Attention, recall and purchase: Experimental evidence on online news and advertising*

Tommaso Valletti† and Andre Veiga‡

March 26, 2021

Abstract

We conduct an experiment where subjects read online news articles and are shown ads for brands next to those articles. Using eye-tracking technology, we measure the attention that each individual devotes to each article and ad. Then, respondents choose between cash or vouchers for the brands advertised. Attention to ads is a predictor both of willingness-to-pay for brands, and brand recall. The main predictors of attention include the type of news and the match between individual political preferences and the media outlet.

JEL Classification Codes: M37, C91, L86

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

*We thank Daniel Ershov, Stephan Seiler and Gokhan Yildirim for very useful comments, as well as participants to seminars at the Bank of Italy, Imperial College London and the CESifo Conference on the Economics of Digitization. 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 (www.NETinst.org).

†Imperial College London, CEPR and CESifo, t.valletti@imperial.ac.uk

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

Contents

1 Introduction

Online advertising has been extremely successful over the past two decades. Around 2017, online advertising overtook television as the medium with the highest global ad spending. Yet, the economic quantification of its impact and the mechanisms at work remain elusive. Intuitively, advertising should work when it captures the 'attention' of the viewers, but a reliable metric to quantify this concept has not been readily available – until recently.

In this paper we use eye-tracking technology and run an online experiment to assess the effectiveness of advertising when people read online news. We find that, once an ad of a particular brand receives the attention of a viewer, this increases the probability of purchasing a voucher that can only be used to purchase from the advertised brand. The effect is positive, and both statistically and economically significant.

We also find that attention is positively associated to brand recall, which is a measure of the effectiveness of advertising widely used in the industry and in the marketing literature, but less so in the economics literature. Hence, we establish a link between an economic concept (purchase) and a marketing concept (brand recall). The latter is easier (and cheaper) to measure than the former, but would not be very meaningful without having established this link, as we do in this article. Our results also allow us to compare advertising effectiveness and its costs, possibly across different media.

In the second part of the paper, we provide some evidence regarding the predictors of attention to ads. First, we consider the effect of "hard news." This line of inquiry was partly motivated by the observation that, during the COVID-19 pandemic, online newspapers saw an increase in their viewership, but experienced a drop in advertising revenues. According to much of the press, this apparent paradox was linked to the fact that ads are served by intermediaries that often block ads from being seen next to certain words associated to "hard news", i.e., news that advertisers perceive as upsetting or could otherwise create a negative image associated with the brand. The rationale behind this blocking is the fear that negative association could hurt the brand's image or the brand's advertising will be less effective, thereby dissuading readers from purchasing from the advertised brand. In practice, it seems that articles reporting on the COVID-19 pandemic or the Black Lives Matter (BLM) movement are often considered to be "hard news" and blocked by intermediaries. We take this industry standard to also be our definition of "hard news". Our experiment included several "hard news" articles related to the topics of the COVID-19 pandemic and the BLM protests during the summer of 2020. We validate our categorizing of an article as "hard news" via an independent survey on Mechanical Turk.

We find that the type of news does directly affects the degree of attention. However, conditional on the amount of attention devoted to the ad, whether an article is hard news or not has no additional effect on recall or purchase. In other words, any effect that hard news articles have on advertising effectiveness seems to be entirely mediated by the amount of attention devoted to the ad, not the content of the article. We discuss the implications of these results for managers and advertisers and discuss why these results should be interpreted with caution.

We also explore how the reader's and newspaper's political affiliation predict attention. We show that attention seems to be driven by the match between individual characteristics (including political preferences) and characteristics of the article. A reader with a certain political leaning is more likely to devote attention to an article (and to the ad shown next to it) from an outlet with a similar political leaning.

The paper is organized as follows. Section 2 reviews the literature, while Section 3 illustrates the experiment and Section 4 describes the resulting dataset we employ. Section 5 presents the main results of the impact of attention and brand recall and purchase. Robustness checks are discussed in Section 6, while Section 7 explores the main drivers of attention. Finally, Section 8 concludes.

2 Literature

The returns to advertising have been studied in the economics and marketing literature for decades, but this literature has faced some struggles. Traditional, offline, advertising data have often been insufficient to measure the impact of advertising on consumer purchasing behavior, because of endogeneity and identification problems (see Bagwell (2007)). Online advertising has been not only extremely successful, but it has also offered a new and large data collection opportunity, allowing researchers to revamp their efforts. Metrics such as click-through rates have been used to match consumers with context and make advertising spend more efficient. For instance, surveys about purchasing intent have been used to study ad intrusiveness, as in Goldfarb and Tucker (2011). See also Lewis and Reiley (2014) for a more general survey.

One limitation of some of the existing research is that it often measures *intent to purchase*, rather than *actual purchases*.[1] Ultimately, what matters for advertisers is uncovering the causal link between dollar sales and dollars spent on an ad. Imai, Kang and Camerer (2019) highlight the bias introduced when hypothetical decisions, such as purchase intent, are considered. Schmidt and Bijmolt (2019) and Ding, Grewal and Liechty (2005) emphasize the importance of incentivized choices when estimating demand. One purpose of our paper is to use a new measure of exposure to ads, namely attention as measured via an eye-tracking technology, and link this measure of attention to purchase decisions.

To be sure, large datasets have become available to link advertising exposure to actual purchases. But these come with their challenges too, as illustrated by Johnson (2020). Lewis and Reiley (2014), Lewis and Rao (2015) and Gordon et al. (2019) argue that measuring the effectiveness of advertising with observational data is difficult due to endogeneity and to the fact that large-scale experiments (required since the effects of online ads are typically small) lack the statistical power to *reject* even a null hypoth-

[1] Goldfarb and Tucker (2011) and Neumann, Tucker and Whitfield (2019) study online ad exposure and ad targeting, and measure its effects on stated purchase intention.

esis of no ad effect. We address these measurement issues through an experiment that can improve on observational methods. Despite the smaller sample size, we recover estimates on the impact of ads on purchase that are both statistically and economically significant, and find that online ads can have big effects, but only if people pay attention to them.[2]

A significant literature (for instance, Chandon et al. (2009)) measures the relationship between attention and purchase. These papers tend to use proxies for attention such as in-store location of products, do not use eye-tracking, and do not focus on online ads as we do.

This article contributes to a growing literature using eye-tracking to measure attention (for instance, Brocas et al. (2014), Camerer et al. (1993), Knoepfle, Wang and Camerer (2009) and Reutskaja et al. (2011) use eye-tracking in settings very different from ours). Several articles examine the role of attention in advertising, but this is often measured as exposure to ads, or ad visibility. For instance, Ghose and Todri (2016) study a quasi-experimental setting and measure the impact of advertising on consumer search for the product, and on purchase, but do not have access to eye-tracking and thus cannot measure attention to the ad, only its potential viewability. Balcombe, Fraser and McSorley (2015) show that stated attention is found to diverge substantively from visual attention to attributes, thus emphasizing the need to carefully measure attention. As in those articles, via the eye-tracking technology, we can measure precisely the amount of time the eye retina actually dwells on a particular ad.[3]

One important finding of our work is to establish a link between brand recall and actual purchases (as opposed to purchase intention), in online environments. A large literature, as well as most analyses by industry practitioners, focus on the concepts of

[2]This article is also related to a relatively smaller literature in economics (e.g., Bertrand et al. (2010)) that uses field experiments to determine the effects of advertising on customer take-up and selection. However, this literature does not consider online environments and does not measure attention in the way we do.

[3]Several articles use eye-tracking to understand consumer purchases off-line. For instance, Takahashi, Todo and Funaki (2018) uses eye-tracking to study the effects of labels on purchases of coffee. Martinovici and Erdem (2021) use eye-tracking to explain the value of different smartphone features. In contrast to these articles, our focus is on attention and online advertising.

brand recognition and recall, as a proxy for attention (see, e.g., Macdonald and Sharp (2000)). Khurram, Qadeer and Sheeraz (2018) examine the role of brand recall and brand recognition in predicting purchases, using a survey. In this paper, we consider brand recall but are particularly interested in actual consumer purchases. We find that attention to the ad drives both brand recall and purchases. To our knowledge, this link has not been previously established and it is important since brand recall is easier and cheaper to measure than purchases. Without having first established the link between recall and purchase, however, one would be left to wonder if brand recall is meaningful for purchase conversion.

Our last contribution is towards the understanding of the drivers of attention when reading online news. While context has always been one of the most important factors for advertising in general, and online targeted advertising in particular (see, for instance, the distinction between obtrusive and unobtrusive ads proposed by Goldfarb and Tucker (2011)), our attention metric allows us to make additional progress. We can relate the amount of attention to the types of news that are shown to participants to the experiment (and, in particular, we focus on the distinction between *soft* and *hard* news), as well as to the match between a media outlet and individual preferences (measured along the political dimension in our setting).[4]

3 Experimental Setting

The experimental design involved a stratified sample of 1,000 people, split into two cells of 500 people each in the United Kingdom (UK) and the United States (US). Each cell was further divided equally according to the device used (desktop or smartphone). The respondents were recruited to match the UK/US online population in terms of age, gender, income, and location. They were recruited via Panelbase, a specialist supplier of

[4] Yan, Miller and Skiera (2020) show evidence that the presence of ads is associated with lower quantity and variety of news consumption, but they rely on an endogenous choice by individuals of whether to use an ad-blocker. By contrast, we use an experimental setting to quantify the (positive) effect of attention on purchase, and the (negative) effect of ads on attention devoted to articles.

research and marketing panels.

Each respondent was first asked to self-report several socio-demographic characteristics (in particular: age, education, income, gender and postcode).[5]

Then participants were invited to read articles from two online newspapers (The Guardian and Daily Mail in the UK, the New York Times and USA Today in the US). In each country, we chose outlets that had a wide readership online but are widely perceived as differing in their political leaning. In each country, articles were split evenly between two outlets (in order to assess the impact of the type of media outlet).

We chose articles split evenly between soft and hard news (as mentioned in the Introduction). To select the latter, we followed the advice of industry experts and focused on articles about the COVID-19 pandemic and the BLM protests of summer 2020. We additionally validate these decisions using an independent survey on Mechanical Turk, which we present in Appendix C.

We chose ads from well-known and widely available product brands. Ads were inserted into the article pages as they would be normally. Each article contained ads for a single brand.

Every individual was exposed to all 9 different articles and all 8 different brands (each article and brand was shown only once). One of the articles, at random, was shown without an ad (to assess the baseline level of interest in each article). The order in which the ads were shown, and the matching between articles and brands, was randomized.

For each individual, measures of attention to each article and ad were recorded. In particular, the experiment recorded the amount of time the article and the ad were visible on screen (this measure does not use eye-tracking). The experiment also recorded, via eye-tracking, the amount of time each individual was actively looking at each article and ad, as we discuss further below.

After reading the articles, individuals were first asked if they could remember the brands whose ads had been shown to them. Individuals were presented with a list of

[5]Respondents were also asked to report their political orientations, but only at the very end of the experiment in order not to frame their attention on this issue.

the 8 brands shown, in addition to 8 "decoy" brands, and were asked to identify which brands they had seen. The decoy brands were chosen to be well known in each country.

Then, individuals were asked to make purchase decisions. Individuals were offered to choose between an e-voucher worth £10 (UK) or $10 (US) specific to a certain brand and some (randomly selected) lower amounts of cash (in the range £3-7 in the UK, and $3-7 in the US). Individuals were asked to make one choice for each brand they had seen, and they were informed they would be sent electronically one outcome of their choice, which was then administered again via Panelbase at the end of the experiment.[6]

The experiment was then set to assess the probability of remembering correctly branded ads shown as well as the probability of choosing a voucher for a certain product (instead of cash), and how this would vary with respect to the intensity of attention paid to a certain ad while reading the articles. The experiment was designed to measure the impact of attention at the intensive margin (as most articles do show ads). Notice, importantly, that consumers made *actual* product choices in our experiment, rather than merely stating their preferences in a survey.

The participants were anonymous to the research team, since all contact was mediated through the recruiting firm (Panelbase). The study received ethical approval of the protocol prior to the start of the experiment. The experiment was run at the end of July 2020.

Appendix A reports more details about the experiment, including the specific articles and brands.

[6]A common approach is to elicit willingness to pay (WTP) through a second-price auction (known as the Becker-DeGroot-Marschak (BDM) mechanism). In the interest of simplicity, and because the experiment took place online (so the researchers were not available to provide clarifications) the approach used in this paper was to estimate WTP by presenting individuals with a take-it-or-leave-it (TIOLI) offer. Berry, Fischer and Guiteras (2020) show that TIOLI has a good performance in practice and is simpler to implement than BDM.

4 Data

We first describe the eye-tracking technology supplied by Lumen Research, a specialist advertising research agency based in London. After receiving the consent of the viewer, this technology employs software that uses the camera of a desktop/mobile phone and measures where on the screen the eye retina is focused on. The experiment measured how long each part of the screen (articles and ads) were *viewable*, which did not require eye-tracking. The experiment also measured how long each individual's sight dwelled on each article and ad, which does require eye-tracking (this variable is called *dwell* in our analysis).[7] These are our two measures of *attention*.

A heat map is provided in Figure 1 as an example of how these metrics are constructed. The figure shows an article, as well as the ad for one brand (one banner is shown at the top, and two are on the side, as a viewer scrolls down the article). The map highlights the pixels on the screen that were actually viewed by the reader. In Figure 2 we illustrate these heat maps for ads of two different brands.

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
Ad Visible (s)	6,431	18.994	17.231	0.000	291.905
Page Visible (s)	6,431	143.301	169.341	20.130	1,894.635
Ad Dwell (s)	4,426	2.700	3.113	0.000	40.214
Page Dwell (s)	4,426	77.256	98.935	0.165	966.945
Price (GBP/USD)	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

Table 1 presents the summary statistics. As mentioned earlier, each respondent was asked to read 9 articles and to make choices involving 8 brands. Observations were split

[7] Eye-tracking, according to Lumen, allows one to see "what people actually do, not what they say they do." More details are in Appendix A.

Figure 1: Heat map of a page (article and ads)

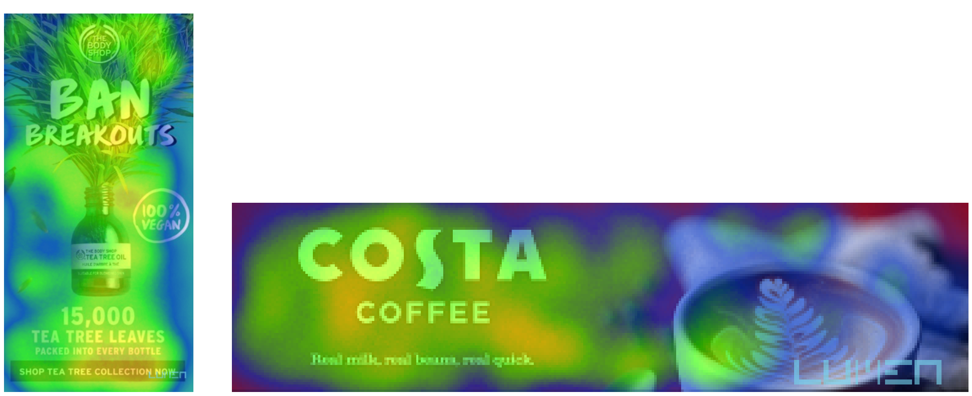

Figure 2: Heat map of ads for two branded products

evenly between US/UK, and between desktop/smartphone.

The attention variables are defined as follows: *Ad Visible* reports the number of seconds a certain ad is 'viewable', according to Media Rating Council (MRC) standards.[8] An ad is counted as 'viewable' if 50% of the pixels of an ad are on the screen for more than 1 second. The sample mean is 19 seconds. *Ad Dwell* reports the time an ad was actually looked at. The sample mean is just short of 3 seconds. Similarly, *Page Visible* reports the number of seconds an article was viewable (the mean is 2 minutes and 23 seconds) while *Page Dwell* is the time an article was actually read (the mean is 1 minute and 17 seconds).

The number of actual valid observations involving dwell are lower than the number of observations related to visibility. This is because the eye-tracking technology is used only for the former, and relies of the respondent not moving too much in front of the camera. If the respondent moves too much (while s/he could still be reading a page), that observation would not be reported for dwell (while visibility would still be reported). The threshold used to determine whether an individual was kept in the sample was an internal measure commonly used by the firm providing the eye-tracking technology. The individuals for whom high-quality eye-tracking data is and is not available are balanced on observables.

In terms of product choices, people were offered the choice between vouchers for products, or random amounts of money. The money amount was chosen uniformly at random between \$3 and \$10 in the US (£3 and £7 in the UK). 35% of the respondents chose an e-voucher worth \$10 (£10) for an actual product, while the rest opted for lower amounts of cash.

Finally, we consider another measure of outcomes, commonly used in the marketing literature: brand recall. In our sample, 48% of the respondents recalled a brand correctly after they had seen an ad for that brand.

Figure 3 shows how the percentage of individuals who purchased a brand's voucher

[8] http://mediaratingcouncil.org/Standards.htm

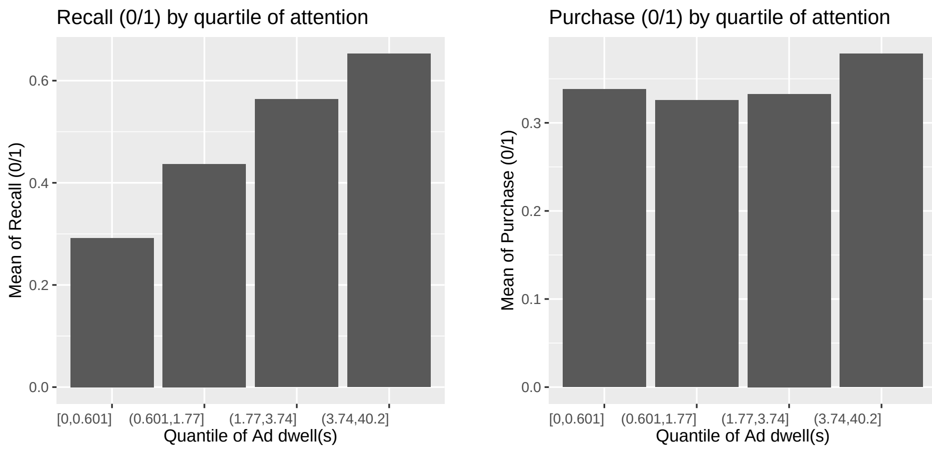

Figure 3: Recall and purchase by quartile of attention

and recall seeing that brand, increases with the quartile of attention devoted to the ad. The pattern is strongly increasing for recall. The pattern for purchase is still increasing, but the effect is smaller.

5 How Attention Affects Recall and Purchase

Each individual i reads article j which show ads for brand k. After reading all articles, the individual is first asked if she recalls the brands shown. Hence, we estimate the following linear probability model

$$r_{ijk} = \beta att_{ijk} + \gamma X_i + o_{ik} + \delta_j + \delta_k + \varepsilon_{ijk} \tag{1}$$

where $r_{ijk} = \{0, 1\}$ is an indicator describing whether or not there is a correct brand recall. att_{ijk} is a measure of attention that individual i devoted to the ad for brand k, on the page of article j where it was shown. As measures of attention we use *Ad Visibility* and *Ad Dwell*.

X_i is a vector of controls for the individual that include country, device used, and various socio-demographic characteristics. $o_{ik} = \{1, 2, ..., 9\}$ are fixed effects for the "Step Order" in which ad k was shown to individual i (i.e., is this the first article shown to individual i, the second, etc.). We include this to capture, for instance, fatigue, which would explain why individuals pay more attention to ads early in the experiment. δ_j and δ_k are respectively article and brand fixed effects, as some articles might be more interesting than others, and some brands more popular than others.

After being asked about brand recall, the individual is then further asked to make actual choices between vouchers for products or cash amounts. We estimate a model for product purchase which is essentially the same as Equation (1), except that the amount that the individuals can obtain by not selecting the voucher (i.e., the price) is also included. Specifically, we estimate the following linear probability model

$$v_{ijk} = \beta att_{ijk} + \gamma X_i + o_{ik} + \delta_j + \eta p_i \times \delta_k + \varepsilon_{ijk} \tag{2}$$

where $v_{ijk} = \{0, 1\}$ is an indicator describing whether or not individual i accepts the voucher for product k. p_i is a variable that captures the price faced by individual i for a voucher for brand k (the random amount of cash (£/\$) offered as an alternative to each voucher). This is the opportunity cost of the voucher. This is interacted with a brand fixed effect δ_k. Hence the specification allows the price elasticity to vary flexibly along the demand curve for each product, and allows these demand curves to be different across products. All other terms are the same as in Equation (1). Our coefficient of interest in both equations is β, which captures the effect of attention on the outcome of interest.

Our key identification argument is the following. Brand ads are assigned to articles randomly, and the order in which these are shown to individuals is also random. We include controls for the order in which a particular ad is shown and the brand. We also include a rich set of individual demographic fixed effects. Conditional on these, our claim is that the attention that an individual devotes to a brand's ad varies exogenously with the (random) article that the brand's ad is placed next to. More specifically, if an article is particularly interesting, and a brand's ad is placed next to that article, then the brand's ad will enjoy more attention from individuals for exogenous reasons.

If the error terms are not correlated with attention (conditional on controls), we can estimate both equations simply with OLS. This is what we do in this section. However, if instead unobserved characteristics drive both attention and the outcome variable (recall or purchase), our results would be biased. We return to this point in the next Section 6, where we confirm our results using an IV specification and a specification with individual fixed effects.

We begin by describing the results on brand recall. These are shown in the first two columns of Table 2, which report results for Equation (1), using two different attention metrics. As shown in column (1), if a brand's ad is visible for one extra second, this increases the probability of the individual remembering that brand by 0.12%. That is, an increase in attention (as measured by the time the ad is visible) of one standard deviation, is associated with an increase in recall of 2.1%.

In line with intuition, results are larger when attention is measured using time individuals spend engaged with the ad (which we measure using eye-tracking). As reported in column (2), if an ad is viewed for one extra second, this increases the probability that the ad is recalled by 3.1%. An increase in one standard deviation of attention, increase recall probability by 9.6%).

The results suggest that attention to ads has a strong and positive impact on brand recall. Hence we confirm previous results from the marketing literature on the relevance of this metric. In addition, we are also able to quantify the impact that an extra second of attention has on recall.

But from an economic point of view, what does this mean? Does advertising indeed lead to *actual* product purchase? In columns (3) and (4) of Table 2 we present the estimates of Equation (2). Attention to an ad has a positive impact on the probability of purchasing the advertised brand. As shown in column (3), if an article is visible for an extra second, this increases the probability of purchasing by 0.13% (an increase in one standard deviation increases the probability of purchase by 2.2%). If an ad is actually

Table 2: Effect of Attention on Recall/Purchase (Linear Probability Model)

	Dependent variable:			
	Recall (0/1)		Purchase (0/1)	
	(1)	(2)	(3)	(4)
Ad Visible	0.0012**		0.0013**	
	(0.0005)		(0.0005)	
Ad Dwell		0.0307***		0.0072**
		(0.0039)		(0.0029)
Brand FE	Y	Y	N	N
Price x Brand FE	N	N	Y	Y
Individual Covariate FE	Y	Y	Y	Y
Observations	5,707	3,925	5,707	3,925
R^2	0.0809	0.1310	0.1343	0.1433

Note: *p<0.1; **p<0.05; ***p<0.01. All specifications include Article (which subsumes Newspaper, Country and Device) and Step Order Fixed Effects. Standard errors clustered at the individual level.

looked at for an extra second, the probability of purchase increases by 0.72% (an increase in one standard deviation in attention increases purchase by 2.2%; see column (4)).

To our knowledge, this link between recall and purchase has not been established in the economics or marketing literature before in a formal way, despite being widely used in research studies and by industry practitioners. "Advertising works" might sound obvious, but here we are giving detailed micro-evidence of its working. We are also able to measure that impact, which could be used to compare advertising effectiveness (perhaps across multiple media) and its costs.

6 Robustness and Heterogeneous Effects

In this section, we first conduct robustness checks of our results. Then we propose an instrumental variable approach. Finally, we consider heterogeneity.

6.1 Robustness Checks

Our main specification assumes that the impact of attention on outcomes is linear. We now add a term that is quadratic in attention to Equations (1) and (2). Our goal is to test if, as it seems intuitive, returns to attention are diminishing. Indeed this turns out to be the case, as reported in Table 3. Results are statistically strong in the case of recall (columns (1) and (2)). Regarding the effects on purchase, the quadratic terms in columns (3) and (4) still have the expected negative sign, but are not statistically different from zero. The first second of attention is valuable, while additional attention would still increase the probability of purchase, or of correct recall, but by a lower margin.[9]

Table 3: Effect of Attention on Recall/Purchase (Linear Probability Model)

	Dependent variable:			
	Recall (0/1)		Purchase (0/1)	
	(1)	(2)	(3)	(4)
Ad Visible	0.0027***		0.0017*	
	(0.0010)		(0.0009)	
Ad Visible sqr.	−0.00001**		−0.000004	
	(0.00001)		(0.00001)	
Ad Dwell		0.0710***		0.0112**
		(0.0073)		(0.0053)
Ad Dwell sqr.		−0.0028***		−0.0003
		(0.0005)		(0.0003)
Brand FE	Y	Y	N	N
Price x Brand FE	N	N	Y	Y
Individual Covariate FE	Y	Y	Y	Y
Observations	5,707	3,925	5,707	3,925
R^2	0.0819	0.1555	0.1344	0.1436

Note: *p<0.1; **p<0.05; ***p<0.01. All specifications include Article (which subsumes Newspaper, Country and Device) and Step Order Fixed Effects. Standard errors clustered at the individual level.

As a second robustness check, Table 2 re-visits the main equations, but in an even more saturated specification with full individual fixed effects (whereas Table 2 used

[9] Please note that we are always on the increasing portion of the attention curve, both for recall and for purchase decisions. For instance, looking at the results on ad dwell in column (4) of Table 3, the attention curve would be declining after the maximum reached at approximately $0.011/(0.0003 \times 2 \approx 18$ seconds, which is well to the right of the median dwell (less than 3 seconds).

fixed effects for individual demographic characteristics such as age, gender, etc.). Results persist, and they are still significant for the impact of ad dwell on recall (column (2)) and for the impact of ad visibility on purchase (column (4)). The remaining coefficient are imprecisely estimated, although the point estimates are quite close to those reported in Table 2.

Table 4: Effect of Attention on Recall/Purchase (Linear Probability Model)

	Dependent variable:			
	Recall (0/1)		Purchase (0/1)	
	(1)	(2)	(3)	(4)
Ad Visible	0.0007		0.0011^{**}	
	(0.0006)		(0.0005)	
Ad Dwell		0.0099^{***}		0.0044
		(0.0034)		(0.0028)
Brand FE	Y	Y	N	N
Price x Brand FE	N	N	Y	Y
Individual FE	Y	Y	Y	Y
Observations	5,707	3,925	5,707	3,925
R^2	0.5124	0.5160	0.4920	0.4871

Note: *p<0.1; **p<0.05; ***p<0.01. All specifications include Article (which subsumes Newspaper, Country and Device), Step Order and Individual Fixed Effects. Standard errors clustered at the individual level.

All the results so far are cast in terms of a linear probability model. We have also considered a logit specification, where all the qualitative effects persist. For brevity, we present this in Appendix B.1.

We return to the fundamental question of the causal impact of attention on outcomes. As written earlier, our results would be biased if an unobservable variable drives both attention and the outcome variable (recall or purchase). This would be the case, for instance, if tall individuals tend to pay more attention to ads and are also wealthier, making them more likely to purchase. Since we cannot observe height, our results would be biased (at least in a specification without individual fixed effects).[10]

[10] We randomized both the order of the articles shown, as well as the ad shown next to each article, which should already take care of such concerns. But we chose to show ads as they would be normally placed next to each article. So we cannot rely on the location of the ads on the page as a random shifter

Our instrument relies on the fact that ads are randomly matched with articles, and articles are of different quality. Some articles are more interesting than others, and this seems, intuitively, a driver of a reader's attention. If readers tend to spend more time on an article, it is also more likely that the ads shown next to that article will be visible and seen by the reader. Instead, the quality of an article should not affect the propensity to purchase or recall a branded product - other than through the attention channel.

Our goal is to obtain a measure of the attractiveness of each article which is independent of each individual's taste over articles. We build our instrument as follows. We consider, for each individual×article, the average amount of attention devoted to that article (but not to the associated ads) by all *other* individuals who were presented that article. We refer to this instrument as the "Leave One Out" (L1O) mean of Article Attention. In Appendix C, we validate this instrument using an independent survey on Mechanical Turk.

We include Country×Device fixed effects since articles were randomized within each country and device. Individuals in one country were not exposed to articles from the other country. The articles shown on desktop and mobile had the same content, but since their format was very different, we effectively consider them to be different articles. This eliminates concerns that, for instance, individuals in the US are more likely to purchase and US articles are on average more interesting (or that US individuals are more likely to devote attention to ads).

This approach is identical to the "random judges" instruments used in Dahl, Kostøl and Mogstad (2014) and Dobbie, Goldin and Yang (2018). In this case, each article is a "judge", and ads are randomly assigned to articles within each country×device.

Beyond this, we control for step order and also include brand fixed effects. We also control for price. Price is randomized in our experimental setting, so its inclusion is not necessary for unbiasedness, but we include it for precision.

The results can be seen in Table 5. As before, we consider two measures of attention:

---

of attention.

Ad Visible and Ad Dwell. For each of these, we present both the first stage and second stage regressions. For Ad Visible, the first stage F-statistic is over 21. For Ad Dwell, the value is approximately 7.3. The first stage regressions confirm that a given individual tends to devote more attention to brand ads randomly placed next to articles which other individuals find more interesting.

The second stage coefficients are estimated imprecisely. However, the point estimates are remarkably close to the estimates in Table 2. The effect of Ad Visible in both specifications is 0.001. The effect of Ad Dwell is 0.007 in Table 2, and 0.011 when attention is instrumented.

Table 5: 2SLS Regression of Purchase on Attention

	Dependent variable:			
	Ad Visible (1st stage)	Buy (0/1) (2SLS)	Ad Dwell (1st stage)	Buy (0/1) (2SLS)
	(1)	(2)	(3)	(4)
Step Order	-0.9172^{***}	0.0012	-0.1873^{***}	0.0021
	(0.0983)	(0.0055)	(0.0192)	(0.0087)
Price	-0.0248	-0.0331^{***}	0.0056	-0.0332^{***}
	(0.1272)	(0.0036)	(0.0248)	(0.0036)
L1O Mean of Article Attention	0.0650^{***}		0.0077^{***}	
	(0.0120)		(0.0023)	
Ad Visible		0.0013		
		(0.0052)		
Ad Dwell				0.0113
				(0.0442)
Observations	3,925	3,925	3,925	3,925
R^2	0.0979	0.0981	0.0361	0.0975
F Statistic	21.1854^{***}		7.3065^{***}	

Note: *p<0.1; **p<0.05; ***p<0.01. Table shows the 1st and 2nd stage of the 2SLS regression. We measure attention using Ad Visible and Ad Dwell. All specifications include Country x Device and Brand Fixed Effects. All specifications include a linear control for price (normalized using PPP) and Step Order.

It is worth noting that step order, which we include as a control, would not be a good instrument. The order in which articles/ads were shown was randomized, and ads shown early in the experiment obtain much more attention. Presumably, this is because individuals become fatigued as the experiment progresses. However, ads seen later in the experiment might also be more salient in individuals' memory, since they were seen closer to the time when individuals were asked to make their recall and purchase decisions. Therefore, this approach could violate the standard IV exclusion restriction.

6.2 Heterogeneity

In this section we comment on heterogeneity in the effects we estimate. We revisit our main specification (Table 2), where we consider the effect of attention devoted to an ad on whether the individual purchases. We re-estimate this effect over subsamples of the data. For brevity, we present the details of the results in the tables of Appendix B.3.

Results on the impact of attention on both recall and purchase persist across those splits. In particular, the effect is present in both countries (US and UK) and relatively stronger in the US (for individuals in the US, an extra second of attention to the ad is associated with a higher increase in the probability of purchase). The effects are also present across both types of devices (desktop and smartphone), and relatively stronger for desktops. The effects are present for both men and women, and relatively stronger for men.

In terms of age, the effect on purchase is strong for young people, and less so for older groups. As we discuss further in the next section, young people are also the group that generally devotes less attention to ads in absolute terms. Hence, conditional on actually getting the attention of younger people, one second of their attention seems to be very valuable. An alternative explanation is that younger people are simply faster at processing information online.

We conclude this section by briefly returning to our main results. We established that attention matters both for recall and for purchase. As mentioned in the Introduction, we are also interested in understanding the impact of the type of news (i.e., hard vs. soft) on outcomes.

When one adds to our main specification a dummy for hard news, we find that, once a viewer devotes attention to an ad, the nature of the article next to it does *not* affect the probability of purchase or recall.[11] In this sense, hard news does not have a negative connotation on brands, as long as ads are viewed and dwelled upon. This result is about the impact of the type of news, over and above the role of attention.

Hard news, however, may receive less attention from readers. The reasons for this could be multiple: perhaps the content is upsetting to readers, or perhaps these stories were frequently repeated in the news at the time of the experiment, so readers are less interested in reading one more article about them. The instrumental variable approach of Section 6.1 indeed shows that individual articles impact attention to ads. We return to this issue in Section 7, where we investigate the drivers of attention.

7 Drivers of Attention

In this section we explore the drivers of attention. Since our setting is about online news, it seems natural that attention is given by readers to articles. Then, this attention devoted to the article can translate to attention to the ad bundled to that article. Indeed, if we correlate our measures of attention to the article with attention to the ad next to it, we find a high and positive correlation both for visibility and dwell (controlling for individual and article fixed effects).[12]

We now explore the drivers of attention to the ads. We note that all results would be qualitatively the same if one investigated the drivers of attention to the article - in fact this is, in our view, the correct interpretation. Attention is given to the article according to its characteristics, and ads are a natural complement to the article.

We consider three types of drivers. First, individual-level characteristics like age, country, and the device being used. Second, article-level characteristics like whether the content was "hard news" or not. Third, we propose an indicator that captures the

[11] Results are presented in the Appendix, Table 10.

[12] Recall that we also asked respondents to read an article without any ad. In Appendix D.1 we regress total time a given article is read (with a full set of individual fixed effects) against a dummy for the presence of an ad next to that article, and we find that the ad decreases dwell time on the article. Article visibility is not impacted by ad visibility, as the two are effectively bundled together.

matching between individual and newspaper, namely in terms of political orientation.

7.1 Individual Characteristics

Table 6 shows results regarding the individual characteristics that drive attention to ads, using a specification that includes article fixed effects (which subsumes country, newspaper and device fixed effects).

Table 6: Attention and Individual Characteristics

	Dependent variable:	
	Ad Visible	Ad Dwell
	(1)	(2)
Female	2.3782***	0.3583**
	(0.8899)	(0.1682)
Age: 25-34	−1.1708	0.5178**
	(1.9192)	(0.2567)
Age: 35-44	0.4733	0.9306***
	(1.9752)	(0.2471)
Age: 45-54	1.0438	0.9153***
	(2.0965)	(0.2658)
Age: 55-65	2.8242	1.2748***
	(2.2586)	(0.3913)
Age: 65+	4.8996**	0.3186
	(2.3899)	(0.3031)
Observations	6,428	4,423
R^2	0.1534	0.1106

Note: *p<0.1; **p<0.05; ***p<0.01. Fixed effects: Income, Education, Politics, Brand, Step Order, Article. Standard errors clustered at the individual level.

We find that some individual characteristics matter, in particular gender and age.[13] Men typically devote less attention than women. Among younger individuals, attention is lower relative to older ones. However, for these individuals, the relationship between attention and purchase is large and statistically significant (see Appendix B.3 on heterogeneous effects), which suggests that these individuals are not simply ignoring ads, but

[13]Because we have included article fixed effects, and each article is defined by country and device, in Table 6 there are no results related to country or device.

appear instead to be processing ads more quickly than other individuals. The remaining individual covariates (e.g., education, income, etc.) were not found to be robust predictors of attention, so they are not reported in Table 6 (but are included as a fixed effects). In particular, some income brackets were actually significant predictors, but no clear pattern was in the data. In a specification without article fixed effects (but with newspaper fixed effects, which subsumes country fixed effects), the device also matters: people on desktops devote much more attention to the articles compared to people using mobile devices.

7.2 Article Characteristics

Second, we look at article characteristics, in a specification that includes full individual fixed effects. We focus on how hard news reduces attention to the article and to the ad. Half of the articles we chose were "hard news" in the sense that they are of topics typically considered to be sensitive by the advertising industry. Based on discussions with industry experts, we included articles about the COVID-19 pandemic and the BLM protests, since these were often blocked by advertising intermediaries. Recall that the experiment took place in late July 2020. We validate these choices in an independent survey on Mechanical Turk, discussed in Appendix C.

From the outcome variables Ad Dwell and Ad Visible, we obtain the amount of attention devoted to the article itself, excluding attention devoted to the ad. That is, we compute

$$\begin{aligned} \text{Article Visible} &= \text{Page Visible} - \text{Ad Visible} \\ \text{Article Dwell} &= \text{Page Dwell} - \text{Ad Dwell} \end{aligned}$$

and then we regress these metrics on a dummy for hard news.

Results are shown in Table 7. To increase precision, we include fixed effects for the step order, brand and newspaper, and individual fixed effects. We find that hard news articles, and ads randomly shown next to these articles, receive less attention compared to 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 attention dwell, there is a reduction of almost 7 seconds for the article (about 10% of the median article dwell), and a reduction of 0.4 seconds for the ad (about 14% of the median ad dwell). Since ads are randomly assigned to articles, these effects are causal.

Table 7: Attention and Hard News

	Measure of attention:			
	Ad Visible	Ad Dwell	Article Visible	Article Dwell
	(1)	(2)	(3)	(4)
Hard News	-1.0674^{***}	-0.3875^{***}	-10.3569^{***}	-6.8927^{***}
	(0.2985)	(0.0742)	(2.4232)	(1.9845)
Observations	6,431	4,426	6,431	4,426
R^2	0.6174	0.5005	0.6153	0.6083

Note: *p<0.1; **p<0.05; ***p<0.01. Fixed Effects: Individual, Step Order, Brand, Newspaper. Standard errors clustered at the individual level.

These results should be interpreted with caution. It is possible that, since there were, at the time of the experiment, many hard news stories, individuals could already have been informed about those stories (the experiment did not allow to test for pre-experiment knowledge), or possibly individuals were weary of such stories. Hence we cannot say if our finding is because people do not like to read about hard news, or because we showed them articles related to news they already knew about, and so they skimmed through them quickly.

7.3 Individual-Article Match

Finally, we consider an indicator of match/mismatch between individuals and newspapers. Recall that individuals are able to see the newspaper from which each story originates (a banner is shown at the top of each article clearly showing the news source). Also, at the very end of the experiment, respondents were asked about their political views (this was done so that the question would not bias the other responses by individuals). 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, while the Daily Mail is on the right. In the US, the NYT is left leaning, while USA Today is centrist. We validate these choices in an independent survey on Mechanical Turk, discussed in Appendix C.

For instance, does an individual with self-reported liberal views react differently to news when such news is shown by a newspaper that leans to the left? We first build an index of "right-wing-ness" for each newspaper and individual. Regarding newspapers, The Daily Mail is assigned +1, USA Today is assigned 0, while The New York Times and The Guardian are assigned -1.[14]

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 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 is created (mismatch = 2) when that person is presented with an article from an outlet at the opposite end of the political spectrum. Intermediate cases can arise from other combinations.

Table 8 shows results with respect to the mismatch between individual political leaning and newspaper-level leaning. The results clearly suggest that, if the match between individual and newspaper is poor, attention to the page is lower. Similar results hold for attention to the ad.

These findings are quite interesting, and go beyond our more limited exercise. They go more to the core of how articles are written, and how news caters for its expected audiences.

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

Table 8: Attention and Political Mismatch

	Dependent variable:			
	Ad Visible	Ad Dwell	Article Visible	Article Dwell
	(1)	(2)	(3)	(4)
Politics Mismatch (0/1/2)	-0.8714^{***}	-0.1455^{*}	-6.0959^{*}	-7.7653^{***}
	(0.2641)	(0.0778)	(3.3199)	(2.4150)
Observations	6,037	4,115	6,037	4,115
R^2	0.6352	0.5139	0.6414	0.6411

Note: *p<0.1; **p<0.05; ***p<0.01. Fixed Effects: Individual, Article, Brand, Step Order. Standard errors clustered at the individual level.

8 Conclusion

This paper has proposed a measure of attention using eye-tracking and used it to estimate advertising effectiveness in online markets. We set up an experiment focusing on display adverting online, where ads are shown next to news articles. We have shown that attention to specific ads increases the probability that the advertised brands are both correctly recalled and subsequently purchased.

One managerial implication of our paper is the alignment between the objectives that the Chief Marketing Officer (CMO) and Chief Financial Officer (CFO) of a firm have with regard to the impact of advertising. CMOs are typically more interested in lifting brand health indicators, such as brand awareness, recall and recognition. These indicators help marketers understand the state of mind of consumers. In contrast, CFOs typically assess advertising impact with measures such as purchases and return on sales. We show there is not a conflict. If an ad is well designed to grab a user's attention, it will be driving positively both sets of metrics.

On the conceptualization side, and related to the previous remark, the managerial relevance of our contribution is that we establish a link between two important metrics that are part of the consumer purchase funnel: recall and purchases. Recall is a metric that assesses 'cognition', which is an upper funnel metric while purchase is a 'behavioral', lower funnel metric. The role of advertising is twofold: (i) get a strong position in customers' minds and (ii) get a sales conversion. We establish the connection between recall and purchase, as they are both driven by attention, that is, the time the user dwells on the ads.

Our experiment suffers from limitations typical of similar experimental settings. First, in our experiment, we asked individuals to make an *immediate* purchase decision, so we are likely overestimating the effect that a real ad would have on purchases. We note, however, that the brand-specific vouchers that individuals could obtain are valid for one year or more, so *consumption* does not need to be immediate, hence possibly mitigating the this bias.

Second, we may be underestimating the impact of ads as 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 for large audiences. Still, we acknowledge that we cannot estimate the effectiveness of targetted ads, as this would require access to one of the algorithms that assign ads to readers online, that we do not possess.

Notwithstanding these limitations, we conclude with a back-of-the-envelope exercise that tries to put ballpark figures on costs and benefits of online display ads.

On the benefits side, in our experiment, each brand had an average dwell time of about 2.7 seconds per individual (i.e., time individuals are attentive to the ad). At the mean, this attention increases the probability of purchase by $2.7 \times 0.007 = 0.0019$ or 1.9%. In the US, for instance, the opportunity cost to individuals of acquiring the voucher (the amount of cash individuals had to forego, or the ”price” of the voucher) was on average \$5. Therefore, we take the revenue to the brand from a purchase to be on average \$5. This implies that an ad is worth $5 \times 0.0019 = 9.5$ cents of revenue per person exposed to the ad, or \$95 for 1,000 people. When we run some heterogeneity analysis by device, the benefit per 1,000 people can be further differentiated between desktop (\$127) and smartphones (\$77).[15]

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

On the cost side, the advertising industry typically uses the metric of a ”cost per mille” (CPM, or cost per thousand impressions). For digital inventory this is difficult to assess because it is 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 (i.e., to ensure that their ads are shown to men, or older people, or people who are assumed to be interested in buying cars), which further influences the cost. Still, Lumen Research shared with us their estimate of the cost per *attentive* thousand views (aCPM), which is $21.88 ($\simeq$ \$30) on desktops and £13.54 ($\simeq$ \$19) on mobile devices.[16] On top of this, one would have to include technology and agency fees, that is, the cost of creating the ads and employing the marketers to manage the advertising agencies. These figures suggest that advertising is likely worth its cost.

We conclude by advocating more research on the drivers of attention. Tools such as eye-tracking software are now available to measure that, which is exciting.

[16]See https://www.lumen-research.com/blog/p653k3atys5ubp58jcydxoyn0d0wik.

References

Bagwell, Kyle. 2007. "The economic analysis of advertising." *Handbook of Industrial Organization*, Vol. 3: 1701–1844.

Balcombe, Kelvin, Iain Fraser, and Eugene McSorley. 2015. "Visual attention and attribute attendance in multi-attribute choice experiments." *Journal of Applied Econometrics*, 30(3): 447–467.

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

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman. 2010. "What's advertising content worth? Evidence from a consumer credit marketing field experiment." *The Quarterly Journal of Economics*, 125(1): 263–306.

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

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

Chandon, Pierre, J Wesley Hutchinson, Eric T Bradlow, and Scott H Young. 2009. "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.

Cinelli, Carlos, Andrew Forney, and Judea Pearl. 2020. "A Crash Course in Good and Bad Controls." *Available at SSRN 3689437.*

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

Ding, Min, Rajdeep Grewal, and John Liechty. 2005. "Incentive-aligned conjoint analysis." *Journal of Marketing Research*, 42(1): 67–82.

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

Ghose, Anindya, and Vilma Todri. 2016. "Towards a digital attribution model: Measuring the impact of display advertising on online consumer behavior." *MIS Quarterly*, 40(4): 889–910.

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

Gordon, Brett R, Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky. 2019. "A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook." *Marketing Science*, 38(2): 193–225.

Imai, Taisuke, Min Jeong Kang, and Colin F Camerer. 2019. "When the eyes say buy: visual fixations during hypothetical consumer choice improve prediction of actual purchases." *Journal of the Economic Science Association*, 5(1): 112–122.

Johnson, Garrett. 2020. "Inferno: A guide to field experiments in online display advertising." *Available at SSRN*.

Khurram, Mehreen, Faisal Qadeer, and Muhammad Sheeraz. 2018. "The Role of Brand Recall, Brand Recognition and Price Consciousness in Understanding Actual Purchase." *Journal of Research in Social Sciences*, 6(2): 219–241.

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

Lewis, Randall A, and David H Reiley. 2014. "Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo!" *Quantitative Marketing and Economics*, 12(3): 235–266.

Lewis, Randall A, and Justin M Rao. 2015. "The unfavorable economics of measuring the returns to advertising." *The Quarterly Journal of Economics*, 130(4): 1941–1973.

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

Martinovici, Ana, Pieters-FGM, and Tulin Erdem. 2021. "Attention trajectories predict brand choices." *Available on SSRN*.

Neumann, Nico, Catherine E Tucker, and Timothy Whitfield. 2019. "Frontiers: How effective is third-party consumer profiling? Evidence from field studies." *Marketing Science*, 38(6): 918–926.

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

Schmidt, Jonas, and Tammo HA Bijmolt. 2019. "Accurately measuring willingness to pay for consumer goods: A meta-analysis of the hypothetical bias." *Journal of the Academy of Marketing Science*, 1–20.

Takahashi, Ryo, Yasuyuki Todo, and Yukihiko Funaki. 2018. "How can we motivate consumers to purchase certified forest coffee? Evidence from a laboratory randomized experiment using eye-trackers." *Ecological Economics*, 150: 107–121.

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

Appendix

A Experimental Details

We selected branded products that would be of general interest to a wide audience. We picked products that would be relatively easy to redeem with an e-voucher. We also chose brands for which we could find brand-specific vouchers.[17] We also tried to ensure that the types of product categories would be similar between the two countries. The table below reports the chosen brands.

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

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 headlines).

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

[17] The vouchers were purchased on the specialized websites GiftPay and Tango Card.

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

The protocol received ethical approval from 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 with in the AEA RCT Registry with RCT ID AEARCTR-0006010.[18]

For the interested reader, we provide below links to the full experiment:

US Desktop   UK Desktop   US Mobile   UK Mobile

The workings of the eye-tracking technology are summarized in the top panel of Figure 4. 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 in the eye model and the gaze point for each image sample is calculated. When the procedure is finished, the quality of the calibration is illustrated by green lines of varying length (see the lower panel of Figure 4 for an example involving one of the authors of this paper). Figure 5 contains a visual summary of the experimental protocol.

B Robustness

B.1 Logit

Table 9 presents a robustness check where we depart from our main specification by using a Logit model (rather than a linear probability model). Results are quantitatively similar.

[18] See https://www.socialscienceregistry.org/trials/6010/history/73163.

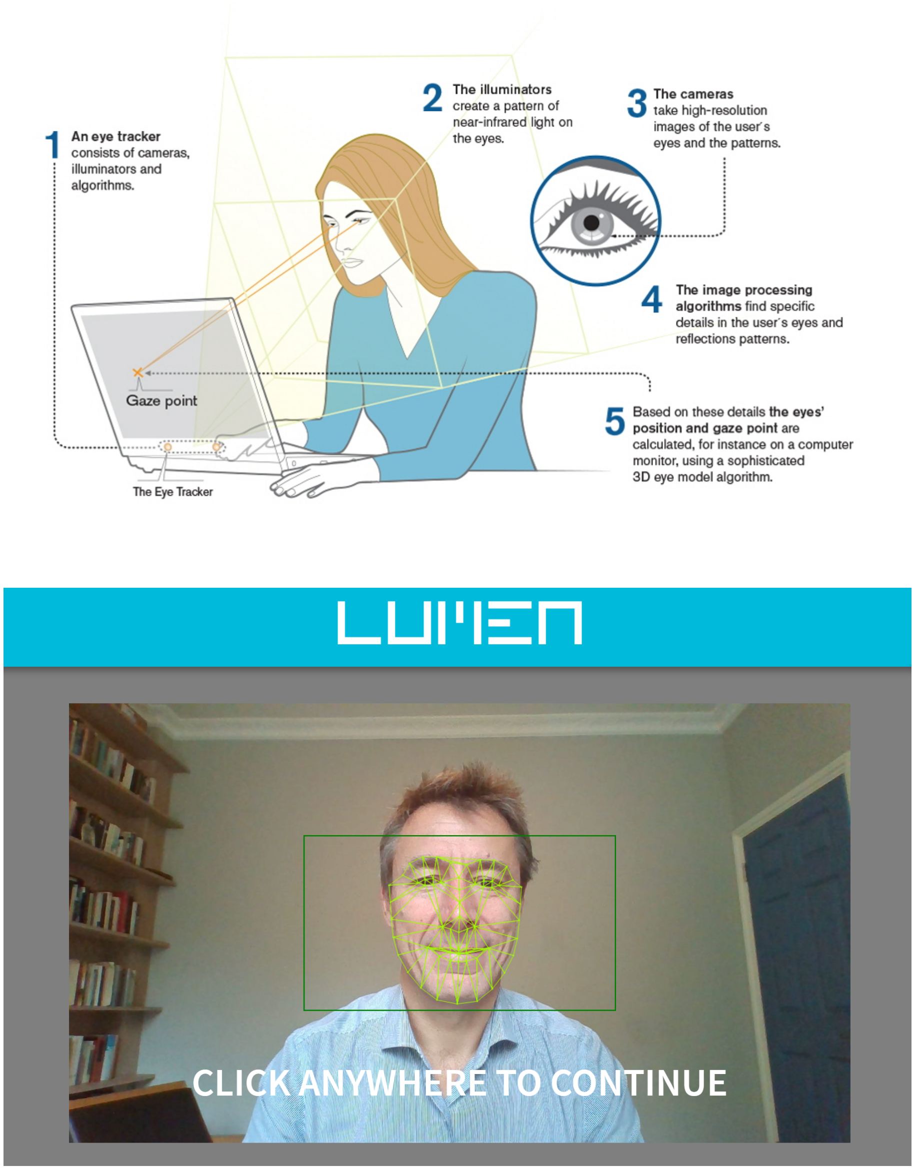

Figure 4: Eye-tracking technology

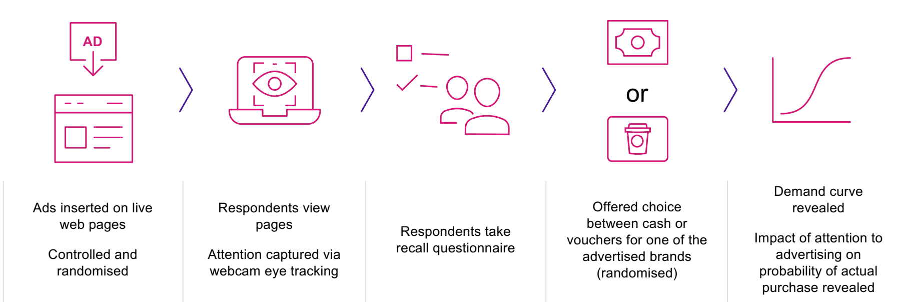

Figure 5: Research protocol

Table 9: Effect of Attention on Recall/Purchase (Logit)

	Dependent variable:			
	Recall (0/1)		Purchase (0/1)	
	(1)	(2)	(3)	(4)
Ad Visible	0.0065***		0.0067***	
	(0.0017)		(0.0018)	
Ad Dwell		0.1598***		0.0324***
		(0.0134)		(0.0118)
Brand FE	Y	Y	N	N
Price x Brand FE	N	N	Y	Y
Individual Covariate FE	Y	Y	Y	Y
Observations	5,707	3,925	5,707	3,925

Note: *p<0.1; **p<0.05; ***p<0.01. All specifications include Article (which subsumes Newspaper, Country and Device) and Step Order Fixed Effects.

B.2 Effect of Hard News on Purchase

Table 10 shows that attention devoted to an ad is predictive of the individual recalling the brand and purchasing the product. However, including a dummy variable for whether the article is hard news (conditional on the country, newspaper and device) has no predictive power. The coefficient is imprecisely estimated but if anything is positive, which constitutes evidence against the commonly held idea that advertising next to hard news reduces brand recall and purchases *per se*.

B.3 Heterogeneity

The Tables below present the results of the impact of attention on brand recall and product purchase when we split the sample by country, type of device, gender, and age, as commented in Section 6.3. In the interest of brevity, we present only the results of the impact of these covariates on ad dwell. The impact of ad visibility is similar, and the results are available from the authors on request.

These tables include (when possible) the following sets of fixed effects: price*brand, article, step order, income, gender, education, age, country, device, political orientation.

Table 10: Effect of Attention on Recall/Purchase (Linear Probability Model)

	Dependent variable:			
	Recall (0/1)		Purchase (0/1)	
	(1)	(2)	(3)	(4)
Ad Visible	0.0012**		0.0013**	
	(0.0005)		(0.0005)	
Ad Dwell		0.0312***		0.0069**
		(0.0038)		(0.0029)
Hard News	0.0054	0.0155	0.0050	0.0151
	(0.0110)	(0.0132)	(0.0108)	(0.0135)
Brand FE	Y	Y	N	N
Price x Brand FE	N	N	Y	Y
Individual Covariate FE	Y	Y	Y	Y
Observations	5,707	3,925	5,707	3,925
R^2	0.0734	0.1250	0.1284	0.1361

Note: *p<0.1; **p<0.05; ***p<0.01. All specifications include Newspaper, Country, Device and Step Order (but not Article) Fixed Effects. Individual covariate fixed effects include: income, gender, education, age (in bins of 10 years), and self-reported political leaning.

Table 11: Heterogeneity by Device

	Dependent variable:			
	Recall			
	Desktop	Mobile	Desktop	Mobile
	(1)	(2)	(3)	(4)
Ad Dwell	0.0344***	0.0242***	0.0094***	0.0057
	(0.0034)	(0.0039)	(0.0034)	(0.0036)
Observations	2,101	1,824	2,101	1,824
R^2	0.2102	0.1512	0.1565	0.2055

Note: *p<0.1; **p<0.05; ***p<0.01. Fixed Effects: Price x Brand, Article, Step Order, Income, Gender, Education, Age, Country, Politics

Table 12: Heterogeneity by Gender

	Dependent variable:			
	Recall			
	Female	Male	Female	Male
	(1)	(2)	(3)	(4)
Ad Dwell	0.0242***	0.0409***	0.0057*	0.0128***
	(0.0033)	(0.0042)	(0.0031)	(0.0039)
Observations	2,249	1,676	2,249	1,676
R^2	0.1672	0.2199	0.1877	0.1950

Note: *p<0.1; **p<0.05; ***p<0.01. Fixed Effects: Price x Brand, Article, Step Order, Income, Education, Age, Country, Device, Politics

Table 13: Heterogeneity by Country

	Dependent variable:			
	Recall			
	UK	US	UK	US
	(1)	(2)	(3)	(4)
Ad Dwell	0.0281***	0.0317***	0.0058*	0.0084**
	(0.0036)	(0.0036)	(0.0033)	(0.0036)
Observations	2,093	1,832	2,093	1,832
R^2	0.1602	0.1530	0.1202	0.1519

Note: *p<0.1; **p<0.05; ***p<0.01. Fixed Effects: Price x Brand, Article, Step Order, Income, Gender, Education, Age, Device, Politics

Table 14: Purchase by age

	Dependent variable: Purchase					
	Age: 18-24	Age: 25-34	Age: 35-44	Age: 45-54	Age: 55-64	Age: 65+
Ad Dwell	−0.001	0.015***	0.008	0.001	0.005	0.011
	(0.010)	(0.005)	(0.005)	(0.005)	(0.008)	(0.011)
Observations	396	941	1,067	845	363	313
R^2	0.570	0.258	0.249	0.314	0.467	0.626
Adjusted R^2	0.320	0.123	0.130	0.170	0.119	0.313
Residual Std. Error	0.394	0.446	0.450	0.437	0.436	0.362

Note: *p<0.1; **p<0.05; ***p<0.01
Fixed Effects: Price*brand, Article, Step Order, Income, Gender, Education, Country, Device, Politics

Table 15: Recall by age

	Dependent variable: Recall					
	Age: 18-24	Age: 25-34	Age: 35-44	Age: 45-54	Age: 55-64	Age: 65+
Ad Dwell	0.030**	0.023***	0.027***	0.031***	0.035***	0.029*
	(0.012)	(0.006)	(0.005)	(0.006)	(0.009)	(0.015)
Observations	396	941	1,067	845	363	313
R^2	0.454	0.214	0.233	0.297	0.479	0.475
Adjusted R^2	0.137	0.070	0.111	0.150	0.138	0.037
Residual Std. Error	0.443	0.481	0.471	0.457	0.464	0.471

Note: *p<0.1; **p<0.05; ***p<0.01
Fixed Effects: Price*brand, Article, Step Order, Income, Gender, Education, Country, Device, Politics

C 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 US. All individuals 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 (we obtained about 2,300 observations). We showed individuals articles without ads. We asked individuals 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 is the political slant of the article. We then compute the right-wing 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.

The outcomes of the AMT survey are shown in Table 16. 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.

Table 16: Validation Summary Statistics

Newspaper	Hard	E[Upset]	SD[Upset]	E[Rightwing]	SD[Rightwing]	E[Interest]	SD[Interest]
Guardian	FALSE	2.216	1.073	-0.121	0.600	3.155	1.074
Guardian	FALSE	2.577	1.162	0.111	0.708	3.082	0.997
Guardian	TRUE	1.871	1.111	0.039	0.572	2.921	0.880
Guardian	TRUE	1.972	1.172	-0.418	0.731	3.333	0.947
Guardian	TRUE	2.924	1.214	-0.194	0.623	3.352	0.951
Mail	FALSE	1.733	1.031	-0.114	0.689	3.104	1.095
Mail	FALSE	1.720	1.147	0.130	0.626	2.704	1.185
Mail	TRUE	2.911	1.162	-0.237	0.788	3.290	0.961
Mail	TRUE	2.105	1.146	0.331	0.838	2.798	1.104
NYT	FALSE	1.894	1.065	-0.214	0.717	3	1.229
NYT	FALSE	1.623	1.051	-0.077	0.478	3.719	1.133
NYT	TRUE	3.008	1.252	-0.542	0.608	3.438	1.182
NYT	TRUE	2.603	1.253	-0.225	0.825	3.087	1.124
USAT	FALSE	1.540	1.061	-0.061	0.551	3.611	1.056
USAT	FALSE	2.536	1.259	-0.148	0.656	3.010	1.220
USAT	TRUE	2.632	1.212	-0.074	0.581	3.379	1.064
USAT	TRUE	3.284	1.310	-0.291	0.734	3.316	1.187
USAT	TRUE	3.420	1.249	-0.250	0.638	3.610	1.180

Table 17 shows a regression of the average score obtained by each one of the 18 articles on the AMT survey on our subjective definition of hard news attributed to those articles. The positive coefficient suggests a validation of our subjective definition.

Table 17: Validation of Hard News Measure

	Dependent variable:
	Article Upsetting (1-5)
Hard News	0.6930***
	(0.2313)
Constant	1.9800***
	(0.1724)
Observations	18
R^2	0.3594

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

We used the same procedure for the political leaning of each newspaper. Table 18 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 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 the way we coded the data.

Table 18: Validation of Political Leaning

	Article is Right-Wing (1-5)	
	US	UK
	(1)	(2)
USAT	0.0999	
	(0.1011)	
Mail		0.1442
		(0.1533)
Constant	-0.2645^{***}	-0.1166
	(0.0754)	(0.1022)
Observations	9	9
R^2	0.1223	0.1122

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

Last, we also asked the survey respondents on AMT to express their views on how ”interesting” each article is. Taking the average score of each article on this dimension, and regressing it on the L1O variable for that article that we used as our instrumental variable, shows a positive coefficient between the two (results not reported for brevity, but available from the authors).

Finally, Table 19 shows that individuals in our online experiment spend more time reading articles (as measured by our eye-tracking tool) that individuals on AMT deemed more interesting. This regression includes only data for individuals doing the original online experiment on desktop and for whom we have high quality eye-tracking data, since the survey on AMT showed only articles in the desktop format.

D Other Results

D.1 The effect of ads on article attention

In this section we investigate what effect ads have on attention devoted to articles. In the experiment, for each individual, one of the 9 articles was randomly presented without

Table 19: Validation of Article Interest

	Dependent variable:
	Article is Interesting (1-5)
Article Dwell	0.0001**
	(0.0001)
Constant	3.1876***
	(0.0071)
Observations	2,371
R^2	0.0026

Note: *p<0.1; **p<0.05; ***p<0.01. One observation per individual x article.

any ads. In this context, our measure of attention is that time in which the article was being actively looked at (e.g., Article dwell, that is, Page dwell minus Ad dwell).

We regress attention on a dummy variable which equals 1 if there is no ad shown next to the article. We also include individual, article and step order fixed effects. Results are shown in Table 20. As expected, we find no effect on Article Visible since, whenever an ad is visible, the page is necessarily also visible (column (1)). More importantly, when an ad next to an article is missing, the total dwell time devoted to that article increases by approximately 6.6 seconds. Since the average dwell on a page is 77.3 seconds, this corresponds to an 8.5% increase in average dwelling time spent on a page.

Table 20: Effects of Ads on Attention to Article

	Dependent variable:	
	Article Visible	Article Dwell
	(1)	(2)
No Ad	0.1886	6.6075*
	(4.1992)	(3.4903)
Observations	6,431	4,426
R^2	0.6390	0.6399

Note: *p<0.1; **p<0.05; ***p<0.01. Fixed Effects: Individual, Article, Step Order. Standard errors clustered at the individual level.

D.2 The effect of recall on purchase

In this article, we have not taken a particular stance on whether recall is an intermediary performance variable, ultimately leading to ad impact, or is determined simultaneously with the decision to purchase. Still, a hierarchical sequence of events, whereby first attention leads to recall, which then eventually converts into actual purchase, seems particularly natural.

While this question is open for further research, below we conduct a simple exercise to shed some light on this question. In Table 21 we re-estimate our main regression equation, but we add recall as an independent variable (compare to columns (3) and (4) of Table 2). Notice that in this setting, the variable recall is an endogenous control variable (i.e., a "bad control"; see Cinelli, Forney and Pearl (2020)). Adding recall as a control reduces the coefficient on attention, suggesting that the effect of attention on purchase indeed operates primarily through recall. Notice that the attention variables are also strong predictors of recall (Table 2).

Table 21: Determinants of Recall/Purchase (Linear Probability Model)

	Dependent variable:	
	Purchase (0/1)	
	(1)	(2)
Ad Visible	0.0010**	
	(0.0005)	
Ad Dwell		0.0034
		(0.0027)
Recall	0.1053***	0.1122***
	(0.0145)	(0.0173)
Observations	5,707	3,925
R^2	0.4979	0.4937

Note: *p<0.1; **p<0.05; ***p<0.01. All specifications include Step Order, Individual, Article (which subsumes Country, Newspaper and Device) and Price x Brand Fixed Effects. Standard errors clustered at the individual level.