CESifo

AREA CONFERENCES 2020

Economics of Digitization

Munich, 19–20 November 2020

Attention, Recall and Purchases: Experimental Evidence on Online News and Advertising

Tommaso Valletti and Andre Veiga

CESifo

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

Tommaso Valletti* and Andre Veiga†

November 4, 2020

Abstract

We conduct an experiment where 1,000 individuals read online news articles and are shown ads for branded goods next to those articles. Using eye-tracking technology, we measure the attention that each individual devotes to reading each article and viewing each ad. Then, respondents choose between cash or vouchers for branded goods. We find that attention is a predictor both of willingness-to-pay for brands, and for brand recall. We also give suggestive evidence of the main drivers of attention. These 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

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

†Imperial College London, a.veiga@imperial.ac.uk. We gratefully acknowledge financial help from the NET Institute (www.NETinst.org)

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 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 than can only be used to purchase from that brand. The effect is small, but positive, and 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 industry and the marketing literatures, but less so in the economics literature. Hence, we establish a substantial 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, which is what we do in this paper. 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 preliminary evidence regarding the drivers of attention. 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 associated to the fact that ads are served by intermediaries that often block ads from being seen next to certain words associated to 'hard news'. The rationale behind this blocking is the fear that the brand might otherwise suffer from a negative impression coming from the article, which would hurt its brand image.

In practice, it seems that articles reporting on COVID-19 or the Black Lives Matter Movement, BLM) are often considered to be “hard news” and blocked by intermediaries. We take this industry standard to be also our definition of “hard news”. We find that the type of news does directly affects the degree of attention. However, controlling for attention, whether an article is hard news or not does not have any additional effect on recall or purchase. We discuss the implications of these results for managers and advertisers and discuss why these results should be interpreted with caution.

We also 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.

2 Literature

Attention to products and advertising has been studied widely in the marketing literature. Chandon et al. (2009) show that in-store location of products is related to attention and purchase. Macdonald and Sharp (2000) show that brand awareness is linked to purchases. Neither of these articles considers online environments, as we do.

Ghose and Todri (2015) study a quasi-experimental setting and measure the impact of advertising on consumer search for the product, and on purchase. This study does not consider eye-tracking, and does not examine how content (e.g., articles) influences attention to ads. Lewis and Reiley (2014) appraise the returns to online advertising, but does not consider the role of attention.

Lewis and Rao (2015) and Gordon et al. (2019) argue that measuring the effectiveness of advertising with observational data is difficult due to lower power and endogeneity, and that RCT can improve on observational methods. Our experimental setting attempts to fill this gap.

Goldfarb and Tucker (2011) and Neumann, Tucker and Whitfield (2019) study online targeting and show that exposure to ads is related to ad recall. They also show that the type of targeting done by the ad affects its effectiveness. These authors do not consider actual purchases.

This article is also related to a 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.

An increasing literature uses eye-tracking in various related contexts (e.g., Brocas et al. (2014), Camerer et al. (1993), Knoepfle, Wang and Camerer (2009), Reutskaja et al. (2011)). These studies do not focus on the effects of online advertising.

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 research and marketing panels.

Each respondent was first asked to self report several socio-demographic characteristics (in particular: age, education, income, gender and postcode).[1]

Then participants were invited 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 differed in their political leaning.

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

We chose articles which were split evenly between 'hard' and 'soft' news (as mentioned in the Introduction, the former were articles related to COVID-19 and BLM).

We chose ads from well-known and widely available product brands. Ads were inserted into the article pages as they would be normally. In each country, articles were split evenly between two outlets (in order to assess the impact of the type of media outlet).

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

For each individual, measures of attention to each article and ad were recorded. In particular, the experiment recorded the amount of time the article was visible and the ad was 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 (more on this 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 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 so they are well known in each of the countries.

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 £7-3 in the UK, and $7-3 in the US). Individuals were asked to make one choice for each brand they had seen, and they were told they would be sent electronically one outcome of their choice, which was then administered again via Panelbase at the end of the experiment.

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 stating their preferences.

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

See Appendix A.1 for more details about the experiment, including the specific articles and brands.

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 a software that uses the camera of a desktop/mobile phone and measures how the eye retina looks at the screen. It can measure both which parts of the screen are viewable (which in our analysis we split between article and ad) and for how long, as well as how much time is actually spent viewing that page (which is called dwell in our analysis).[2] These are our measures of 'attention'.

A heat map is provided in Figure 1 as an example of how these metrics can be 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 report heat maps of two different types of ads.

Table 1 presents the summary statistics. As mentioned earlier, each respondent was

[2] Eye-tracking, according to Lumen, allows to see "what people actually do, not what they say they do". More details are in Appendix A1.

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

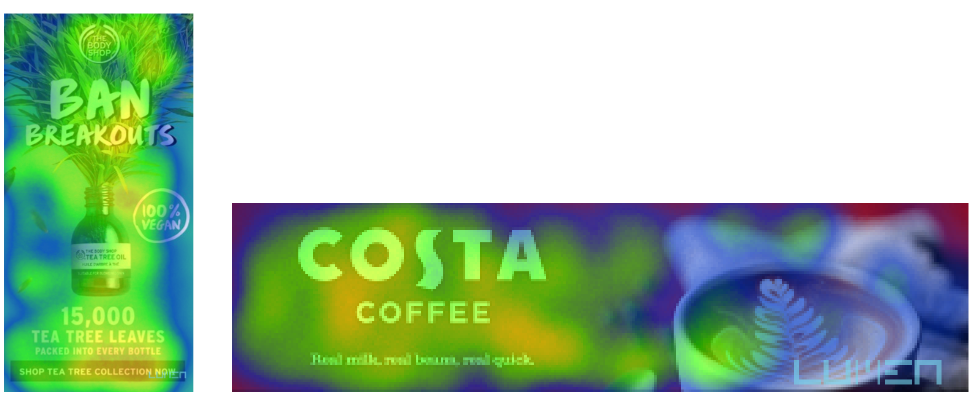


Figure 2: Heat map of ads for two branded products

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

asked to read 9 articles and to make choices involving 8 brands. Observations were split evenly between US/UK, 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.[3] An ad is counted as 'viewable' if 50% of the pixels of an ad are on the screen for more than 1 second. The mean in our sample is 19 seconds. *Ad Dwell* reports the time an ad was actually looked at. The 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). 48% of the respondents recalled a brand correctly after they had seen an ad for that brand.

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

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

5 The impact of attention on recall and purchase

Each individual i reads articles j which show ads for branded products k. The individual is first asked if she recalls the brands shown. Hence, we estimate the following 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 (that is, we use a linear probability model). 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. 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 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 offered is also included. In particular, we estimate

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

where $v_{ijk} = \{0,1\}$ is an indicator describing whether or not individual i accepts the voucher for product k. $p_i = \{3,4,5,6,7\}$ refers to the "price", that is, the random amount of cash (£/\$) offered in alternative to each voucher. In other words, 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. We have a rich set of controls, and we can rely on randomization; hence, 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. If instead unobserved characteristics drive both attention and recall (or purchase), our results would be biased. We return to this point in the next Section 6.

We begin by describing the results on brand recall, which is a metric widely used in the marketing literature. 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), an increase of one second an ad about a product is visible, allows the respondent to increase the probability of remembering that product by 0.1%. That is, an increase in attention of one standard deviation, is associated with an increase in recall of 1.7%.

In line with intuition, results are larger when it comes to dwell. As reported in column (2), one extra second an ad is actually being viewed, is associated with an increase in the probability that the ad is recalled of 2.9%. 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 measure the impact that an extra second of attention has on recall.

But from an economics point of view, what does this mean? Does advertising indeed

Table 2: Determinants of Recall/Purchase (Linear Probability Model)

	Dependent variable:			
	Recall (0/1)		Purchase (0/1)	
	(1)	(2)	(3)	(4)
Ad Visible	0.001***		0.001***	
	(0.0004)		(0.0004)	
Ad Dwell		0.031***		0.007***
		(0.003)		(0.002)
Article FE	Y	Y	Y	Y
Step Order FE	Y	Y	Y	Y
Brand FE	Y	Y	N	N
Price*Brand FE	N	N	Y	Y
Individual Covariate FE	Y	Y	Y	Y
Observations	5,707	3,925	5,707	3,925
R^2	0.081	0.131	0.134	0.143

Note: *p<0.1; **p<0.05; ***p<0.01

Individual covariate fixed effects include: income, gender, education, age (in bins of 10 years), and self-reported political leaning. Country and Device fixed effects are subsumed into the Article fixed effects. Step Order fixed effects include an indicator for the article being shown first in the experiment, second, etc.

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 product. As shown in column (3), if an article is visible for an extra second, this increases the probability of purchasing by 0.1% (an increase in one standard deviation increases the probability of purchase by 1.7%). If an ad is actually *looked at* for an extra second, the probability of purchase increases by 0.7% (and the increase in one standard deviation would increase it 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 practitioners. "Advertising works" might sound obvious, but here we are giving detailed micro-evidence of its working. We are also able to associated some figures to that impact, which could be used to compare advertising effectiveness (perhaps across multiple media) and its costs.

6 Robustness

In this section, we conduct robustness of our results in several directions.

Our main specification assumes that the impact of attention on outcomes is linear. We now add an attention quadratic term in Equations (1) and (2), 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 with recall (columns (1) and (2)). Regarding the effects on purchase, attention is also not monotonic: 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 a purchase, or of correct recall, but by a lower margin.[4]

As a second robustness check, Table 2 re-visits again the main equations, but in an

[4]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 $0.011/(0.0003 \text{ x } 2) \approx 18$ seconds, which is well to the right of the median dwell (less than 3 seconds).

Table 3: Determinants of Recall/Purchase (Linear Probability Model)

	Dependent variable:			
	Recall (0/1)		Purchase (0/1)	
	(1)	(2)	(3)	(4)
Ad Visible	0.003^{***}		0.002^{**}	
	(0.001)		(0.001)	
Ad Visible sqr.	-0.00001^{**}		-0.00000	
	(0.00001)		(0.00001)	
Ad Dwell		0.071^{***}		0.011^{**}
		(0.005)		(0.004)
Ad Dwell sqr.		-0.003^{***}		-0.0003
		(0.0003)		(0.0003)
Article FE	Y	Y	Y	Y
Step Order FE	Y	Y	Y	Y
Brand FE	Y	Y	N	N
Price*Brand FE	N	N	Y	Y
Individual Covariate FE	Y	Y	Y	Y
Observations	5,707	3,925	5,707	3,925
R^2	0.082	0.156	0.134	0.144

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Individual covariate fixed effects include: income, gender, education, age (in bins of 10 years), and self-reported political leaning. Country and Device fixed effects are subsumed into the Article fixed effects. Step Order fixed effects include an indicator for the article being shown first in the experiment, second, etc.

Table 4: Determinants of Recall/Purchase (Linear Probability Model)

	Dependent variable:			
	Recall (0/1)		Purchase (0/1)	
	(1)	(2)	(3)	(4)
Ad Visible	0.001		0.001^{**}	
	(0.0005)		(0.0005)	
Ad Dwell		0.010^{***}		0.004
		(0.003)		(0.003)
Article FE	Y	Y	Y	Y
Step Order FE	Y	Y	Y	Y
Brand FE	Y	Y	N	N
Price*Brand FE	N	N	Y	Y
Individual FE	Y	Y	Y	Y
Observations	5,707	3,925	5,707	3,925
R^2	0.512	0.516	0.492	0.487
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$			

even more saturated specification with full individual fixed effects (whereas Table 2 used fixed effects for individual 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 other results are not statistically significant, though we observe that the point estimates are close to those reported in Table 2, while there is an increase in the standard errors.

All the results so far are cast in terms of a linear probability model. We have also consider a logit specification, where all the qualitative effects persist (see Appendix A.2).

6.1 Instrumental Variables

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 recall or purchase.

We propose an instrumental variable (IV) approach which 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 than the ad bundled 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 individual tastes over articles. We build our instrument as follows. We consider, for each individual and article, the average amount of attention devoted to the article (total page dwell time minus ad dwell time) by all *other* individuals who were presented that article.

We include Country $\times$ 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).

Beyond this, we include fixed effects for Step Order and Brand. We also control for price (which we normalize across countries using the PPP measure). Price is randomized in our experimental setting, so its inclusion is not necessary, but this is done to increase precision and to make the IV specification more similar to the specification in Table 2.

The results can be seen in Table 5. As before, we consider two measures 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 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	-0.917^{***}	0.001	-0.187^{***}	0.002
	(0.098)	(0.006)	(0.019)	(0.009)
Price	-0.025	-0.033^{***}	0.006	-0.033^{***}
	(0.127)	(0.004)	(0.025)	(0.004)
Leave-Out Mean of Artcile Attention	0.065^{***}		0.008^{***}	
	(0.012)		(0.002)	
Ad Visible		0.001		
		(0.005)		
Ad Dwell				0.011
				(0.044)
Constant	23.034^{***}	0.563^{***}	2.792^{***}	0.562^{***}
	(1.749)	(0.149)	(0.341)	(0.152)
Country*Device FE	Y	Y	Y	Y
Brand FE	Y	Y	Y	Y
Observations	3,925	3,925	3,925	3,925
R^2	0.098	0.098	0.036	0.097
F Statistic	21.185^{***}		7.307^{***}	
Note:	*p<0.1; **p<0.05; ***p<0.01			

We note 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 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 would violate the standard IV exclusion restriction.

6.2 Heterogeneity

In this section we briefly comment on heterogeneity in the effects we estimated in Section 5. We run our main specification as in Table 2, but now over subsamples that are split across several dimensions (all the Tables are in Appendix A.3). Results on the impact of attention on both recall and purchase persist across those splits. In particular we find that:

- Effects are strong in both countries, and relatively stronger in the US.
- Effects are strong both on desktops and smartphones, and relatively stronger for desktops.
- Effects are present for both among women and men, 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 will see in the next Section, though, 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 seems to matter both for recall and for purchase. As mentioned in the Introduction, we are also interested to understand if the type of news (i.e., hard vs soft) could have an impact on outcomes. In a specification (not reported for space constraints, but available from the authors) where we remove the article fixed effect, but we introduce a dummy for hard news, we find that, once a viewer gives attention to an ad, the nature of the article next to it does not matter. In this sense, hard news do 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. But hard news might affect attention directly. In fact, the instrumental variable approach that we followed shows that individual articles do impact attention to ads. This is what we study in the next Section where we investigate the drivers of attention.

7 What drives attention

In this section we investigate the drivers of attention. Since our setting is about online news, it seems natural that attention is given by readers to articles. And then this attention 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).[5]

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.

With this clarification in mind, 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 matching between individual and newspaper, namely in terms of political orientation.

Table 6 shows results of individual characteristics, in a specification that controls for article fixed effects. We find that some individual characteristics (age/income/gender) matter.[6] Men give less attention than women, and reading from a mobile phone results in less attention than from a desktop. Education does not show significant results (not

[5]Recall that we also asked respondents to read an article without any ad. In Appendix A.4 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.

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

reported in the Table, but included as a fixed-effect).

Something interesting is that, 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 A.3), which suggests that these individuals are not simply ignoring ads, but might simply be processing ads more quickly than other individuals.

Table 6: Attention and Individual Characteristics

	Dependent variable:	
	Ad Visible	Ad Dwell
	(1)	(2)
Female	2.378***	0.358***
	(0.417)	(0.095)
Age: 25-34	−1.171	0.518***
	(0.813)	(0.174)
Age: 35-44	0.473	0.931***
	(0.795)	(0.169)
Age: 45-54	1.044	0.915***
	(0.817)	(0.174)
Age: 55-65	2.824***	1.275***
	(0.883)	(0.207)
Age: 65+	4.900***	0.319
	(0.967)	(0.216)
Observations	6,428	4,423
R^2	0.153	0.111
Note:	*p<0.1; **p<0.05; ***p<0.01 Fixed effects: Income, Education, Politics, Brand, Step Order, Article.	

Second, we look at article characteristics, in a specification that controls for full individual fixed effects. 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. In our case, and based on discussions with industry experts, we included articles about the COVID-19 pandemic and the BLM protests, since these are often blocked by advertising intermediaries. Recall that the experiment took place in late July 2020.

In addition to the usual outcome variables (Ad Dwell and Ad Visible), we also define the amounts of attention devoted to the article itself, excluding attention devoted to the ad. That is, we compute

$$\text{Article Visible} = \text{Page Visible} - \text{Ad Visible},$$

$$\text{Article Dwell} = \text{Page Dwell} - \text{Ad Dwell}.$$

Using these, we measure how hard news reduces attention to the article and to the ad.

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 lower 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, for instance, 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). Notice that ads are randomly assigned to articles, so these effects have a causal interpretation.

However, these results should be interpreted with caution. It is possible that, since there were, at the time the experiment took place, 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 skipped them.

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). And also recall that, 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 re-

Table 7: Attention and Hard News

	Measure of attention:			
	Ad Visible	Ad Dwell	Article Visible	Article Dwell
	(1)	(2)	(3)	(4)
Hard News	-1.067^{***}	-0.388^{***}	-10.357^{***}	-6.893^{***}
	(0.300)	(0.075)	(2.773)	(2.088)
Observations	6,431	4,426	6,431	4,426
R^2	0.617	0.500	0.615	0.608
Note:	*p<0.1; **p<0.05; ***p<0.01			
	Fixed Effects: Individual, Step Order, Brand, Newspaper			

sponses by individuals). The newspapers we chose have a wide online readership, but are also quite politically oriented. In the U.K., the Guardian has a political alignment on the left, while the Daily Mail is on the right. In the U.S., the NYT is left leaning, while USA Today lies more at the centre.

Does an individual with self-reported liberal views (respectively, moderate or conservative) react differently to news when such news are shown by a newspaper that leans on the left (respectively, centre or right)? 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.[7]

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 on 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.

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

Figure 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 cater for their expected audiences.

Table 8: Attention and political mismatch

	Dependent variable:	
	Ad Visible	Ad Dwell
	(1)	(2)
Politics Mismatch (0/1/2)	-0.871^{***}	-0.145^{**}
	(0.281)	(0.070)
Observations	6,037	4,115
R^2	0.635	0.514
Note:	*p<0.1; **p<0.05; ***p<0.01 Fixed Effects: Individual, Article, Brand, Step Order.	

8 Conclusion

In our experiment we asked individuals to make an immediate purchase, so we are likely overestimating the effect of ads on purchase (though notice that the vouchers individuals obtained were valid for one year, so consumption does need to be immediate). At the same time, we may be underestimating the impact as our ads are not targeted to specific individuals, whereas we showed ads to individuals at random. We acknowledge that we cannot respond to the question of targeted ads, as this would require access to one of the algorithms that assigns ads to readers online, that we do not possess.

Notwithstanding these limitations, which are typical of an experimental setting and of its external validity, we conclude with a back-of-the-envelope exercise that tries to put ballpark figures on costs and benefits of online ads.

On the benefits side, in our experiment we showed ads to individuals which had an average dwell of 2.7 seconds on each set of ads. So, at the mean, attention increases the probability of purchase by 2.7 x 0.007 = 1.9%. Purchase is related to a voucher worth 10 (£/$), which is chosen instead of the alternative of an average amount of cash of 5 (£/$), so the average net profit from purchase is of 5 (£/$). That is, the ad is worth 5 x 0.0019 = 9.5 (p/c) per person, or 95 (£/$) for 1,000 people. When we run some heterogeneity analysis by device, the benefit for 1,000 can be further differentiated between desktop (121 £/$) and smartphones (81 £/$).

On the cost side, the advertising industry refers to 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 also 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 warps the cost. Still, Lumen Research gave us their estimate of cost per thousand views, which is $3 on mobile devices and $18 on desktops.

We finish by advocating more research on what drives attention. Tools are now becoming available to measure that, which is exciting.

References

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.

Ghose, Anindya, and Vilma Todri. 2015. "Towards a digital attribution model: Measuring the impact of display advertising on online consumer behavior." *Available at SSRN 2672090*.

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.

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.

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.

A Appendix

A.1 Experimental Details

We selected branded products that would be of general interest to a wide audience. We also selected products that would be relatively easy to redeem with an e-voucher. We also chose brands for which we could find brand-specific vouchers).[8] We also tried to make sure that the types of product categories would be similar between the two countries. The table below reports the chosen brands.

Type of product/Country	U.S.	U.K.
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 also indicate with an asterisk (*) those articles that we classified as 'hard news'. We also provide the url to retrieve the full article (click on the headlines).

The following articles were sourced from the New York Times (USA):

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

What if the U.S. Bans TikTok?

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

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

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.[9]

For the interested reader, we provide below links to the full experiment (a cell denotes a country-device combination).

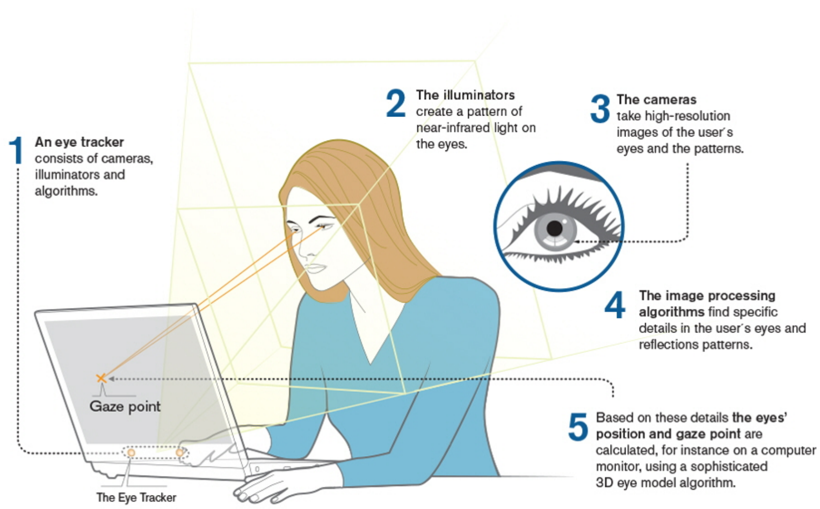
Test Cell and URL
US Desktop
UK Desktop
US Mobile
UK Mobile

The workings of the eye-tracking technology are summarized in the top panel of Figure 3. 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 3 for an example involving one of the authors of this paper).

A.2 Robustness: Logit

We present here the main specification as in Table 2, where we employ a Logit specification instead of a linear probability model.

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



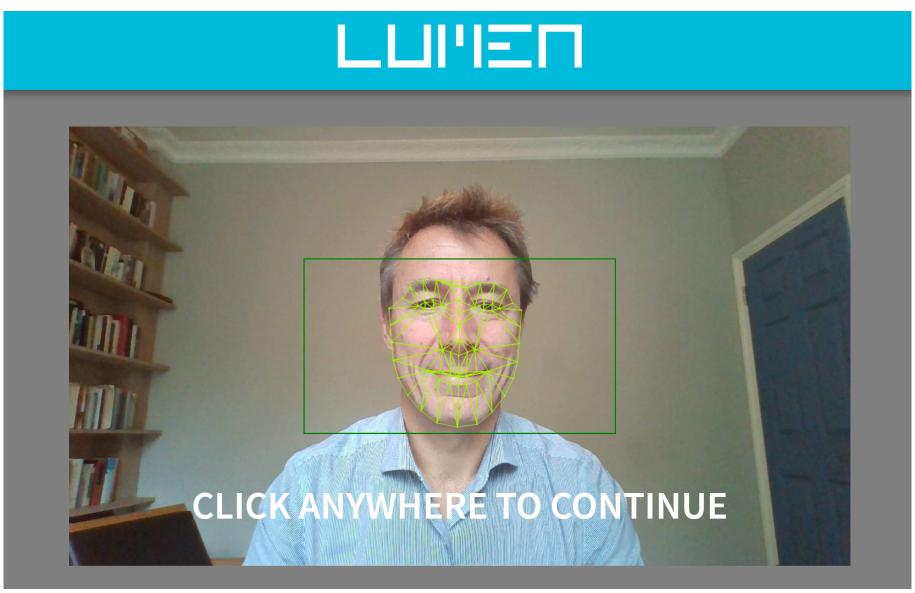


Figure 3: Eye-tracking technology

Table 9: Determinants of Recall/Purchase (Logit Model)

	Dependent variable:			
	Recall (0/1)		Purchase (0/1)	
	(1)	(2)	(3)	(4)
Ad Visible	0.005***		0.007***	
	(0.002)		(0.002)	
Ad Dwell		0.159***		0.036***
		(0.014)		(0.012)
Article FE	Y	Y	Y	Y
Step Order FE	Y	Y	Y	Y
Brand FE	Y	Y	N	N
Price*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			

Individual covariate fixed effects include: income, gender, education, age (in bins of 10 years), and self-reported political leaning. Country and Device fixed effects are subsumed into the Article fixed effects. Step Order fixed effects include an indicator for the article being shown first in the experiment, second, etc.

A.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 (desktop vs mobile), gender, and age, as commented in Section 6.2. To save some space, we present only the results of the impact of 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: Purchase by country

	Dependent variable: Purchase	
	UK	US
Ad Dwell	0.006*	0.008**
	(0.003)	(0.004)
Observations	2,093	1,832
R^2	0.120	0.152
Adjusted R^2	0.083	0.110
Residual Std. Error	0.432	0.464
Note:	*p<0.1; **p<0.05; ***p<0.01 Fixed Effects: Price*brand, Article, Step Order, Income, Gender, Education, Age, Device, Politics	

Table 11: Recall by country

	Dependent variable: Recall	
	UK	US
Ad Dwell	0.028***	0.032***
	(0.004)	(0.004)
Observations	2,093	1,832
R^2	0.160	0.153
Adjusted R^2	0.125	0.111
Residual Std. Error	0.467	0.471
Note:	*p<0.1; **p<0.05; ***p<0.01 Fixed Effects: Price*brand, Article, Step Order, Income, Gender, Education, Age, Device, Politics	

Table 12: Purchase by device

	Dependent variable: Purchase	
	Desktop	Mobile
Ad Dwell	0.009^{***}	0.006
	(0.003)	(0.004)
Observations	2,101	1,824
R^2	0.156	0.205
Adjusted R^2	0.100	0.143
Residual Std. Error	0.451	0.440
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ Fixed Effects: Price*brand, Article, Step Order, Income, Gender, Education, Age, Country, Politics	

Table 13: Recall by device

	Dependent variable: Recall	
	Desktop	Mobile
Ad Dwell	0.034^{***}	0.024^{***}
	(0.003)	(0.004)
Observations	2,101	1,824
R^2	0.210	0.151
Adjusted R^2	0.157	0.085
Residual Std. Error	0.459	0.476
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ Fixed Effects: Price*brand, Article, Step Order, Income, Gender, Education, Age, Country, Politics	

Table 14: Purchase by gender

	Dependent variable: Purchase	
	Female	Male
Ad Dwell	0.006^{*}	0.013^{***}
	(0.003)	(0.004)
Observations	2,249	1,676
R^2	0.188	0.195
Adjusted R^2	0.130	0.116
Residual Std. Error	0.452	0.431
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ Fixed Effects: Price*brand, Article, Step Order, Income, Education, Age, Country, Device, Politics	

Table 15: Recall by gender

	Dependent variable: Recall	
	Female	Male
Ad Dwell	0.024***	0.041***
	(0.003)	(0.004)
Observations	2,249	1,676
R^2	0.167	0.220
Adjusted R^2	0.108	0.144
Residual Std. Error	0.472	0.460
Note:		*p<0.1; **p<0.05; ***p<0.01 Fixed Effects: Price*brand, Article, Step Order, Income, Education, Age, Country, Device, Politics

Table 16: 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 17: 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

A.4 The effect of ads on articles

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 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 18. 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 18: Effects of Ads on Attention to Article

	Dependent variable:	
	Article Visible	Article Dwell
	(1)	(2)
No Ad	0.189	6.607**
	(4.267)	(3.183)
Observations	6,431	4,426
R^2	0.639	0.640
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Fixed Effects: Individual, Article, Step Order.	