jarmo Kuisma, Anssi Öörni, Liisa Uusitalo, and Jukka Hyönä. The impact of salient advertisements on reading and attention on web pages. *Journal of Experimental Psychology: Applied*, 17(2):174, 2011.

Andrey Simonov and Shawndra Hill. Competitive advertising on brand search: Traffic stealing and customer selection. *Marketing Science, Forthcoming*, (18-59), 2021.

Andrey Simonov, Chris Nosko, and Justin M Rao. Competition and crowd-out for brand keywords in sponsored search. *Marketing Science*, 37(2):200–215, 2018.

Statista. Online advertising revenue in the united states from 2000 to 2021. https://www.statista.com/statistics/183816/us-online-advertising-revenue-since-2000/ [Acccessed: January 17, 2023], 2023.

Iréne Stenfors, Jan Morén, and Christian Balkenius. Behavioural strategies in web interaction: A view from eye-movement research. In *The Mind's Eye*, pages 633–644. Elsevier, 2003.

George J Stigler and Gary S Becker. De gustibus non est disputandum. *The american economic review*, 67(2):76–90, 1977.

Thales S Teixeira, Michel Wedel, and Rik Pieters. Moment-to-moment optimal branding in tv commercials: Preventing avoidance by pulsing. *Marketing Science*, 29(5):783–804, 2010.

The Guardian. Uk publishers losing digital ad revenue due to content 'blacklists'. https://www.theguardian.com/media/2020/jan/20/uk-publishers-losing-digital-ad-revenue-due-to-content-blacklists [Acccessed: January 17, 2023], 2020.

The Media Leader. Havas: news, science and sports sites get most user engagement. https://the-media-leader.com/havas-news-science-and-sports-sites-get-most-user-engagement/ [Acccessed: January 31, 2023], 2022.

Joan Treistman and John P Gregg. Visual, verbal, and sales responses to print ads. *Journal of Advertising Research*, 19(4):41–47, 1979.

Anna E Tuchman, Harikesh S Nair, and Pedro M Gardete. Television ad-skipping, consumption complementarities and the consumer demand for advertising. *Quantitative Marketing and Economics*, 16(2):111–174, 2018.

Christina Uhl, Nadia Abou Nabout, and Klaus Miller. How much ad viewability is enough? the effect of display ad viewability on advertising effectiveness. *arXiv preprint arXiv:2008.12132*, 2020.

Michel Wedel. Attention research in marketing: A review of eye-tracking studies. 2015.

Michel Wedel and Rik Pieters. Eye fixations on advertisements and memory for brands: A model and findings. *Marketing science*, 19(4):297–312, 2000.

Michel Wedel and Rik Pieters. A review of eye-tracking research in marketing. *Review of marketing research*, pages 123–147, 2017.

Michel Wedel, Rik Pieters, and John Liechty. Attention switching during scene perception: how goals influence the time course of eye movements across advertisements. *Journal of Experimental Psychology: Applied*, 14(2):129, 2008.

Kenneth C Wilbur. A two-sided, empirical model of television advertising and viewing markets. *Marketing science*, 27(3):356–378, 2008.

Kenneth C Wilbur. Advertising content and television advertising avoidance. *Journal of Media Economics*, 29(2):51–72, 2016.

Rick T Wilson, Daniel W Baack, and Brian D Till. Creativity, attention and the memory for brands: an outdoor advertising field study. *International Journal of Advertising*, 34(2):232–261, 2015.

Shunyao Yan, Klaus M Miller, and Bernd Skiera. How does the adoption of ad-blockers affect news consumption? *Available at SSRN*, 2020.

Xiaozhi Yang and Ian Krajbich. Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, 16(6):1486, 2021.

Xuebai Zhang and Shyan-Ming Yuan. An eye tracking analysis for video advertising: Relationship between advertisement elements and effectiveness. *IEEE access*, 6:10699–10707, 2018.

Online Appendix

A Experimental Details

A.1 Branded Ads

We showed ads for brands, rather than specific products. We selected brands that would be of general interest to a wide audience and that would be relatively easy to redeem with an e-voucher. We also chose brands for which we could find brand-specific vouchers.[35] We also ensured that the types of product categories would be similar between the two countries. The table below lists the brands chosen.

Type of product/Country	US	UK
Coffee shop	Starbucks	Starbucks
Coffee shop	Dunkin' Donuts	Costa
Clothing	Banana Republic	Primark
Clothing	GAP	H&M
Food	Domino's Pizza	Pizza Express
Food	Burger King	Wagamama
Bath products	Bath & Body Works	The Body Shop
DIY/Home improvement	Home Depot	B&Q

A.2 Articles

We report below the headlines of the articles that were chosen, split by country and by newspaper. We indicate with an asterisk (*) those articles that we classified as 'hard news'. We provide the URL to retrieve the full article (click on the headline).

The following articles were sourced from *The New York Times* (US):

Trump Aides Undercut Fauci as He Speaks Up on Virus Concerns*

Qualified Immunity Protection for Police Emerges as Flash Point Amid Protests*

Technology Bridges the Gap to Better Sight

[35] The vouchers were purchased on the GiftPay (https://www.giftpay.co.uk/) and Tango Card (https://www.tangocard.com/) websites.

What if the U.S. Bans TikTok?

The following articles were sourced from *USA Today* (US):

CDC adds runny nose, nausea to the growing list of COVID-19 symptoms*

'I thought this was a hoax': Patient, 30, dies after attending 'COVID party,' doctor says*

California officer under investigation for allegedly sharing 'vulgar image' of George Floyd; NAACP San Diego calls for his firing*

Johnny Depp accuses Amber Heard of hitting him with 'roundhouse punch' near end of their marriage

Pour by phone: Coca-Cola introduces contactless technology to pour your beverage

The following articles were sourced from *The Guardian* (UK):

NHS data reveals 'huge variation' in Covid-19 death rates across England*

Boris Johnson says face masks should be worn in shops in England*

Police apologise to woman told to cover up anti-Boris Johnson T-shirt*

Johnny Depp tells high court libel case how he lost $650m in earnings

How we met: 'It's 1,300 miles to Romania – the same as the number of pounds my phone bill was'

The following articles were sourced from the *Daily Mail* (UK):

People living in England's poorest areas are TWICE as likely to die of coronavirus than those in the wealthiest neighbourhoods, statistics show*

Two-thirds of Britons back Boris Johnson's refusal to 'take the knee' because people should not be 'bullied' into making 'gestures'*

Scooby Who? Great Dane's popularity falls to its lowest level in 50 years after peaking in the 1980s thanks to the Scooby Doo TV series

Are you a victim of 'batterygate?' Users with older iPhones may be eligible for a $25 settlement if their device was covertly slowed by the tech giant

A.3 Eye-tracking Technology

Details of the eye-tracking technology are summarized in the top panel of Figure 6. No hardware is needed. The software is Javascript code which is entirely removed from the participant's device after completion of the experiment. Before an eye-tracking session is started, the user is taken through a calibration procedure. During this procedure, the eye-tracker measures characteristics of the user's eyes and uses them together with an anatomical 3D eye model to calculate the gaze data. During the calibration, the user is asked to look at specific points on the screen (calibration dots). Several images of the eyes are collected and analyzed. The resulting information is then integrated into the eye model and the gaze point for each image sample is calculated. When the procedure is finished, the calibration process is illustrated by green lines of varying length (see the lower panel of Figure 6 for an example). Figure 7 contains a visual summary of the experimental protocol.

During the calibration procedure at the beginning of the experiment, as well as at two other points in the study – after the third and sixth articles – users do another validation procedure where the software measures the reliability of the data collected by eye-tracking. During these validation steps, the user is asked to look at five specific dots on the screen sequentially. Three main metrics determine the reliability of the eye-tracking data: "accuracy", "precision" and "gaze duration", as described in the main text. If on average per respondent one of these metrics climbs above the pre-specified thresholds – 300 CSS pixels for the accuracy and precision, and 100 milliseconds for gaze duration – the eye-tracking data in the articles before and after the validation step is typically considered "invalid" and is not used for the analysis. We rely on Lumen's rules for flagging participants for whom the data is deemed invalid and later confirm that valid observations indeed look much better on these metrics.

Figure 8 presents the densities of accuracy, precision, and gaze duration metrics for data deemed valid and invalid by Lumen Research. Figures on the left correspond to desktop devices, and figures on the right to mobile devices. For desktop devices, respondents flagged by Lumen Research as invalid have low quality data across all three metrics. A substantial share of these users tend to have values of accuracy and gaze duration higher than the thresholds, and some have precision higher than the threshold. In contrast, the vast majority of users have average

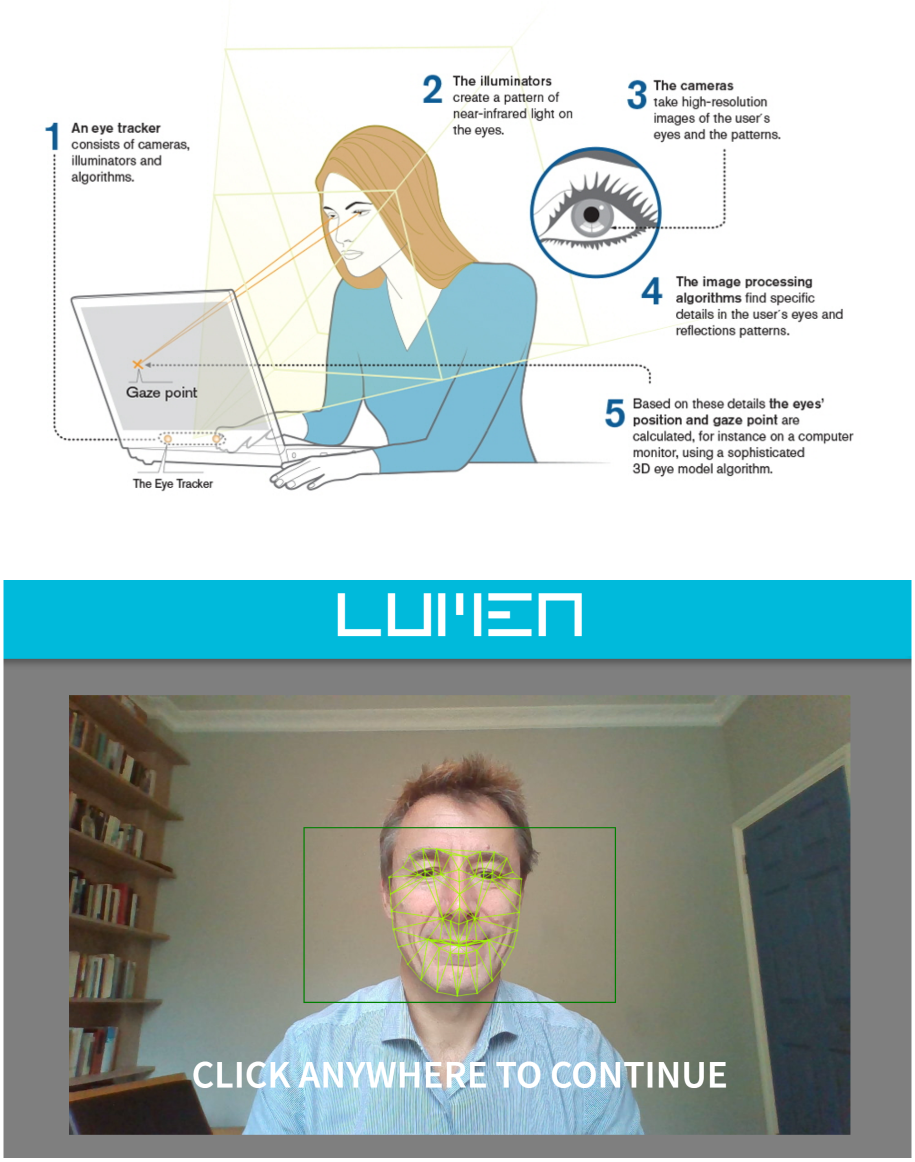


Figure 6: Eye-tracking technology

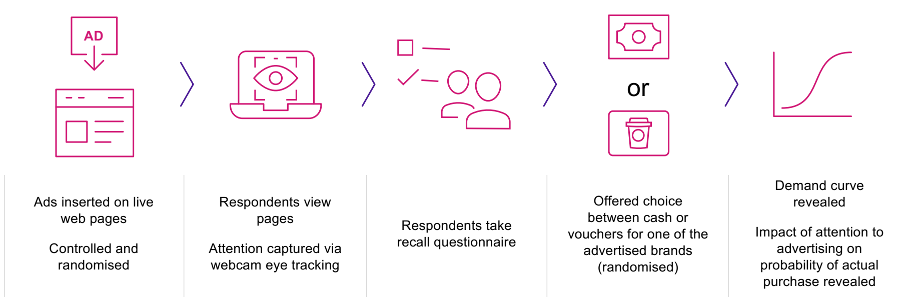
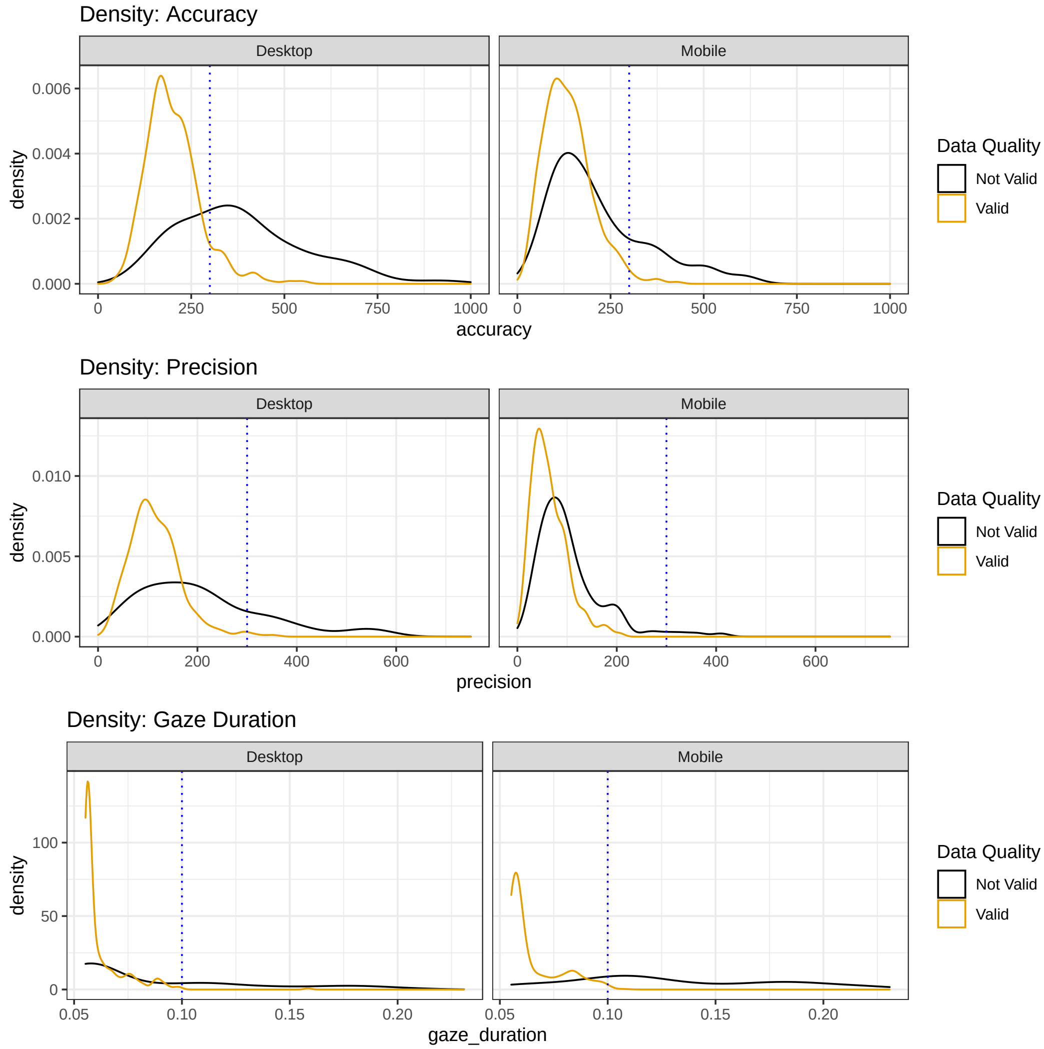


Figure 7: Research protocol

Figure 8: Densities of precision, accuracy, and gaze duration



Values of accuracy, precision, and gaze duration are computed as an average across the available validation steps per respondent. Values of accuracy and precision are in CSS pixels, and in seconds for gaze duration. The blue dotted line corresponds to thresholds above which Lumen typically flags data as invalid for analysis.

values of accuracy, precision, and gaze duration below the thresholds.

Similarly, on mobile devices, users that were deemed invalid tend to have worse metrics of accuracy, precision, and gaze duration compared to the respondent who are flagged as valid. For mobile users, gaze duration is the primary metric that crosses the quality threshold and disqualifies users from the analysis.

Comparing precision metrics across device types, we can see that eye-tracking is able to capture gazes more accurately and precisely on mobile devices – on average, accuracy for respondents flagged as valid on mobile devices is 137, compared to 202 on desktops, and precision is 64 on mobile devices, compared to 115 on desktops. However, if we normalize the metrics by devices' screen sizes, relative accuracy and precision on mobile devices are approximately the same as on desktops.

Table 8: Average Eye-Tracking Metrics across Devices for Valid Respondents

Device	Accuracy	Precision	Gaze_Duration
Desktop	201.75	115.13	0.06
Mobile	136.66	64.47	0.07

Values are computed as an average across the available validation steps per valid respondent. Values of accuracy and precision are in CSS pixels, and in seconds for gaze duration.

Table 9 presents the quality metrics of eye-tracking data across validation steps. As the study progresses, the accuracy and precision of eye-tracking data decrease. Average accuracy drops from 128 to 240, and average precision drops from 51 to 144. This is explained by the fact that even though respondents were instructed to sit still, eventually, as respondents read articles, they change their position. We check the robustness of our main results to accounting for the measurement error induced by this noise in Section 8.

A.4 Ethics Approval

The protocol received ethical approval from the Imperial College Research Ethics Committee (ICREC) and the Science Engineering Technology Research Ethics Committee (SETREC); SETREC reference: 20IC6104. The study was approved by SETREC on 12/06/20 and by the Joint Research Compliance Office on 19/06/20. The study was registered in the AEA RCT Registry as

Table 9: Average Eye-Tracking Metrics across Validation Steps for Valid Respondents

Validation Steps	Accuracy	Precision	Gaze_Duration
1	127.56	51.47	0.06
2	182.96	102.67	0.07
3	207.67	132.12	0.07
4	240.21	144.11	0.07

Values are computed as an average with the available validation steps per valid respondent. Values of accuracy and precision are in CSS pixels, and in seconds for gaze duration.

RCT ID AEARCTR-0006010.

B Validation

In this section we describe how we validated our measure of hard news and political slant. We issued a survey on Amazon Mechanical Turk (AMT) to 250 individuals in the UK and another 250 in the US. All participants used their desktops to take the survey.

In each country, we asked individuals to read the same articles used in the original experiment, in the desktop format. To avoid the effects of tiredness, we asked each individual to read only a random subset of 5 articles. We obtained about 2 thousand observations.

We showed individuals articles without ads and asked them to express their opinion about each article along three dimensions. First, how "upsetting" the article was on a Likert 1/5 scale. Second, how "interesting" the article was, again on a Likert 1/5 scale. Third, what the individual perceived as the political slant of each article (Left/Neutral/Right). We then computed the political slant of each article by assigning that article a score of +1, 0 or -1 if a participant considered the article to have, respectively, a slant that was right-wing, neutral, or left-wing. We then computed the extent to which each article was upsetting, interesting, and right-wing slanted, based on the average response from the AMT survey.

Table 10: Validation Summary Statistics

Newspaper	Hard	E[Upset]	SD[Upset]	E[Rightwing]	SD[Rightwing]	E[Interest]	SD[Interest]
Guardian	FALSE	0.304	0.268	-0.121	0.600	0.539	0.268
Guardian	FALSE	0.394	0.291	0.111	0.708	0.521	0.249
Guardian	TRUE	0.218	0.278	0.039	0.572	0.480	0.220
Guardian	TRUE	0.243	0.293	-0.418	0.731	0.583	0.237
Guardian	TRUE	0.481	0.304	-0.194	0.623	0.588	0.238
Mail	FALSE	0.183	0.258	-0.114	0.689	0.526	0.274
Mail	FALSE	0.180	0.287	0.130	0.626	0.426	0.296
Mail	TRUE	0.478	0.290	-0.237	0.788	0.573	0.240
Mail	TRUE	0.276	0.287	0.331	0.838	0.450	0.276
NYT	FALSE	0.223	0.266	-0.214	0.717	0.500	0.307
NYT	FALSE	0.156	0.263	-0.077	0.478	0.680	0.283
NYT	TRUE	0.502	0.313	-0.542	0.608	0.609	0.296
NYT	TRUE	0.401	0.313	-0.225	0.825	0.522	0.281
USAT	FALSE	0.135	0.265	-0.061	0.551	0.653	0.264
USAT	FALSE	0.384	0.315	-0.148	0.656	0.503	0.305
USAT	TRUE	0.408	0.303	-0.074	0.581	0.595	0.266
USAT	TRUE	0.571	0.328	-0.291	0.734	0.579	0.297
USAT	TRUE	0.605	0.312	-0.250	0.638	0.652	0.295

The outcomes of the AMT survey are shown in Table 10. The table shows, for each article, the newspaper and whether it was considered to be hard news. The table also includes the mean and standard deviation of the answers in the AMT survey regarding the extent to which each article was upsetting, right-wing, and interesting. The "upsetting" and "interesting" articles variables are rescaled to be between 0 and 1.

Table 11 shows a regression of the "upsetting" score that participants gave to the articles in the AMT survey on our subjective definition of hard news with respect to those articles. Observation are on the article-respondent pairs level. The positive coefficient suggests a validation of our subjective definition. We get a similar result if we include respondent fixed effects.

Table 11: Validation of Hard News Measure from AMT survey

	Dependent variable:
	Article Upsetting (0-1)
Hard News	0.1804***
	(0.0139)
Constant	0.2360***
	(0.0103)
Observations	2,016
R^2	0.0774

Note: *p<0.1; **p<0.05; ***p<0.01. One observation per article-respondent pair.

We used the same procedure for the political leaning of each newspaper. Table 12 shows that in the US, *USA Today* is deemed more right-wing than *The New York Times* (omitted). In the UK, the *Daily Mail* is deemed more right-wing than *The Guardian* (omitted). Also, the "political distance" between *The Guardian* and the *Daily Mail* is bigger than the distance between *The New York Times* and *USA Today*. These results are in line with how we coded the data. We get a similar result if we include respondent fixed effects.

Table 12: Validation of Political Leaning

	Article is Right-wing (from -1 to 1)	
	US	UK
	(1)	(2)
USAT	0.1121^{**}	
	(0.0489)	
Mail		0.2130^{***}
		(0.0621)
Constant	-0.2881^{***}	-0.1634^{***}
	(0.0331)	(0.0430)
Observations	778	588
R^2	0.0067	0.0197

Note: *p<0.1; **p<0.05; ***p<0.01. One observation per article-respondent pair.

C Sample Balance

In this section we should that the data is balanced on the basis of which observations were retained due to the connectivity and eye-tracking issues.

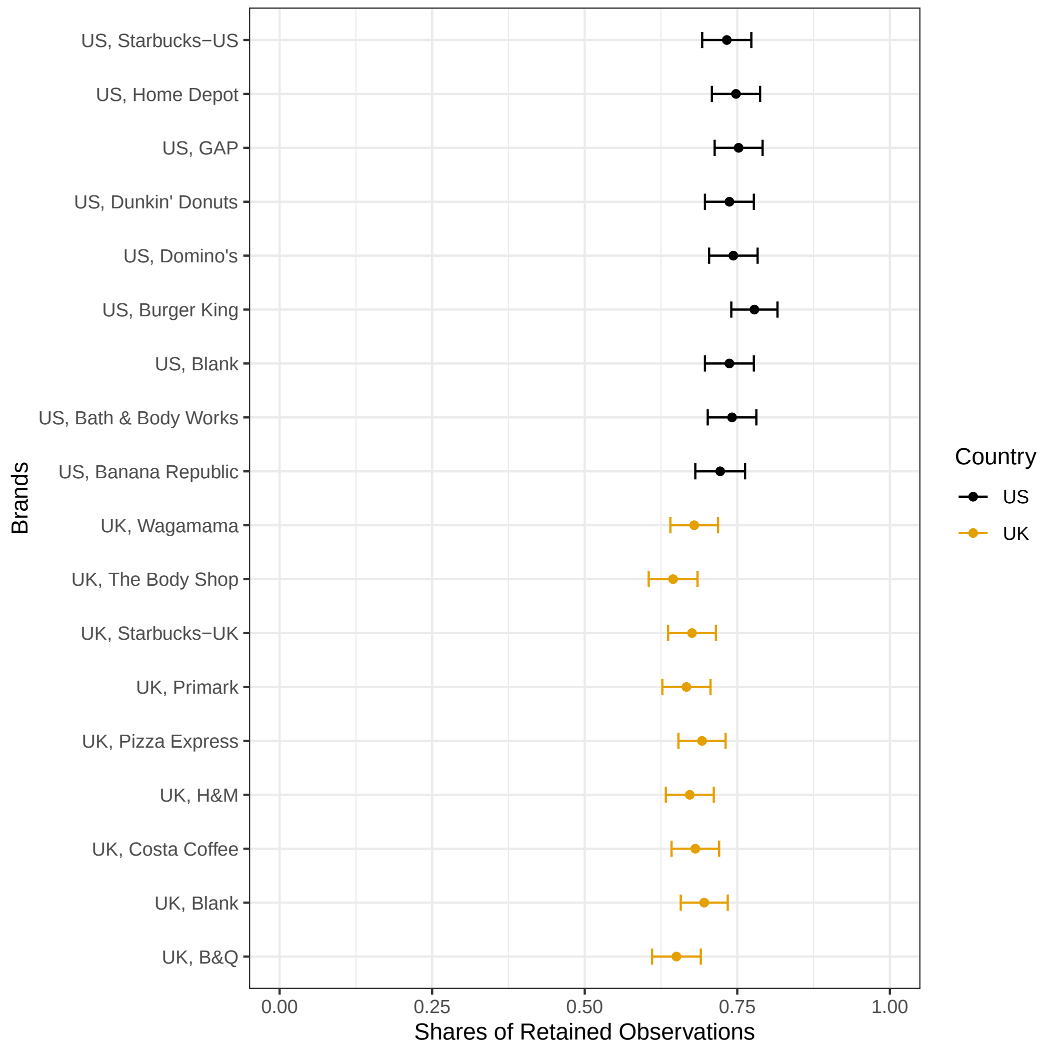
C.1 Connectivity Issues

In Table 13 we compare the sub-sample of individuals for whom we have data on all articles shown to them, relative to the individuals for whom we see only a subset of articles (due to connectivity issues). As described in the main text, an internet connectivity issue during the experiment could result in some lost data for that individual. However, each individual was still shown all articles and made all choices. Table 13 shows that individuals who did not experience any connectivity issues are more likely to be on desktops, more likely to be in the US, have slightly higher measures of *Ad Visible* (but not *Ad Dwell*) and slightly more likely to recall the ad. Otherwise, the two sub-samples are balanced on observables.

We further compare the balance of the retained and missed observations per each article and advertised brand. Since the order of articles and the ads was random, observations should be balanced along both dimensions after we account for the country of origin and device type. Figures 9 and 10 present the results. Observations are indeed well balanced on brands and articles once we make comparisons within the country and device. One notable exception are three articles on mobile phones in the UK (articles 3093, 3090, and 3089), all of which having systematically more connectivity issues (around 20% of data retained) than other UK mobile articles (around 70% of data retailed). We confirm that all our results hold if we exclude observations corresponding to these articles from the main analysis.

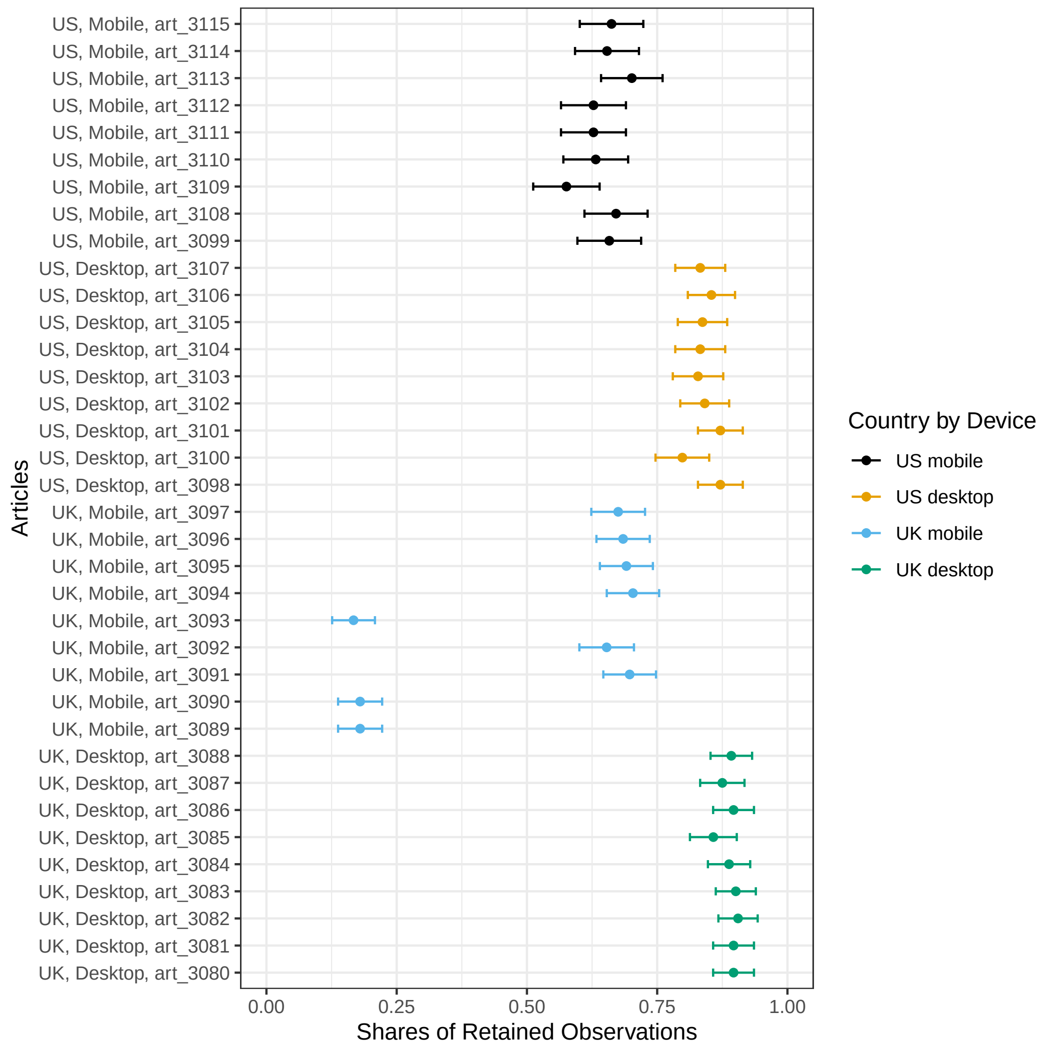
Figure 11 shows that connectivity issues were more pronounced later on in the study – around 90% of observations are retained during the readership of the first three articles (before the second validation step), around 62% of observations are retained before the third validation step (after the sixth article), and around 58% of observations are retained from the last three articles.

Figure 9: Balance of Observations per Brand Retained in the Study due to the Connectivity Issues



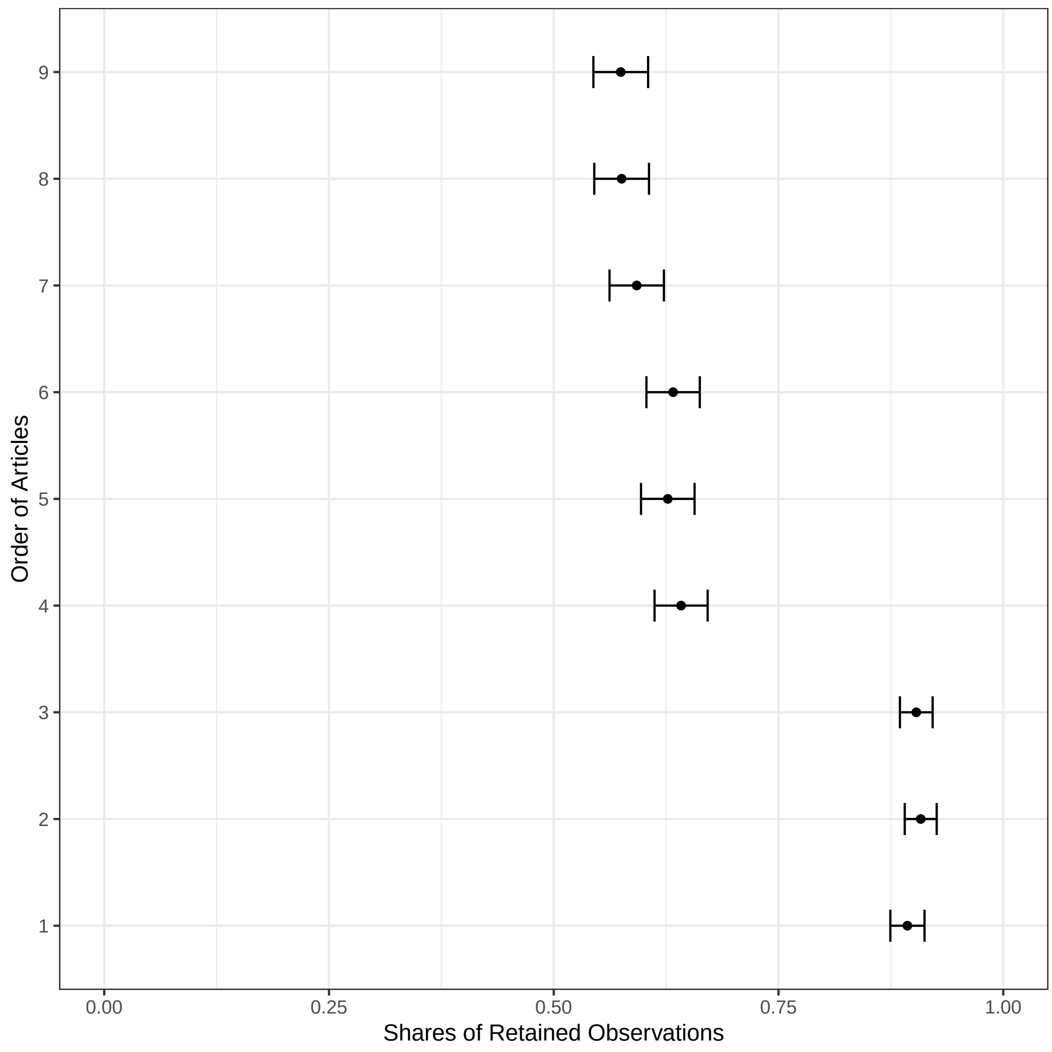
Shares are computed directly from the data. Bars correspond to 95% confidence intervals.

Figure 10: Balance of Observations per Article Retained in the Study due to the Connectivity Issues



Shares are computed directly from the data. Bars correspond to 95% confidence intervals.

Figure 11: Balance of Observations per step-order Retained in the Study due to the Connectivity Issues



Shares are computed directly from the data. Bars correspond to 95% confidence intervals.

Table 13: Individuals with and without connectivity issues.

	With issues		Without issues	
Variables	N	Mean	N	Mean
Desktop	538	0.22	466	0.73
		(0.02)		(0.02)
Female	538	0.55	466	0.58
		(0.02)		(0.02)
U.S.	538	0.39	466	0.54
		(0.02)		(0.02)
Hard News	538	0.55	466	0.55
		(0.01)		(0)
Article Visible (s)	538	149.42	466	142.24
		(5.58)		(6.16)
Ad Visible (s)	538	16.78	466	20.26
		(0.58)		(0.62)
Price (USD/GBP)	538	5.03	466	5.02
		(0.04)		(0.02)
Recall	538	0.42	466	0.51
		(0.02)		(0.02)
Buy	538	0.33	466	0.35
		(0.01)		(0.01)
Article Dwell (s)	381	79.17	327	73
		(4.17)		(3.89)
Ad Dwell (s)	381	2.87	327	2.7
		(0.13)		(0.11)
All valid eyetracking	538	0.71	466	0.7
		(0.02)		(0.02)

Standard errors are in brackets. One observation per individual.

C.2 Eye-Tracking Issues

In Table 14 we compare the sub-sample of individuals who had valid eyetracking data to those individuals that were deemed to have low-quality eye-tracking data, and therefore for whom we use only visibility data. As discussed in the main text and in Appendix A.3, this happens primarily if the accuracy, precision, and gaze duration frequency measures are too high, making the eye-tracking data unreliable (for instance, if the individuals move their head too much). We find that individuals with valid eye-tracking data are less likely to be on desktop and have slightly lower ad visible (but not lower article visible). Otherwise, the two subsamples are balanced on observables.

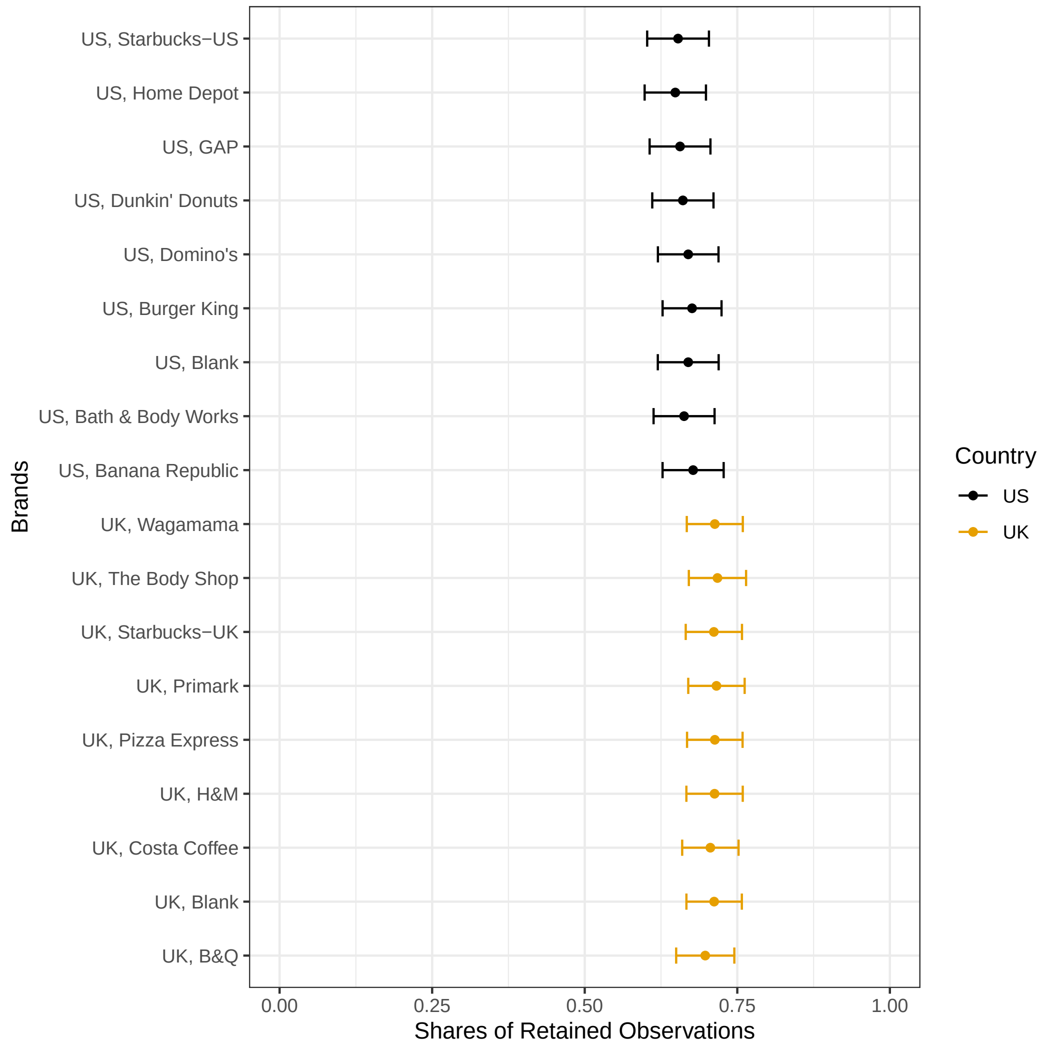
Table 14: Individuals with invalid and valid eyetracking data.

	Valid		Invalid	
Variables	N	Mean	N	Mean
Desktop	708	0.41	296	0.57
		(0.02)		(0.03)
Female	708	0.58	296	0.52
		(0.02)		(0.03)
U.S.	708	0.44	296	0.5
		(0.02)		(0.03)
Hard News	708	0.55	296	0.56
		(0.01)		(0.01)
Article Visible (s)	708	142.11	296	155.59
		(4.67)		(8.46)
Ad Visible (s)	708	17.51	296	20.51
		(0.48)		(0.89)
Price (USD/GBP)	708	5.01	296	5.06
		(0.03)		(0.04)
Recall	708	0.46	296	0.47
		(0.01)		(0.02)
Buy	708	0.34	296	0.35
		(0.01)		(0.02)
Article Dwell (s)	708	76.32	0	NaN
		(2.87)		(NA)
Ad Dwell (s)	708	2.79	0	NaN
		(0.09)		(NA)
All observations	708	0.46	296	0.47
		(0.02)		(0.03)

Standard errors are in brackets. One observation per individual.

We further compare the balance of the observations with reliable and unreliable eye-tracking data across articles and advertised brand. Since the order of articles and the ads was random,

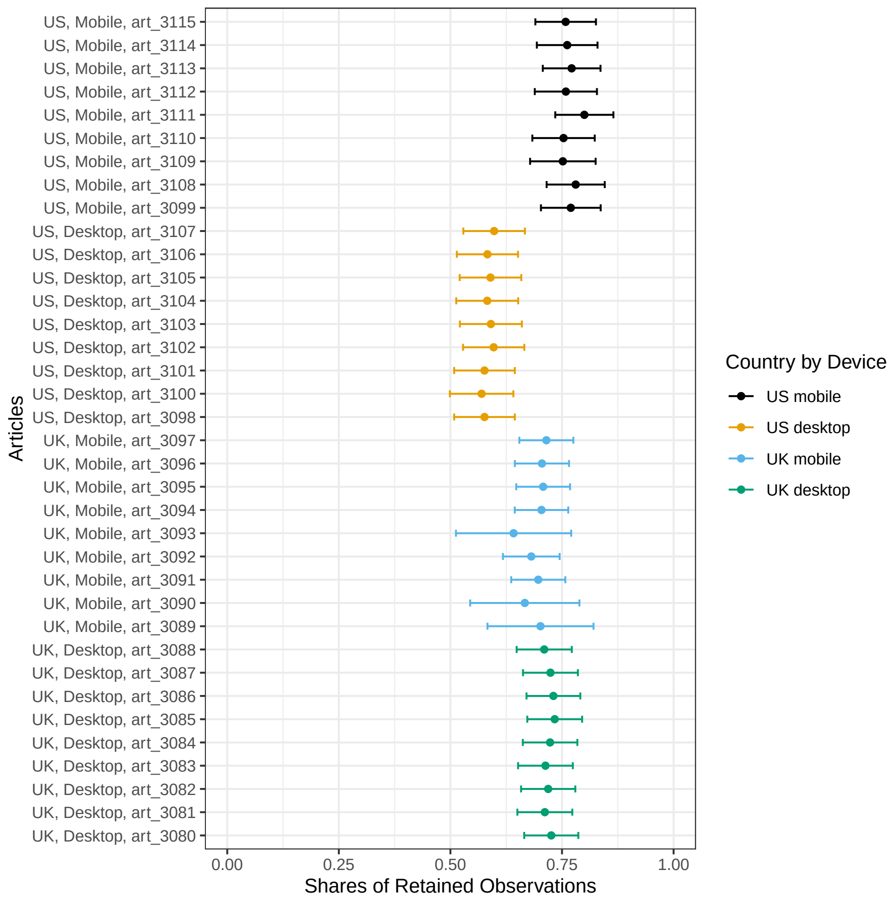
Figure 12: Balance of Observations per Brand Retained in the Study due to the Eye-Tracking Issues



Shares are computed directly from the data. Bars correspond to 95% confidence intervals.

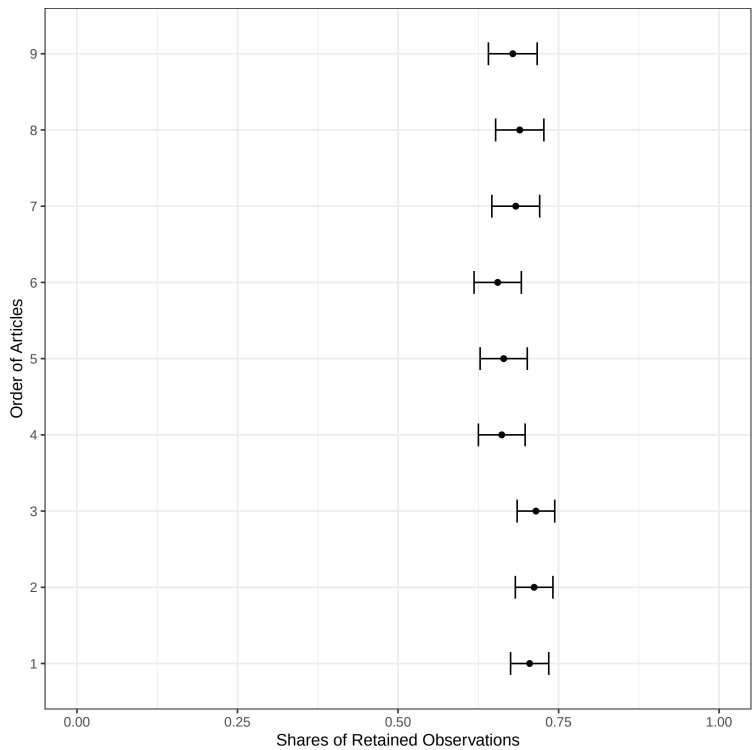
observations should be balanced along both dimensions after we account for the country of origin and device type. Figures 12 and 13 confirm this balance. Figure 14 further shows that eye-tracking issues were slightly more pronounced later on in the study, although differences are not statistically significant.

Figure 13: Balance of Observations per Article Retained in the Study due to the Eye-Tracking Issues



Shares are computed directly from the data. Bars correspond to 95% confidence intervals.

Figure 14: Balance of Observations per step-order Retained in the Study due to the Eye-Tracking Issues



Shares are computed directly from the data. Bars correspond to 95% confidence intervals.

D Summary Statistics By Device Type

Tables 15 and 16 show summary statistic for our sample by device type.

Table 15: Summary Statistics: Mobile

Statistic	N	Mean	St. Dev.	Min	Max
Desktop	2,810	0.000	0.000	0	0
Female	2,810	0.554	0.497	0	1
U.S.	2,810	0.478	0.500	0	1
Hard News	2,810	0.542	0.498	0	1
Article Visible (s)	2,810	144.957	171.568	20.130	1,894.635
Ad Visible (s)	2,495	13.436	12.608	0.000	291.905
Price (USD/GBP)	2,495	4.999	1.437	3.000	7.000
Recall	2,495	0.457	0.498	0.000	1.000
Buy	2,495	0.347	0.476	0.000	1.000
Article Dwell (s)	2,055	65.452	88.727	0.112	966.945
Ad Dwell (s)	1,824	2.690	3.134	0.000	40.214

Each observation is at the individual x article level.

Table 16: Summary Statistics: Desktop

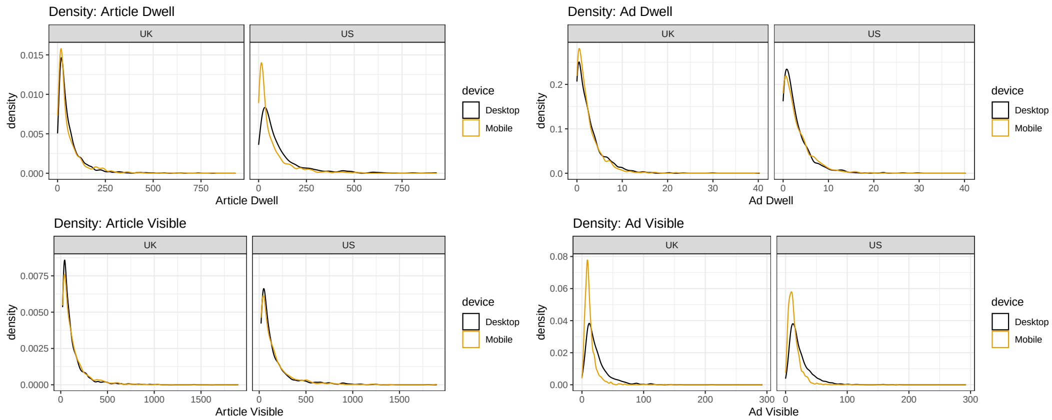
Statistic	N	Mean	St. Dev.	Min	Max
Desktop	3,621	1.000	0.000	1	1
Female	3,621	0.558	0.497	0	1
U.S.	3,621	0.487	0.500	0	1
Hard News	3,621	0.556	0.497	0	1
Article Visible (s)	3,621	142.015	167.606	22.599	1,805.904
Ad Visible (s)	3,212	23.370	19.226	0.000	210.103
Price (USD/GBP)	3,212	5.031	1.436	3.000	7.000
Recall	3,212	0.506	0.500	0.000	1.000
Buy	3,212	0.347	0.476	0.000	1.000
Article Dwell (s)	2,371	82.927	104.578	0.120	928.984
Ad Dwell (s)	2,101	2.810	3.184	0.000	29.895

Each observation is at the individual x article level.

E Additional Data Descriptives

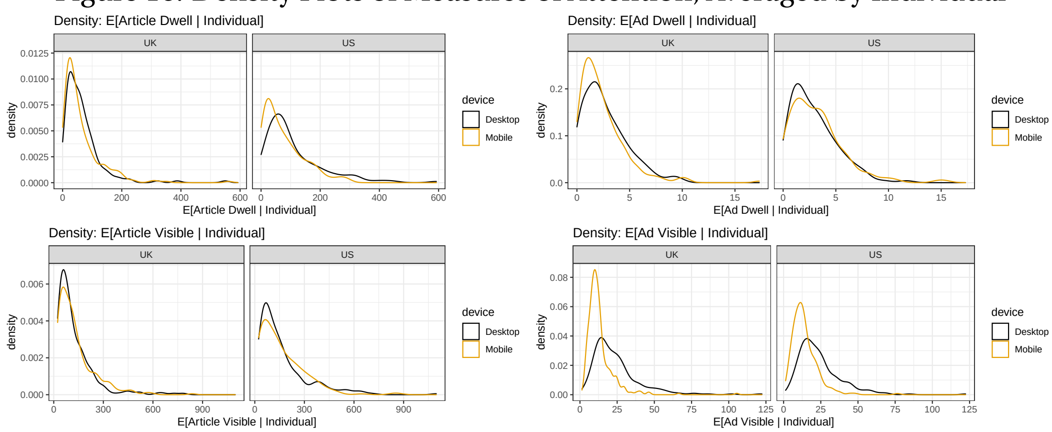
In this Section we provide additional descriptives of our raw data. Figure 15 shows density plots of our attention measures, computed across all observations in the data. Figure 16 shows the same density plots, but averaged by individual.

Figure 15: Density Plots of Measures of Attention, by Country and Device



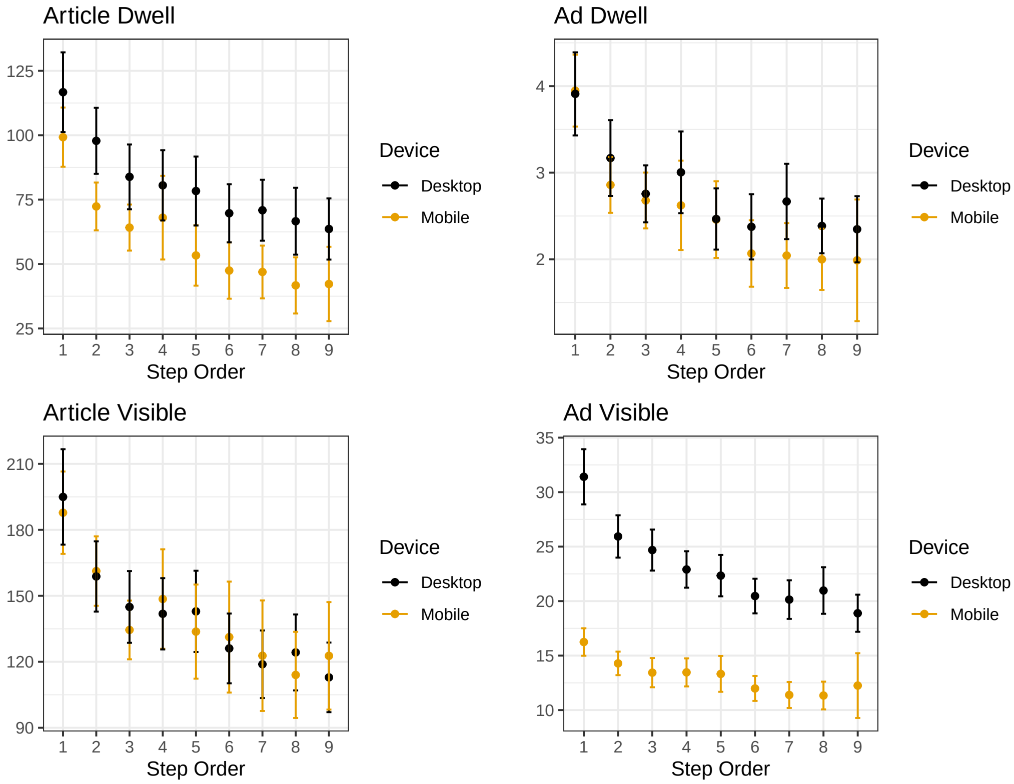
The plots show Article and Ad Dwell and Visible, computed across all observations in the data.

Figure 16: Density Plots of Measures of Attention, Averaged by Individual



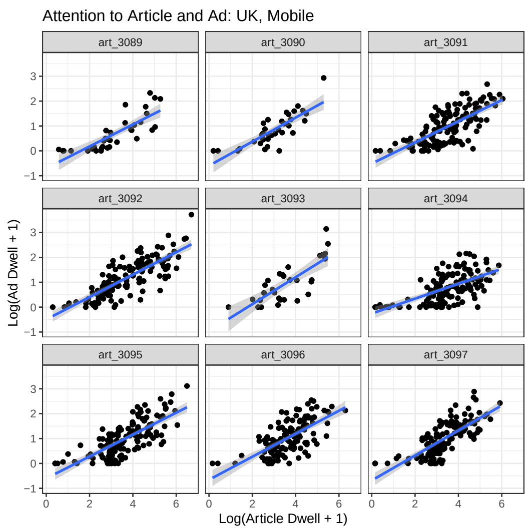
Density plots of measures of attention, averaged by individual and broken down by country and device.

Figure 17: Attention to Articles and Ads is Decreasing Throughout the Study

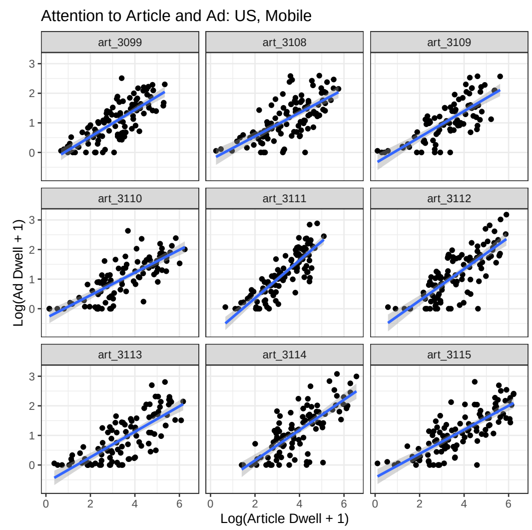


Attention devoted to articles and ads as a function of the "step" at which they are shown in the experiment. "step-order" refers to the order in the experiment in which an article and ad were shown. Bars correspond to 95% confidence intervals.

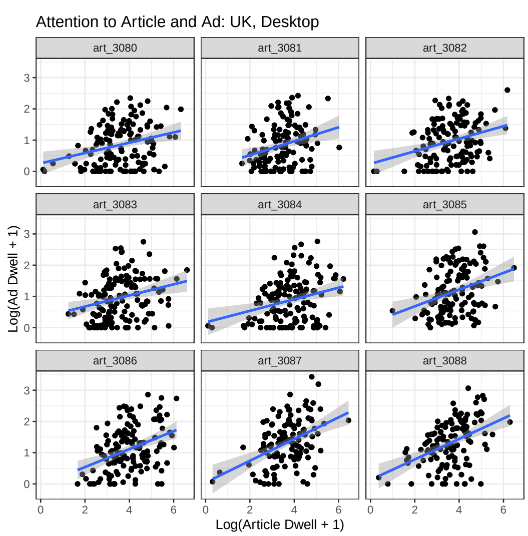
Figure 18: Correlation in Article and Ad Dwell, By Articles



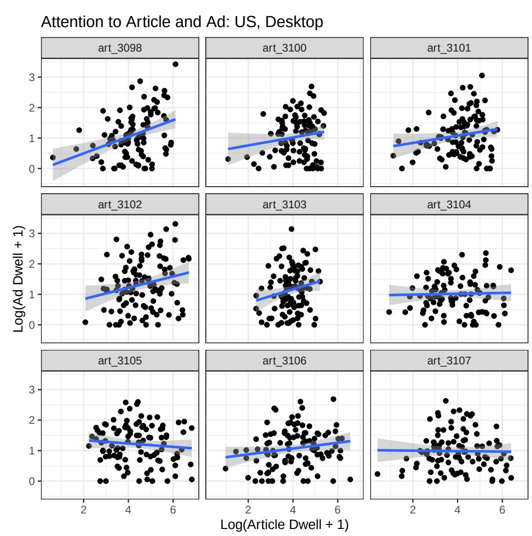
(a) UK, Mobile



(b) US, Mobile



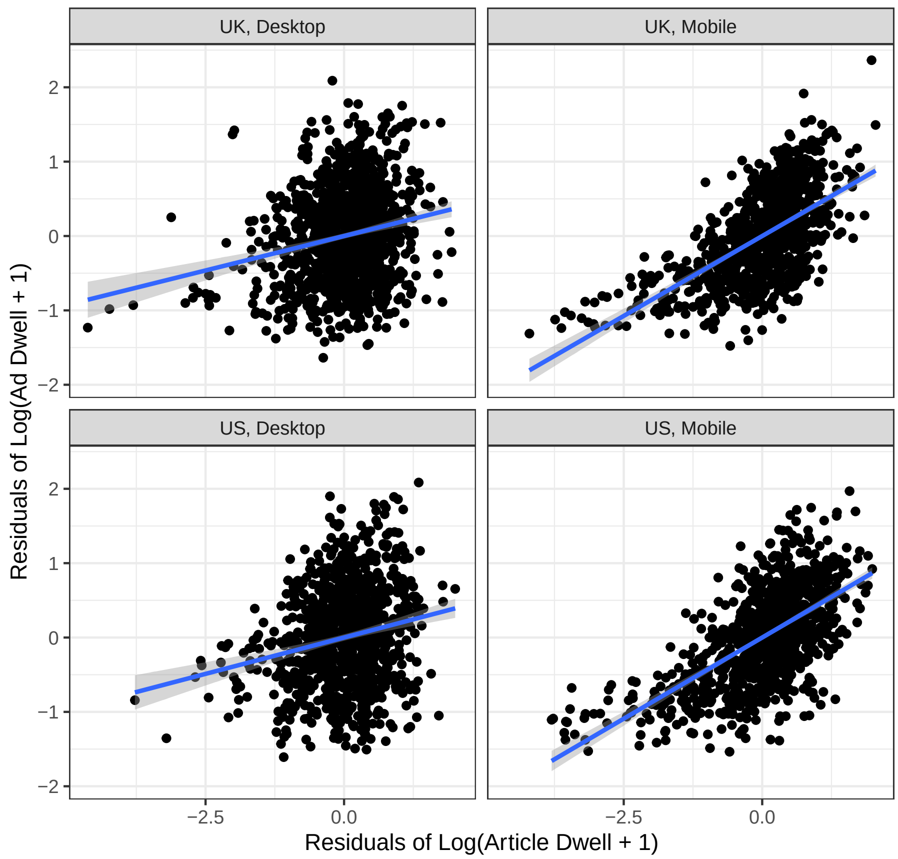
(c) UK, Desktop



(d) US, Desktop

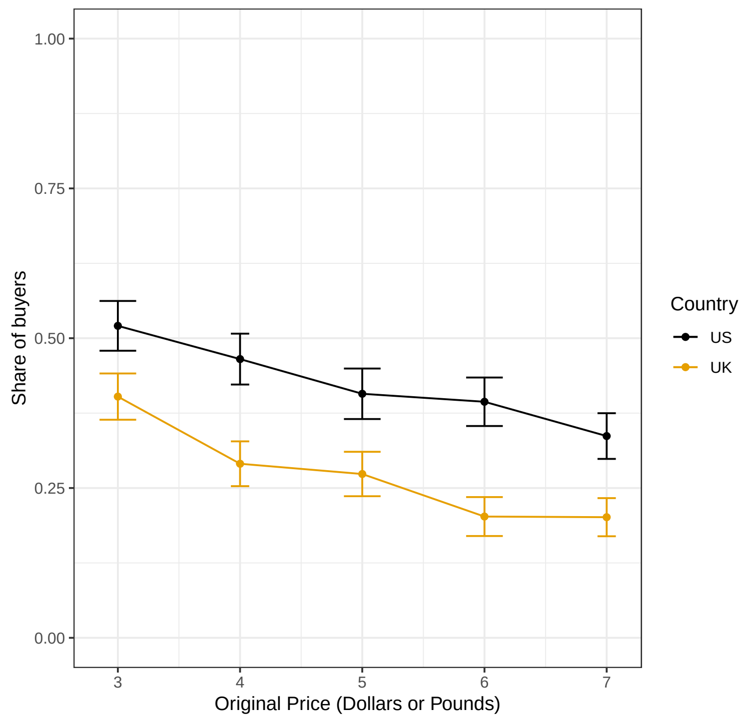
Correlation between attention to article and attention to ad, for each article shown to UK and US individuals. Each panel corresponds to a single article. Ad and article dwell times are transformed into the logarithmic scale to make the visualization easier to read. Blue line corresponds to the best linear prediction of the variable on the vertical axis by the variable on the horizontal axis.

Figure 19: Positive Correlation in the Residualized Article and Ad Dwell



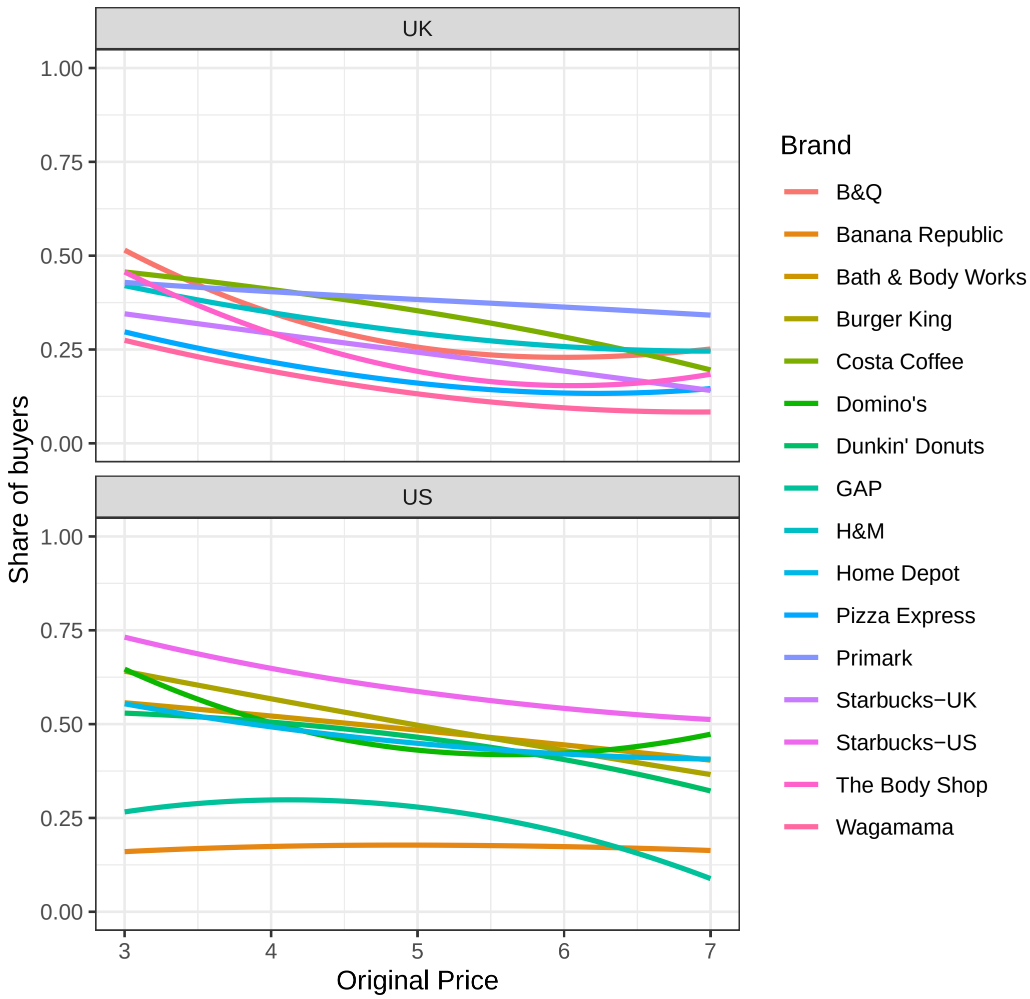
Correlation between attention to article and attention to ad, for each article shown to UK and US individuals. Each panel corresponds to a set of articles in this country on this device type. Ad and article dwell times are transformed into the logarithmic scale to make the visualization easier to read, and then residualized using the same FE as in our main empirical specification (e.g. Table 3). Blue line corresponds to the best linear prediction of the variable on the vertical axis by the variable on the horizontal axis.

Figure 20: Demand Curves by Country



Demand curves are computed in each country as an average purchase probability across brands. The currency is dollars in the US and pounds in the UK.

Figure 21: Demand curves by brands



Demand curves are computed for each brand in each country. The plots show a non-parametric regression of purchase decisions on the price shown to the individual. Demand curves are broadly downward sloping for each brand.

F Advertising Effects on Recall and Purchase: Functional Form Robustness

In this section we present a number of robustness checks with respect to the functional form of our regressions Table 17 presents a specification which allows for non-linear effects of attention. Table 18 presents the results of a specification with individuals fixed-effects. Table 19 presents the results of a logit specification.

Table 17: Effect of Attention on Recall/Purchase

	Dependent variable:			
	Recall		Purchase	
	(1)	(2)	(3)	(4)
Ad Visible	0.0063***		0.0016*	
	(0.0009)		(0.0009)	
Ad Visible sqr.	−0.00003***		−0.000002	
	(0.00001)		(0.00001)	
Ad Dwell		0.0747***		0.0104**
		(0.0069)		(0.0053)
Ad Dwell sqr.		−0.0028***		−0.0002
		(0.0005)		(0.0003)
Brand FE	Y	Y		
Price x Brand FE			Y	Y
Observations	5,707	3,925	5,707	3,925
R^2	0.0946	0.1678	0.1300	0.1361

Note: *p<0.1; **p<0.05; ***p<0.01. All specifications include a quartic polynomial in log of average time that each article was visible for by each individual. All specifications include Step Order and Device x Country FE. All specifications include FE for individual covariates (income, gender, education, age, and self-reported political leaning). Standard errors clustered at the individual level.

Table 18: Effect of Attention on Recall/Purchase

	Dependent variable:			
	Recall		Purchase	
	(1)	(2)	(3)	(4)
Ad Visible	0.0007		0.0010**	
	(0.0006)		(0.0005)	
Ad Dwell		0.0106***		0.0036
		(0.0033)		(0.0027)
Brand FE	Y	Y		
Price x Brand FE			Y	Y
Observations	5,707	3,925	5,707	3,925
R^2	0.5069	0.5092	0.4863	0.4789

Note: *p<0.1; **p<0.05; ***p<0.01. All specifications include a quartic polynomial in log of average time that each article was visible for by each individual. All specifications include Step Order and Device x Country FE. All specificatinos include Individual FE. Standard errors clustered at the individual level.

Table 19: Effect of Attention on Recall/Purchase (Logit)

	Dependent variable:			
	Recall		Purchase	
	(1)	(2)	(3)	(4)
Ad Visible	0.0148***		0.0067***	
	(0.0023)		(0.0023)	
Ad Dwell		0.1780***		0.0342***
		(0.0142)		(0.0123)
Brand FE	Y	Y		
Price x Brand FE			Y	Y
Observations	5,707	3,925	5,707	3,925

Note: *p<0.1; **p<0.05; ***p<0.01. All specifications include a quartic polynomial in log of average time that each article was visible for by each individual. All specifications include Step Order and Device x Country FE. All specifications include FE for individual covariates (income, gender, education, age, and self-reported political leaning).

G Attention Allocation Model Estimates: Mobile and Desktop Devices

In this section, we present the results of estimating the coefficients of the attention allocation model separately by device type (ie, mobile vs desktop).

Table 20: Estimates of attention spillovers and ad avoidence: Mobile Devices

	Ad Dwell					
	OLS			IV		
Panel I	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\delta}_1$	1.627***			2.989***		
	(0.195)			(0.364)		
$\hat{\gamma}$	0.023***	0.023***	0.020***	0.010***	0.008**	0.009***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)
1st Stage Incr. F-Stat				69.55	74.93	80.89
Observations	1,824	1,824	1,824	1,824	1,824	1,824
R^2	0.437	0.447	0.679	0.301	0.277	0.645
	Article Dwell - $\hat{\gamma}$ Ad Dwell					
Panel II	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\alpha}_1$	95.273***			95.325***		
	(5.460)			(5.462)		
$\hat{\beta}$	4.950	3.320	5.445	4.913	3.278	5.418
	(6.139)	(5.703)	(5.487)	(6.139)	(5.703)	(5.487)
Observations	2,055	2,055	2,055	2,055	2,055	2,055
R^2	0.041	0.095	0.607	0.041	0.095	0.607
FE:						
Step Order	Y	Y	Y	Y	Y	Y
Article	N	Y	Y	N	Y	Y
Brand	N	Y	Y	N	Y	Y
Individual	N	N	Y	N	N	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include step order fixed effects, with step order = 1 normalized to zero. Estimates in Panel I represent coefficients from a regression of Ad Dwell on Article Dwell. In the IV specification, Article Dwell is instrumented for by the average amount of attention devoted to that article by all but this individual (Leave One Out IV). Estimates in Panel II represent coefficients from a regression of Article Dwell on an indicator of whether the ad is present on the news article. We subtract $\hat{\gamma}$ Ad Dwell from Article Dwell in Panel II to control for the attention spillover from ad to article. Standard errors clustered at the individual level.

Table 21: Estimates of attention spillovers and ad avoidence: Desktop Devices

	Ad Dwell					
	OLS			IV		
Panel I	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\delta}_1$	3.425***			3.133***		
	(0.258)			(0.464)		
$\hat{\gamma}$	0.004***	0.004***	0.006***	0.007*	0.009***	0.009***
	(0.001)	(0.001)	(0.002)	(0.004)	(0.003)	(0.003)
1st Stage Incr. F-Stat				36.6	57.12	60.31
Observations	2,101	2,101	2,101	2,101	2,101	2,101
R^2	0.042	0.050	0.471	0.036	0.024	0.468
	Article Dwell - $\hat{\gamma}$ Ad Dwell					
Panel II	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\alpha}_1$	118.576***			118.566***		
	(7.436)			(7.436)		
$\hat{\beta}$	8.502*	10.311**	11.563**	8.509*	10.326**	11.571**
	(5.006)	(4.930)	(4.549)	(5.006)	(4.930)	(4.549)
Observations	2,371	2,371	2,371	2,371	2,371	2,371
R^2	0.030	0.112	0.657	0.030	0.112	0.657
FE:						
Step Order	Y	Y	Y	Y	Y	Y
Article	N	Y	Y	N	Y	Y
Brand	N	Y	Y	N	Y	Y
Individual	N	N	Y	N	N	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include step order fixed effects, with step order = 1 normalized to zero. Estimates in Panel I represent coefficients from a regression of Ad Dwell on Article Dwell. In the IV specification, Article Dwell is instrumented for by the average amount of attention devoted to that article by all but this individual (Leave One Out IV). Estimates in Panel II represent coefficients from a regression of Article Dwell on an indicator of whether the ad is present on the news article. We subtract $\hat{\gamma}$ Ad Dwell from Article Dwell in Panel II to control for the attention spillover from ad to article. Standard errors clustered at the individual level.

H Advertising Effects on Recall and Purchase: L1O IV

In this section, we present our estimates of the effects of attention on purchase and recall, when we instrument attention (*Ad Visible* and *Ad Dwell*) using the L1O measure of attention (the average attention devoted to each article by all other individuals in our sample).

Table 22: Estimates of advertising effects on recall and purchase: L1O IV

	Recall ($\hat{\rho}$)					Purchase ($\hat{\lambda}$)				
	All	Device		News Type		All	Device		News Type	
		Mobile	Desktop	Hard	Soft		Mobile	Desktop	Hard	Soft
Panel I	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ad Visible	0.001	0.003	0.003	−0.022	0.004	0.002	0.009	−0.003	0.016	−0.0002
	(0.004)	(0.007)	(0.006)	(0.031)	(0.004)	(0.005)	(0.007)	(0.006)	(0.037)	(0.004)
Observations	3,925	2,165	1,760	1,824	2,101	3,925	2,165	1,760	1,824	2,101
R^2	0.136	0.167	0.156	−0.131	0.146	0.109	0.102	0.087	0.009	0.141
						First Stage				
L1O Article Dwell	0.068***	0.072***	0.074***	0.019	0.099***	0.069***	0.074***	0.070***	0.015	0.104***
	(0.010)	(0.016)	(0.011)	(0.013)	(0.013)	(0.010)	(0.016)	(0.011)	(0.013)	(0.014)
Observations	3,925	2,165	1,760	1,824	2,101	3,925	2,165	1,760	1,824	2,101
R^2	0.414	0.407	0.466	0.292	0.509	0.405	0.389	0.450	0.268	0.488
1st Stage Incr. F-Stat	49.1	19.47	44.25	2.21	57.24	48.94	20.67	42.04	1.24	57.28
Panel II	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ad Dwell	0.009	0.023	0.037	−0.050	0.046	0.016	0.068	−0.052	0.030	−0.002
	(0.037)	(0.058)	(0.078)	(0.061)	(0.057)	(0.040)	(0.050)	(0.109)	(0.065)	(0.049)
Observations	3,925	2,165	1,760	1,824	2,101	3,925	2,165	1,760	1,824	2,101
R^2	0.136	0.162	0.116	0.091	0.098	0.132	0.147	−0.125	0.132	0.136
						First Stage				
L1O Article Dwell	0.008***	0.009***	0.006*	0.009**	0.008***	0.008***	0.010***	0.004	0.008**	0.008***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	3,925	2,165	1,760	1,824	2,101	3,925	2,165	1,760	1,824	2,101
R^2	0.158	0.179	0.203	0.280	0.146	0.143	0.150	0.171	0.252	0.117
1st Stage Incr. F-Stat	12.95	10.34	2.78	5.52	6.77	12.73	13.27	1.78	5.31	7.83
FE:										
Brand	Y	Y	Y	Y	Y					
Price x Brand						Y	Y	Y	Y	Y
Step Order	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Country x Device	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Dem. Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include a quartic polynomial in log of average time that an average article was visible for by each individual. All specifications include step order and device x country fixed effects. All specifications include fixed effects for individual covariates (income, gender, education, age, and self-reported political leaning). Regressions with recall as an outcome include brand fixed effects, and regressions with purchase as an outcome include brand by price fixed effects. Standard errors clustered at the individual level.

I Ideological Mismatch as News Attention Shifter

In this section we examine whether the mismatch between individuals and newspapers (in terms of their political ideology) affects the extent to which advertising is effective. Above, we have established that incremental attention to news content increases readers' exposure to display ads, and this, in turn, increases ad effectiveness. Past work suggests that alignment of readers' beliefs about politics could facilitate such incremental attention: news readers prefer articles with an ideological slant more aligned with their own views [Schmuck et al., 2019].

To measure political alignment of news readers and articles, we rely on the political orientation of news outlets. Recall that individuals are able to see the newspaper from which each story originates, since a billboard is displayed at the top of each article that clearly shows the news source. Also, at the end of the experiment, respondents were asked about their political views.[36]

The outlets we chose, and their political slants, are well known in each country. The newspapers we chose have a wide online readership, but are also quite politically oriented. In the UK, *The Guardian* has a political alignment on the left, and the *Daily Mail* is on the right. In the US, *The New York Times* is left leaning, while *USA Today* is centrist. An independent survey on AMT confirms these choices of each newspaper's political slant (see Appendix B).

Does an individual with self-reported conservative views react differently to the news published by a newspaper that leans politically to the left? We first build an index of "right-wingness" for each newspaper and individual. Regarding newspapers, the *Daily Mail* is assigned +1, *USA Today* is assigned 0, and *The New York Times* and *The Guardian* are assigned -1.[37]

Similarly, individuals who described themselves as Conservative, Moderate, and Liberal are assigned +1, 0 and -1 respectively. We then compute, for each observation, the "political mismatch" between each individual and newspaper article shown as absolute value of the difference between these two variables. There is no mismatch (mismatch = 0) between a person who places her/himself to the right of the political spectrum when reading the *Daily Mail* (or a left-wing person reading *The Guardian*), while a large mismatch occurs (mismatch = 2) when that person is presented with an article from an outlet at the opposite end of the political spectrum.

[36] This was done in the end of the experiment so that this question would not bias the behavior of participants.

[37] This classification is also broadly confirmed by sites that regularly conduct media bias ratings; e.g., https://www.allsides.com/media-bias/media-bias-ratings.

Intermediate cases (mismatch = 1) arise from other combinations.

We first examine whether a politically mismatched individual-article pair is associated with the individual paying less attention to the article and display advertising on this article's page. Table 23 presents the results. A higher political mismatch is associated with lower attention to the article: going from fully aligned views (mismatch = 0) to completely misaligned views (mismatch = 2) decreases the time people read the article by around 18 seconds (Columns 1 and 2). This, in turn, decreases the attention people pay to ads on the page, with ads becoming visible for $0.82 \cdot 2 = 1.64$ seconds less (Column 3) and attracting $0.12 \cdot 2 = 0.24$ seconds less active attention time (Column 4).[38]

Table 23: Attention and Political Mismatch

	Dependent variable:			
	Article Visible	Article Dwell	Ad Visible	Ad Dwell
	(1)	(2)	(3)	(4)
Political Mismatch (0/1/2)	-8.8681^{**}	-9.2912^{***}	-0.8216^{***}	-0.1190
	(3.5830)	(2.5026)	(0.2846)	(0.0810)
Observations	5,360	3,652	5,360	3,652
R^2	0.6587	0.6579	0.6515	0.5285

Note: *p<0.1; **p<0.05; ***p<0.01. Fixed Effects: Article, Brand, Step Order. Includes Individual FE. Includes quartic polynomial in total time page is visible for each individual. Includes Price x Brand FE. Excludes observations for which no brand was shown. Standard errors clustered at the individual level.

We now confirm that this incremental attention to advertising – generated by an article's and a reader's ideological match – converts into higher advertising effectiveness. For this, we re-estimate our IV specifications (Equations 6 and 7) using the political mismatch as an instrument for ad attention. Compared to our main model, we use a more stringent specification with individual FE, to increase the estimates' precision.[39]

Table 24 presents the estimates of the effects of ad attention on purchase probabilities. As before, Columns 1 and 2 present the first stage and IV estimates using ad visible as a measure of ad attention. While the power of the first stage is lower compared to our other IV (F-statistic of 8.44), the IV estimate still produces a significant positive estimate, lining up with the rest of our conclusions on the role of incremental attention to display ads.

Columns 3 and 4 of Table 24 present the first stage and IV estimates using ad dwell as a mea-

[38]Table 23 has fewer observations than some of the previous tables. This is because we allowed individuals to opt out of reporting their political orientation. In these cases (about 6% of the data), the political mismatch could not be

Table 24: 2SLS Regression of Purchase on Attention, IV: Pol Mismatch

	Dependent variable:			
	Ad Visible (1st stage)	Purchase (2SLS)	Ad Dwell (1st stage)	Purchase (2SLS)
	(1)	(2)	(3)	(4)
Pol Mismatch	-0.6178^{**}		0.0105	
	(0.2716)		(0.0838)	
Ad Visible		0.0400^{*}		
		(0.0229)		
Ad Dwell				-2.6920
				(21.7459)
1st Stage Incr. F-Stat	5.18		0.02	
Observations	5,360	5,360	3,652	3,652
R^2	0.6290	-0.2910	0.5030	-161.9133

Note: *p<0.1; **p<0.05; ***p<0.01. All specifications include a quartic polynomial in log of average time that each article was visible for by each individual. All specifications include Country x Device and Step Order FE. All specificatinos include Individual FE. Standard errors clustered at the individual level. All specifications include Price x Brand FE.

sure of ad attention. In this specification, the first stage estimates are too imprecise to generate conclusive IV estimates.

We present the results for ad recall in Table 25. As before, these results are inconclusive on the effect of incremental ad attention on recall.

computed. The same holds for Table 24.

[39]We get imprecise estimates in the first stage without individual FE.

Table 25: 2SLS Regression of Recall on Attention, IV: Pol Mismatch

	Dependent variable:			
	Ad Visible (1st stage)	Recall (2SLS)	Ad Dwell (1st stage)	Recall (2SLS)
	(1)	(2)	(3)	(4)
Pol Mismatch	-0.5807^{**}		-0.0114	
	(0.2762)		(0.0847)	
Ad Visible		-0.0194		
		(0.0187)		
Ad Dwell				-0.2519
				(2.2441)
1st Stage Incr. F-Stat	4.42		0.02	
Observations	5,360	5,360	3,652	3,652
R^2	0.6236	0.3133	0.4914	-0.9061

Note: *p<0.1; **p<0.05; ***p<0.01. All specifications include a quartic polynomial in log of average time that each article was visible for by each individual. All specifications include Country x Device and Step Order FE. All specificatinos include Individual FE. Standard errors clustered at the individual level. All specifications include Brand FE.

J Robustness of Eye-Tracking Metrics by Device

In Section 8 we analyzed the robustness of our findings with respect to the quality of the eye-tracking data. We repeat that analysis here, broken down by device type. Tables 26-27 present the results for OLS, while Tables 28-29 present the results for IV.

Table 26: Estimates of advertising effects on recall and purchase: OLS, Robustness in Ad Dwell Measurements, Mobile

	Recall ($\hat{\rho}$)							
	Main	Re-weighted observations by					Adjusted Ad Dwell	
		1/precision	1/accuracy	1/gaze duration	Hit Rate	1/time since calibr.	Interior	Away from Text
Panel I	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ad Dwell	0.028***	0.007	0.032***	0.026***	0.030***	0.027***	0.026***	0.027***
	(0.006)	(0.010)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)
Observations	1,824	1,785	1,785	1,824	1,824	1,824	1,824	1,824
R^2	0.133	0.239	0.144	0.140	0.144	0.163	0.141	0.143
	Purchase ($\hat{\lambda}$)							
Panel II	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ad Dwell	0.009**	0.005	0.011***	0.009**	0.012**	0.006	0.010***	0.008**
	(0.004)	(0.006)	(0.004)	(0.004)	(0.005)	(0.006)	(0.003)	(0.004)
Observations	1,824	1,785	1,785	1,824	1,824	1,824	1,824	1,824
R^2	0.200	0.383	0.241	0.201	0.222	0.252	0.202	0.200
FE:								
Price x Brand	Y	Y	Y	Y	Y	Y	Y	
Step Order	Y	Y	Y	Y	Y	Y	Y	Y
Country x Device	Y	Y	Y	Y	Y	Y	Y	Y
Dem. Controls	Y	Y	Y	Y	Y	Y	Y	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include a quartic polynomial in log of average time that an average article was visible for by each individual, step order and device x country fixed effects, fixed effects for individual covariates (income, gender, education, age, and self-reported political leaning), and brand by price fixed effects. A few observations have missing values of precision and accuracy variables; we drop those observations from the analysis in specifications with 1/precision and 1/accuracy weights. Standard errors clustered at the individual level.

Table 27: Estimates of advertising effects on recall and purchase: OLS, Robustness in Ad Dwell Measurements, Desktop

	Recall ($\hat{\rho}$)							
	Main	Re-weighted observations by					Adjusted Ad Dwell	
		1/precision	1/accuracy	1/gaze duration	Hit Rate	1/time since calibr.	Interior	Away from Text
Panel I	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ad Dwell	0.036***	0.037***	0.036***	0.036***	0.033***	0.026***	0.024***	0.023***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)	(0.003)	(0.003)
Observations	2,101	2,090	2,090	2,101	2,101	2,101	2,101	2,101
R^2	0.188	0.186	0.205	0.192	0.188	0.209	0.179	0.177
	Purchase ($\hat{\lambda}$)							
Panel II	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ad Dwell	0.008**	0.012***	0.008*	0.007*	0.009*	0.009	0.003	0.005
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Observations	2,101	2,090	2,090	2,101	2,101	2,101	2,101	2,101
R^2	0.153	0.211	0.190	0.155	0.197	0.218	0.151	0.152
FE:								
Price x Brand	Y	Y	Y	Y	Y	Y	Y	
Step Order	Y	Y	Y	Y	Y	Y	Y	Y
Country x Device	Y	Y	Y	Y	Y	Y	Y	Y
Dem. Controls	Y	Y	Y	Y	Y	Y	Y	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include a quartic polynomial in log of average time that an average article was visible for by each individual, step order and device x country fixed effects, fixed effects for individual covariates (income, gender, education, age, and self-reported political leaning), and brand by price fixed effects. A few observations have missing values of precision and accuracy variables; we drop those observations from the analysis in specifications with 1/precision and 1/accuracy weights. Standard errors clustered at the individual level.

Table 28: Estimates of advertising effects on recall and purchase: Article Dwell IV, Robustness in Ad Dwell Measurements, Mobile

	Purchase ($\hat{\lambda}$)							
	Main	Re-weighted observations by					Adjusted Ad Dwell	
		1/precision	1/accuracy	1/gaze duration	Hit Rate	1/time since calibr.	Interior	Away from Text
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ad Dwell	0.015**	0.009	0.015*	0.014**	0.013	0.006	0.018**	0.019**
	(0.007)	(0.010)	(0.009)	(0.007)	(0.008)	(0.012)	(0.008)	(0.009)
Observations	1,824	1,785	1,785	1,824	1,824	1,824	1,824	1,824
R^2	0.199	0.382	0.241	0.201	0.221	0.252	0.198	0.195
	First Stage							
Article Dwell	0.023***	0.026***	0.020***	0.023***	0.020***	0.019***	0.018***	0.018***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)
Observations	1,824	1,785	1,785	1,824	1,824	1,824	1,824	1,824
R^2	0.500	0.843	0.474	0.503	0.463	0.497	0.269	0.291
1st Stage Incr. F-Stat	183.4	90.43	155.16	169.06	233.08	158.62	114.08	48.22
FE:								
Price x Brand	Y	Y	Y	Y	Y	Y	Y	Y
Step Order	Y	Y	Y	Y	Y	Y	Y	Y
Country x Device	Y	Y	Y	Y	Y	Y	Y	Y
Dem. Controls	Y	Y	Y	Y	Y	Y	Y	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include a quartic polynomial in log of average time that an average article was visible for by each individual, step order and device x country fixed effects, fixed effects for individual covariates (income, gender, education, age, and self-reported political leaning), and brand by price fixed effects. A few observations have missing values of precision and accuracy variables; we drop those observations from the analysis in specifications with 1/precision and 1/accuracy weights. Standard errors clustered at the individual level.

Table 29: Estimates of advertising effects on recall and purchase: Article Dwell IV, Robustness in Ad Dwell Measurements, Desktop

	Purchase ($\hat{\lambda}$)							
	Main	Re-weighted observations by					Adjusted Ad Dwell	
		1/precision	1/accuracy	1/gaze duration	Hit Rate	1/time since calibr.	Interior	Away from Text
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ad Dwell	0.052	0.052	0.036	0.054	0.020	0.044	0.038	0.071
	(0.037)	(0.036)	(0.033)	(0.040)	(0.031)	(0.030)	(0.027)	(0.059)
Observations	2,101	2,090	2,090	2,101	2,101	2,101	2,101	2,101
R^2	0.079	0.154	0.160	0.072	0.193	0.168	0.065	−0.157
	First Stage							
Article Dwell	0.005***	0.005***	0.005***	0.005***	0.006***	0.007***	0.007***	0.004**
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Observations	2,101	2,090	2,090	2,101	2,101	2,101	2,101	2,101
R^2	0.156	0.178	0.186	0.158	0.194	0.246	0.128	0.111
1st Stage Incr. F-Stat	10.41	16.05	16	9.87	19.08	15.29	4.55	12
FE:								
Price x Brand	Y	Y	Y	Y	Y	Y	Y	Y
Step Order	Y	Y	Y	Y	Y	Y	Y	Y
Country x Device	Y	Y	Y	Y	Y	Y	Y	Y
Dem. Controls	Y	Y	Y	Y	Y	Y	Y	Y

*p<0.1; **p<0.05; ***p<0.01

All specifications include a quartic polynomial in log of average time that an average article was visible for by each individual, step order and device x country fixed effects, fixed effects for individual covariates (income, gender, education, age, and self-reported political leaning), and brand by price fixed effects. A few observations have missing values of precision and accuracy variables; we drop those observations from the analysis in specifications with 1/precision and 1/accuracy weights. Standard errors clustered at the individual level.