The Effects of Exposure Time on Memory of Display Advertisements

Daniel G. Goldstein
Yahoo! Research
New York, NY, USA
dgg@yahoo-inc.com

R. Preston McAfee
Yahoo! Research
Burbank, CA, USA
mcafee@yahoo-inc.com

Siddharth Suri
Yahoo! Research
New York, NY, USA
suri@yahoo-inc.com

ABSTRACT

Display advertising is a multi-billion dollar industry that has traditionally used a pricing scheme based on the number of impressions delivered. The number of impressions of an ad is simply the number of downloads of that ad. One impression, however, does not differentiate between an ad that is in view for five seconds or five minutes. Since advertisers seek brand recognition and recall, we ask whether a time-based accounting of advertising can better align with advertisers' goals. This work aims to model the basic relationship between ad exposure time and the probability that a viewer will remember an advertisement. We investigate this question via two behavioral experiments, conducted using Amazon Mechanical Turk, in which people viewed Web pages accompanied by ads. The amount of time the ads were in view was either determined endogenously (as a function of reading speed) or exogenously (as a function of a timer and random assignment). Our results suggest that for exposure times of up to one minute, there is a strong, causal influence of exposure time on ad recognition and recall, with the marginal effects diminishing at durations beyond this level. Simple models describing memory response as a function of the logarithm of exposure time provide a good fit. In addition, we find that advertisements that are displayed when the Web page loads attain greater marginal increases in recognition per unit time than do ads that come into view second in a sequence. Nonetheless, for both types of ads, exposure time has a substantial effect. A psychologically-informed accounting system based on ad exposure duration, sequence and onset time may more closely align with advertiser goals than the industry standard of impression-based accounting.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Economics

General Terms

Economics, Experimentation

Keywords

advertising, memory, display, recall, recognition, time, exposure

1. INTRODUCTION

Internet display advertising is a business that generates over $20 billion in annual revenue [8] from advertisers paying to place graphical ads on publisher's websites. Customarily, advertisers pay for a certain number of impressions, where one impression is one download of the ad. These impressions are typically sold in "guaranteed contracts", which specify a number of impressions and a set of targeting variables, for example 10 million impressions on automobile-related pages sold to women between 30 and 50 years of age, or 7 million impressions targeted to people interested in women's fashion. In addition, display ads are also sold in exchanges like Google's DoubleClick exchange (GDC) or Yahoo!'s Right Media exchange (RMX). In these exchanges, display ads can be individually targeted to a specific user, such that when a targeted user clicks on a page, the exchange sells the right to show an ad to that particular user as that page loads.

Since pricing based on thousands of impressions delivered was familiar to advertisers who bought and sold newspaper ads, Internet display advertising may be sold on an impression basis for reasons that are largely historical. Newspapers and early Web sites were both fairly static, making impression-based pricing straightforward. However, the rapid adoption of AJAX and similar technologies are making the practice of tying impressions to the loading of a page obsolete. For example, popular email programs like Gmail and Yahoo! Mail update page content as new mail comes in, without any action by the user. When a page automatically updates its content in a continuous fashion, without changing its URL, it becomes difficult to define what it means for the page to be replaced and consequently what constitutes an impression.

One alternative to selling display ads based on impressions is to sell them based on time, much like how television and radio ads are traditionally sold. In fact, it has already been shown that increasing the duration of a television ad slightly increases the probability the ad will be remembered in an aided recall task [11]. However, advertising on the web is different than advertising on TV and radio because ads on the web are usually placed alongside the content whereas ads on TV and radio appear instead of the content. As a result, many Internet users claim to ignore ads completely, and studies around the concept of "banner blindness" frequently suggest that they do [1]. It is also possible that just seeing

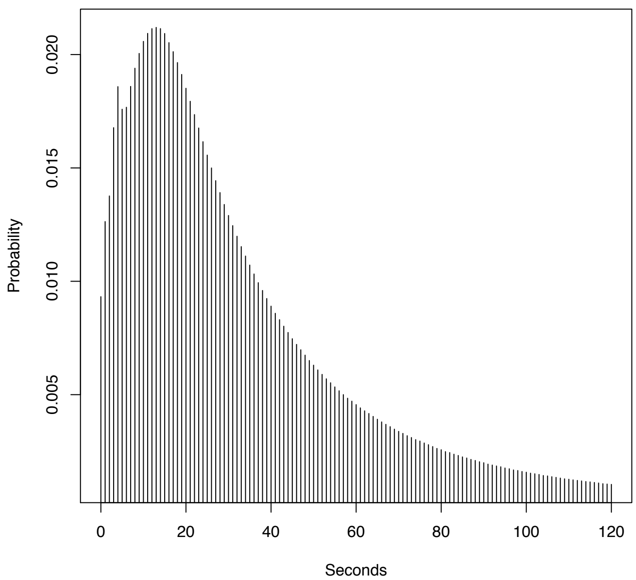


Figure 1: Distribution of time spent on a popular Yahoo! site

an advertisement produces recognition and recall, but that continued exposure does not lead to substantial increases in these measures. Indeed, if users initially scan the page, taking in ads, and then proceeded to focus narrowly on the page content, the duration of exposure would not matter much, beyond the initial few seconds. For these reasons, it is not clear *a priori* that ad exposure time will have an effect on memory, and if it does, what the magnitude of the effect will be.

Measuring the amount of time a user spends on a web page, which is an upper bound on how much time that user is exposed to an ad, is straightforward: simple Javascript code can track the amount of time a page is open, or better still, remains in focus. The result of such measurements is illustrated in Figure 1 for a popular Yahoo! property. Time spent ranges from a couple of seconds to well over two minutes. In practice, one would need to cap the amount of time spent on the page to use it as a measure of ad exposure time. For example, the person who spends 10 hours on a page may not even be twice as likely to recall an ad as the person who spends ten seconds.

Should advertisements be sold by impressions or by ad exposure time? Or more fundamentally, does amount of ad exposure time actually matter? Theoretically, the right metric is the one that most closely approximates advertiser value, because this induces the efficient use of resources [6]. Advertisers primarily judge campaigns by two criteria: ad recall and ad recognition. Thus the central focus of this work is to study if and how exposure time affects memory for advertisements.

2. RELATED WORK

Dreze and Husshe rr [4] ran a two-part study on the effectiveness of banner advertisements, which are long skinny ads that often appear at the top of a web page. In the first part, they used eye-tracking technology to show that users focus directly on banner ads rather infrequently, and argued that since those who do not directly focus on a banner ad are unlikely to click on it, click-through rate is not a good metric of ad effectiveness. In the second part of their study, users did simple tasks on web pages—such as typing a word into a box—with a banner ads present. Twenty-four hours after the task was completed, the subjects were asked unaided and aided recall questions, as well as various recognition and awareness questions. The authors found, for example, an 11% unaided recall rate, a 30% aided recall rate and a 19% aided recognition rate. As a result, Dreze and Husshe rr argue that brand awareness measures, such as recall and recognition, are more appropriate than clicks as metrics for display ads. In what follows, we shall use the metrics these authors advocate to measure the effect of display ad exposure time.

Perhaps the work most closely related to ours is a two-part study by Danaher and Mullarkey [3]. In the first part of their study, participants looked at a series of five web pages for 20, 40, or 60 seconds each. Crucially, the participants were not allowed to navigate away from the page before the assigned time expired. Participants were next asked recognition and recall questions. In the second part of the study, the same participants navigated a small, closed set of web pages containing advertisements for a fixed duration (again without the option to end the task when they finished reading), after which they were again asked questions about ad recognition and recall.

While this study found that higher exposure time was correlated with increased recognition and recall, there are a few design issues that would prevent drawing strong conclusions, especially causal ones. In the first part of the study, participants who finished the text quickly had more unoccupied time to look around the page and notice the ads. Because of this, differences in reading speed could account for differences in memory for ads in a way that would not arise during natural Web browsing. Furthermore, this design should disproportionately increase the recall and recognition rate of the longer duration treatments in which more people would have more "dead" time to look at ads.

In addition, Danaher and Mullarkey presented participants with memory questions for ads between the two parts of their study. After having seen questions about ads at the end of the first part of the study, participants would have payed greater attention to the ads in the second party of the study, thereby compromising the second set of tests. Lastly, since subjects could surf among a limited set of pages in the second part of the study, the number of ad impressions delivered was confounded with the time of exposure, leaving open the question of whether impressions or time of exposure matters more for predicting memory. To improve upon past investigations, in what follows we: manipulate the time of ad exposure exogenously, eliminate the confounding effect of reading speed, give no advance warning that memory for ads will be tested, and hold the number of impressions delivered constant.

3. RELATING MEMORY TO EXPOSURE TIME

Both of the experiments described in this paper were conducted on Amazon's Mechanical Turk[1], a crowdsourcing website where requesters can post jobs, called human in-

[1] `http://www.mturk.com`

telligence tasks (HITs), and workers can do those HITs for pay. After a worker submits their work for a HIT, requesters review the work and either accept it or reject it. Mechanical Turk was originally built for humans to do tasks that are hard for machines to do such as extracting data from images, audio transcription and filtering adult content. We used Mechanical Turk by posting our experiment as a HIT and using workers as our experimental participants. (There is a burgeoning literature that shows that the behavior of Mechanical Turk workers is comparable to laboratory studies [10, 7, 12].) Throughout this experiment, we used the Mechanical Turk API to restrict our subject pool to participants who live in the United States and who have a high approval rating (over 90%). By tracking the IDs of the workers who accepted our HITs, we could enforce that participants were only allowed to participate in one of two experiments, and they were only allowed to do that experiment one time. Next, we describe the results of our first experiment, which formed the basis for our second experiment, which exhibits the main results of this paper.

3.1 Experiment One: Time of Exposure Varied Endogenously

The first experiment aimed to establish a correlation between endogenous (self-determined) time spent on a web page and memory for a display ad on that page. Participants were recruited from Mechanical Turk by means of a HIT offering a fixed payment in addition to a smaller payment per question answered (to avoid the possibility of collusion, payment was not based on the correctness of answers). The preview page for the HIT consisted of a consent form and instructions indicating that the task involved reading and answering questions about a Web page.

In order to spread out the distribution of reading times, participants were assigned to conditions in which the article was either short or long. Randomization occurred at the moment the HIT was undertaken, precluding any confound between time of day and condition assignment. In addition to article length, random assignment also governed which of two display ads (for either Netflix or Jeep) and which of two articles (about either schoolteachers or an oil spill) would be shown. Thus, this experiment had a $2 \times 2 \times 2$ design in which there were two treatments each for the subject of the article, the length of the article, and the ad shown next to the article. Manipulation of ad and article was undertaken towards the aim of generalizable results under the methodological principle of representative design [2].

In the task, participants were shown high-resolution screen captures of web pages, as shown in Figure 2. In all conditions, the assigned display ad appeared next to the text. The image of each web page was chosen to be 600 pixels high. This height was chosen because 99% of screens on the Web are able to show at least 600 pixels in height[2] ensuring that the user could see the pages and the display ad, in their entirety without scrolling, and that the display ad was always in view.

After reading the first page of their assigned article, participants clicked a button to continue on to a second page, which featured the same display ad as the first page. In the short condition, the second page contained just one sentence of text, while in the long condition the text extended to the bottom of the second page. After the participants read both pages at their own pace, they proceeded on to the questionnaire and were unable to direct the browser to display the ad or article again.

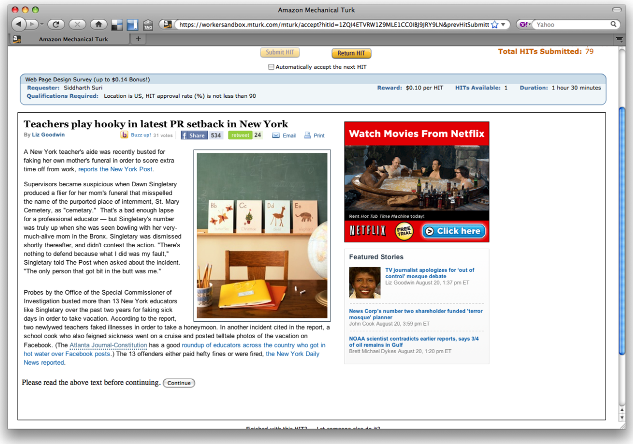


Figure 2: Screenshot of the first page of the New York public school teachers article with the Netflix ad.

The questionnaire first presented two, five-alternative multiple choice reading comprehension questions to assess to what extent participants paid attention to the text. After that, it contained one unaided recall question: "Which advertisements, if any, did you see on the page during the experiment? Type the name of any advertisers here if you can remember seeing their ads on the last page, or indicate that you are unable to remember any." On the subsequent page, participants were asked "Did you see the following advertisement during this HIT?", accompanied by a picture of the Netflix ad, and were next presented the same question for the Jeep ad. See Appendix A for pictures of the ads. Note that participants were shown only one of these two ads in the experiment, making the other ad a lure which is useful for establishing the false positive recall and recognition rates. After submitting, the task was completed and answers could not be changed.

Before proceeding with the analysis, the initial sample of 479 was reduced to the inner 98% of reading times to eliminate outliers, reducing the sample by 10, after which 16 participants (4.6%) were excluded for answering both reading comprehension questions incorrectly, leaving 453 participants with a mean reading time of 123 seconds. The length manipulation was effective in spreading out reading times as reflected by the per-condition means (143 and 103 in the long and short conditions) and the interquartile ranges: the 75th percentiles were 181 and 125 seconds, and the 25th percentiles were 96 and 72 seconds in the long and short conditions, respectively.

Each participants' questionnaire responses were coded as four binary items. The first two binary items coded whether the unaided recall question mentioned the two advertisements. The latter two binary items coded the answer to the pictorial recognition task.

3.2 Experiment One Results

Figure 3 shows the result of Experiment 1. The false pos-

[2]See http://www.w3schools.com/browsers/browsers_display.asp for the distribution of screen sizes on the web.

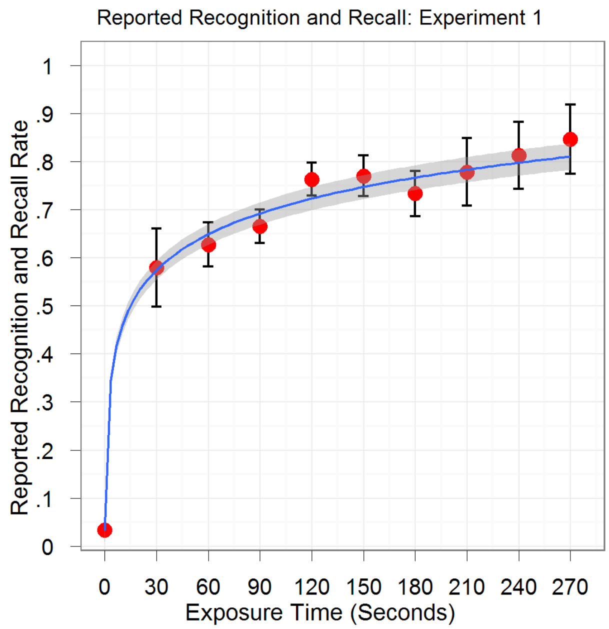


Figure 3: Memory as a function of reading time. Observe the steep increase in the 0–60 second range followed by the slower, but sustained increase after 60 seconds. The points are the mean of the recognition and recall items binned at 30 second intervals with error bars at ± 1 standard error. The blue curve is the best fitting one in its family, with the gray region showing a 95% confidence interval over its parameters.

itive data (claims of remembering the "lure" ad which was not shown) make up the point on the x-axis at 0 seconds of exposure time. The rest of the data contribute to points on the x-axis with an exposure time strictly greater than 0 seconds. Figure 3 also shows the result of fitting a curve of the form $y = a + b\log(x + c)$ to the data along with the 95% confidence interval, where x is the reading time and a, b and c are free parameters. Previous psychological studies on memory have found that memory rates are well described by functions of this form [5]. To indicate the quality of the fit, we also plot the average rate of recognition and recall binned at 30 second intervals. There were 10 responses that took upwards of 270 seconds, which were excluded from the plot owing to sparsity of data within each bin. Note the standard errors are smallest where data are most plentiful, in the region around 120 seconds.

Figures 4(a) and 4(b) are the same type of plots except with the recall and recognition data separated. The same functional form is fit with different parameters. Again, to indicate the quality of the fit, we also plot the mean recognition or recall rate. Since these plot each have half of the data shown in Figure 3 the mean recall and recognition rates are binned in 60 second intervals.

There are a few trends that emerge from examination of Figures 3 and 4. First, if one compares Figure 4(a) to Figure 4(b), one can easily see that unaided recall is harder then recognition yet the unaided recall rate is still reasonably high over a long range of exposure times: roughly 50% or more with 60 seconds or more of exposure. False positive rates are less than 10% for recognition measures and, as would be expected, 0% for the recall measures; this gives some indication that the reported memory rates are reasonably accurate, though false recognition rates always depend on the properties of the ad in question. In the next study we use lures that deliberately resemble the targets to approximate an upper bound on false recognition for the targets. Since the binned means lie so close to the curve that was fit to the data one can easily see that the logarithmic function described above describes the data well. Finally, for both the recall, recognition, and combined measures there is a sharp increase from the false positive rate at 0 seconds to 30 seconds. This suggests that most of an ad's effect on memory is likely to take effect soon after the page loads.

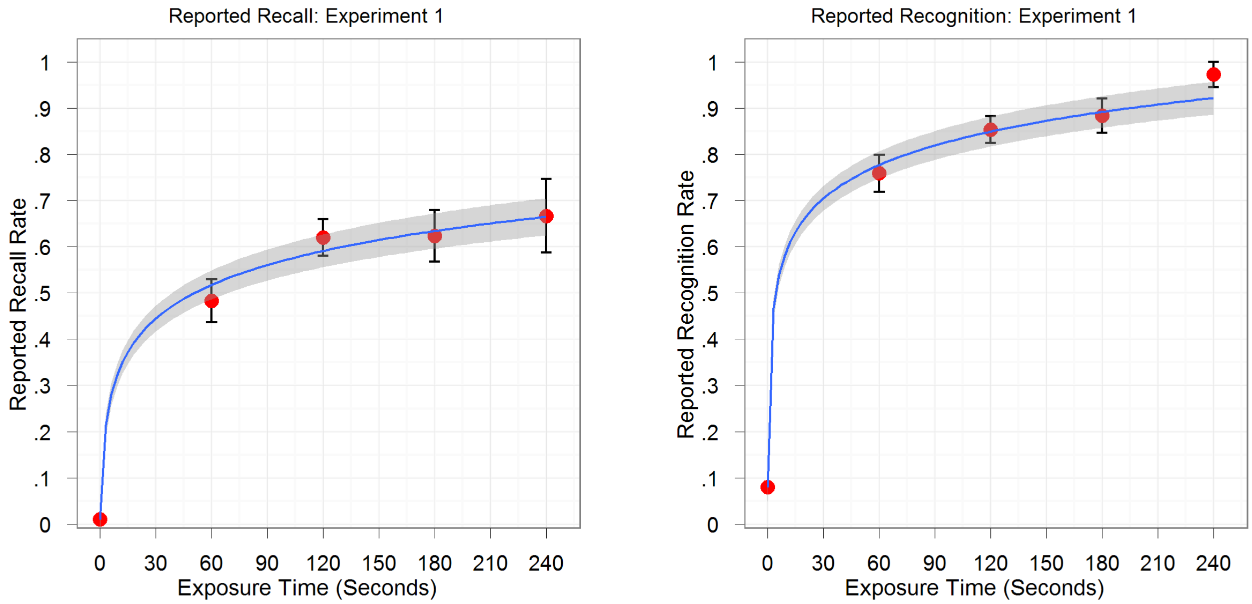
In the experiment described previously the exposure time of the advertisement was determined by the participants' reading time. Figures 3 and 4 show that the more time the participant spent reading, the higher the recall and recognition rates. Because this experimental manipulation of ad exposure depends on a participant's reading speed, one cannot establish a causal link between exposure time and memory. For example, it could be the case that slower readers are more careful individuals who examine the entire page, including the text and the display ad, more closely than faster readers. Or there could be a latent variable that causes both slow reading and good memory in select individuals. Such a variable could create the appearance of a relationship between exposure time and memory across individuals, even though it may not exist within any homogeneous set of individuals. Were this the case, our result would be of little interest to advertisers who wish to target homogeneous segments.

Experiment 2 will address the above concerns by manipulating exposure time exogenously. Because there was a steep rise in the recall and recognition rates in the 0–60 second range we will focus our attention there.

3.3 Experiment Two: Exogenously Varying Time of Ad Exposure

Whereas in the previous study, ad exposure time was roughly equivalent to reading time, this experiment manipulates exposure time exogenously to estimate the causal effect of ad exposure time on memory. This external manipulation was accomplished by using a Javascript timer to display ads for a pre-determined number of seconds while people are reading and then having them disappear from view. To make the scenario as realistic as possible, when the first ad shown disappears, it is immediately replaced with a second ad, as is the case on many Web sites that rotate multiple ads into one slot.

Participants were again drawn from Amazon Mechanical Turk as in Experiment 1. Instructions again indicated that the task involved reading and answering questions about a Web page. Participants were again paid a flat rate for completing the experiment, plus a specified amount per question answered. The article used is depicted in Figure 2. Participants were randomly assigned into one of four treatments which governed how long they were exposed to two display ads (one for Jeep and one for Netflix). Simultaneous with the loading of the article, one ad was displayed, depending on condition, for either 5, 10, 25, or 40 seconds. The 5 and



(a) (b)

Figure 4: The recall and recognition rates, respectively, as a function of reading time. Both show the steep increase in the 0–60 second range followed by the slower, but sustained increase after 60 seconds. The points are the mean of the recognition and recall items binned at 60 second intervals with error bars at $\pm$ 1 standard error. The blue curve is the best fitting one in its family, with the gray region showing a 95% confidence interval over its parameters.

25, as well as the 10 and 40 second exposure times, were paired in the following manner. If the first ad was shown for 5 seconds, then the second ad was shown for 25 seconds; conversely, if the first ad was shown for 25 seconds, then the second ad was shown for 5 seconds. Similarly, if the first ad was shown for 10 seconds, then the second ad was shown for 40 seconds; conversely, if the first ad was shown for 40 seconds, then the second ad was shown for 10 seconds. After the display time for the second ad expired, it was replaced with a white rectangle, giving the appearance of it disappearing altogether. Data collection proceeded in two waves, one comprising 300 subjects that randomly assigned participants to the 10 and 40 second treatments, and the latter one randomly assigning 250 individuals to the 5 and 25 second treatments. Within the four conditions, participants were randomly assigned to see the Netflix ad before the Jeep ad or *vice versa*, resulting in a 4×2 design (four exposure time sequences by two orderings of the advertisements).

To stabilize forgetting time, which is the amount of time between the first ad disappearing and the questionnaire, the reading task was followed by a buffer task in which participants played a game for a pre-determined, condition-specific amount of time. Tetris, a visual game consisting of primary shapes, was chosen to avoid ad-specific linguistic memory interference. The game was rendered in black and white to reduce visual memory interference with the colors present in the stimuli ads. The duration of this buffer task was calculated to equalize forgetting time for the average participant in each condition and ranged between 20 and 55 seconds. After the buffer task time expired, participants were automatically forwarded to the questionnaire, and were not able to exit the game early.

In the questionnaire phase, the same two, five-alternative multiple choice reading comprehension questions were asked followed by the same unaided recall question as in Experiment 1, which asked: "Which advertisements, if any, did you see on the page during this HIT? Type the name of any advertisers here if you can remember seeing their ads on the last page, or indicate that you are unable to remember any." The subsequent page presented four recognition questions with textual cues of the form "Did you see a ___ ad?", with Netflix, Jeep, American Express and Avis being the advertisers filling in the blank. Netflix and Jeep were the "target" advertisers whose ads were indeed shown, while American Express and Avis were "lure" advertisers, whose advertisements were not shown. For pictures of the ads and their respective lures see Appendix A. Inclusion of lures was particularly relevant for the next four questions, which were recognition questions with pictorial cues asking "Did you see this ad?" and showing the actual two ads from the experiment as well as the two lure ads. The lures were chosen for bearing a strong visual resemblance to the targets in order to approximate an upper bound on the false positive rate. The Avis lure is primarily red, much like the Netflix ad, and the American Express lure is primarily black, much like the Jeep ad.

Each participant's data was coded as twelve binary responses. The first four responses coded mentions of the two target ads and lures from the unaided recall question. The next four were the recognition questions with textual cues, and the final four were the recognition items with pictorial cues (the ads themselves).

Before analyzing the effect of exposure on the memory items, 14 participants (2.5%) were excluded for not completing the task, after which 7 (1.3%) were excluded for answering both reading comprehension questions incorrectly. The very small percentages of incomplete or careless responses speak to the quality of data which can be collected via Mechanical Turk.

3.4 Experiment Two Results

Experiment 1 showed a correlation between exposure time and memory for advertisements. Will a similar relationship arise when ad exposure time is manipulated exogenously?

Based on reading time estimates from Experiment 1, we designed this experiment such that about 80% of people would take long enough reading the article that they would be exposed to both ads for the prescribed amounts of time. This estimate turned out to quite accurate as the fastest 20% of readers finished in 51 seconds, and only 50 seconds were needed to see both ads in the longest condition (in which two ads are shown for 10 and 40 seconds). The median reading time was 77 seconds, the 75th percentile was 98 seconds and the 25th percentile was 55 seconds. The purest exogenous analysis would exclude the fastest 20% of readers from all conditions to ensure the sample consists only of those who saw both first and second ads for the intended amount of time. However, since even fast readers were likely exposed to at least one ad for the intended amount of time, and since only three out of 12 questions attributable to a participant are tied to each target ad, one may simply drop the questions (as opposed to the participants) for which full intended treatment was not received and thus retain 96%, as opposed to 80%, of the observations. Doing so may introduce a slight bias, so as a preventive measure, all analyses were run dropping the fastest 20% of individuals, dropping the affected 4% of observations for which the full treatment was not received, or doing neither. The effect of such filters on derived estimates was negligible. For example, in Figure 5 after the application of such filters, the average point estimate moved by less than one percentage point, and no estimate moved by as many as two percentage points. Accordingly, in the analyses which follow, we retain the 96% of items for which the intended level of exposure was received.

As shown in Figure 5, there is a clear increasing trend in the probability of reporting memory of an advertisement as a function of manipulated exposure time. The combined recall and recognition measure increases sharply in the first 10 seconds and then still increases, albeit at a slower rate after that. Thus there are diminishing returns to increased exposure time suggesting that the first seconds of exposure are more valuable to advertisers then later seconds, a topic we will revisit in Section 3.5. Nonetheless, increased exposure time increased the probability that the ad was remembered.

The best fitting curve took on the form $y = -.009 + 0.106 \log(x + 1.681)$, where y is the memory response and x is the time of exposure. The basic shape of this curve, which shows the causal effect of exposure time on memory, can shed light on the results of Experiment 1. Substituting in the mean reading times from Experiment 1's two conditions, 143 and 103 seconds, this curve predicts a difference of 3.4 percentage points, a modest difference, not unlike that realized in this range in Experiment 1. While the relation in

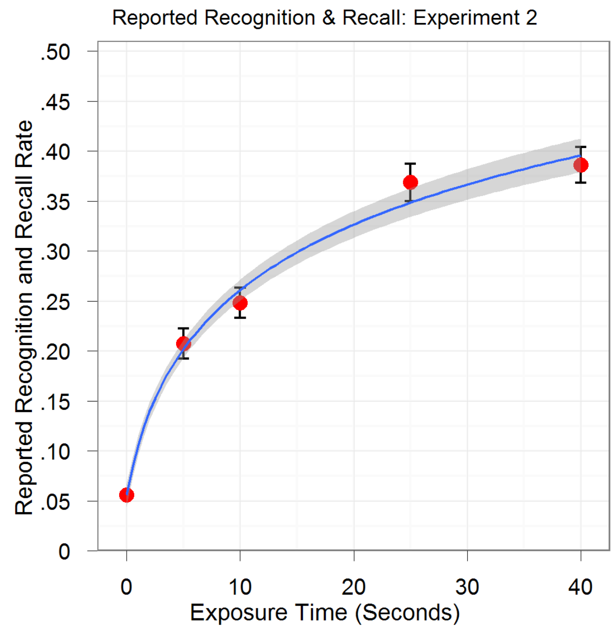


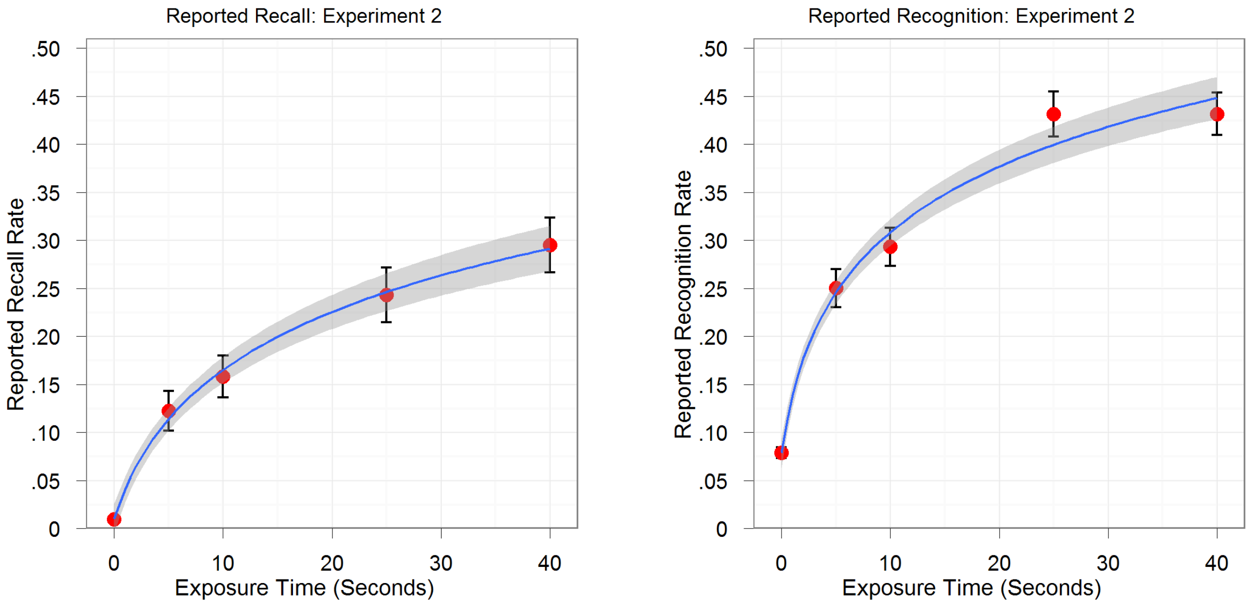
Figure 5: Memory as a function of time of exposure time which was exogenously imposed on the subjects. The points are the mean of the recognition and recall items with error bars at ± 1 standard error. The blue curve is the best fitting one in its family, with the gray region showing a 95% confidence interval over its parameters.

Figure 5 does not make point predictions about Experiment 1—the two curves occupy different ranges, possibly due to the buffer task in Experiment 2 and the two impressions delivered in Experiment 1 versus one impression delivered in Experiment 2—the same basic logarithmic relationship predicts that large relative differences in recall and recognition are difficult to attain beyond exposure times of 100 seconds. For this reason, and for the reason that most Web pages are viewed for less than this long (see again Figure 1), a prudent accounting system may discount or disregard exposure times beyond this threshold.

As in Experiment 1, decomposing the aggregate curve into recall and recognition measures shows relationships of the same basic shape, with recognition, as is most often the case [3], proving to be an easier task than recall. Recall and recognition are shown separately in Figure 6. Despite the lures bearing a strong resemblance to the target ads, false alarm rates were low: 6% on text recognition items, and reaching a maximum of 7.8% for the pictorial recognition items. As expected, the false alarm rate for recall items was 0%. A logistic random effects model, with participants and advertisers as random effects, confirmed that both both time of exposure and presentation order to be strong predictors of what is remembered, as treated in the next section.

3.5 The Effects of an Advertisement Appearing First or Second

Next we discuss the effect of showing an ad first versus second. Recall that users were exposed to an ad for 10 sec-



(a) (b)

Figure 6: The recall and recognition rates, respectively, as a function of the exposure time which was exogenously imposed on the subjects. The points are the mean of the recognition and recall items with error bars at $\pm$ 1 standard error. The blue curve is the best fitting one in its family, with the gray region showing a 95% confidence interval over its parameters.

onds then another for 40 seconds, or *vice versa*, or they were shown an ad for 5 seconds then another for 25 seconds, or *vice versa*. One might expect that due to their sudden appearance on the page, ads which appear second might have an attentional advantage [9] over ads that load with the rest of the page's content. However, the opposite appears to be the case. Figure 7 shows the result of plotting the combined recall and recognition for the first and second ads. Observe that the recall and recognition rate for the second ad does not have the steep increase at low exposure times as first ads do. Moreover, the combined recognition and recall rate for the second ad displayed for 40 seconds was still 30%.

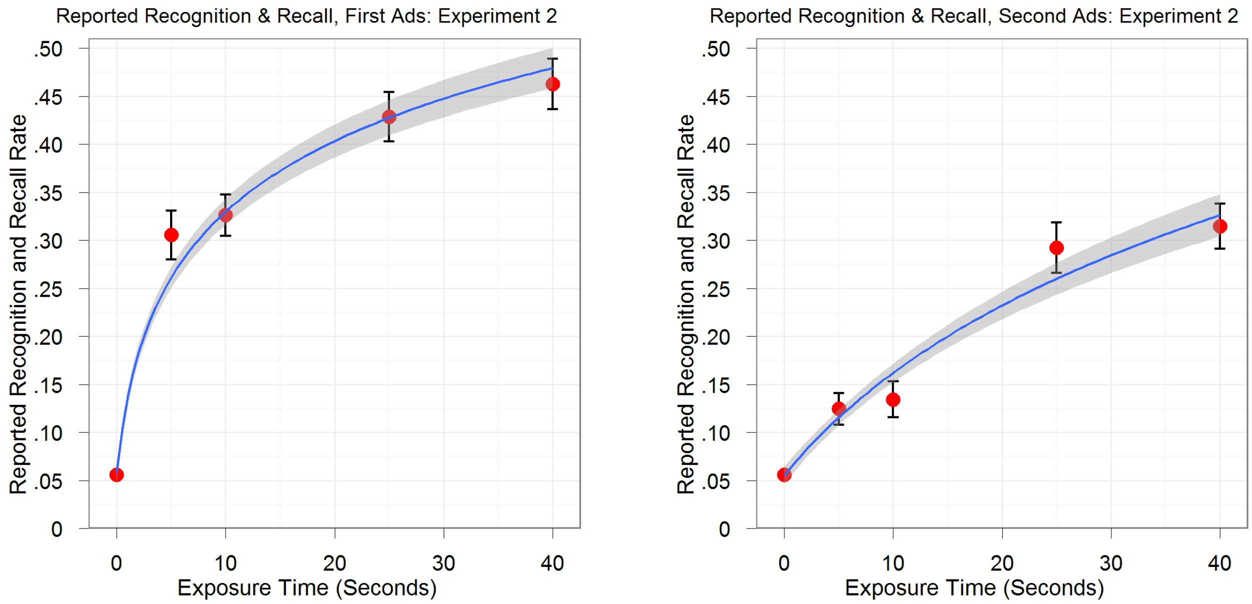
Figure 7(a) shows that the seconds soon after the page loads cause a higher increase in recall and recognition than do later seconds. Consider a 10 second ad shown second, which appeared 40 seconds after the page loaded. Also consider a 5 second ad shown second, which appeared after 25 seconds. The 5 second ad will be seen during a higher recall period of time than the 10 second ad. Thus, in our design, the recall of the 5 second ad will be slightly inflated relative to that of the 10 second ad. One can make a symmetric argument which shows that the recall and recognition of the 25 second ad will also be slightly inflated relative to a 40 second ad shown second. Despite this observation, Figure 7 shows the first ad has a higher recall and recognition rate then the second. Figure 7 also shows that 5–10 seconds of exposure of the first ad shown is roughly equivalent to 40 seconds of exposure of the second ad shown. From an advertiser's perspective, having an ad shown second is less valuable the being shown first. Nonetheless, in either case, increasing the exposure time adds value.

4. CONCLUSION

The main result of this work is giving evidence for a causal effect of exposure time on the recall and recognition of display advertisements. More specifically, we showed that display ad recognition and recall increase as exposure time increases, with the steepest effect occurring for low values of exposure time. A log-linear function is a good approximation of the psychological response. The secondary result of this work is that only the first ad shown in a sequence benefits from this steep increase of recognition and recall in the first seconds of exposure, but that both the first and second ads shown benefit from more exposure time.

As mentioned in the introduction, display ads are currently sold by the impression. This pricing scheme ignores how long ads are in view. Advertisers value ad recognition and ad recall, and time of exposure causally impacts both. Pricing based on better measures of advertiser value than impressions offers three distinct advantages. First, it permits efficient allocation. When pricing is based on an average, allocating pieces to different buyers is impossible, since the pieces have the same average value. Similarly, with noisy measures of value, more allocation mistakes will be made. Second, pricing based more closely on value mitigates buyer risk, since pricing corresponds more closely to the value actually delivered. Risk matters most in environments with great supply variability like digital advertising. Finally, pricing based on value permits price discrimination—charging advertisers based on the value delivered may let the seller capture a larger fraction of the value generated.

Thus this work provides evidence that publishers should charge based on the amount of time users spend with an ad



(a) (b)

Figure 7: The combined recognition and recall rates for the first and seconds ads shown. Observe that the first ad shown has a steep increase in memory in the first few seconds and the second ad shown does not. Yet both show that increasing exposure time increases memory. The points are the mean of the recognition and recall items with error bars at ± 1 standard error. The blue curve is the best fitting one in its family, with the gray region showing a 95% confidence interval over its parameters.

in view. There are two obvious systems by which publishers could achieve this. First, publishers could simply measure how long a user spends with an ad in view, then charge based on a functional transformation of this. A second method could involve advertisers buying 30-second "spots", for example. In this system, even as the user navigated the site, the ad would stay in view for 30 seconds. Moreover, if a user leaves the site before the 30 seconds has expired, publishers could either serve the ad to the user when he or she returns, or prorate the purchase of that spot for the amount of time the user actually spent with the ad in view. Publishers could differentially price spots of different lengths. A clear direction for future work is to analyze theoretical pricing models of systems that take the time of exposure into account.

Acknowledgments

We would like to thank Randall A. Lewis, David H. Reiley and Duncan J. Watts for many helpful conversations.

5. REFERENCES

[1] Jan P. Benway. Banner blindness: The irony of attention grabbing on the world wide web. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, volume 1, pages 463–467, October 1998.

[2] Egon Brunswik. *Perception and the representative design of experiments (2nd ed.)*. University of California Press, Berkeley, CA, 1956.

[3] Peter J. Danaher and Guy W. Mullarkey. Factors affecting online advertising recall: A study of students. *Journal of Advertising Research*, 43:252–267, 2003.

[4] Xavier Drèze and François-Xavier Hussherr. Internet advertising: Is anybody watching? *Journal of Interactive Marketing*, 17(4), 2003.

[5] Fernand Gobet and Herbert A. Simon. Five seconds or sixty? presentation time in expert memory. *Cognitive Science*, 24(4):651–682, 2000.

[6] Bengt Holmstrom and Paul Milgrom. The firm as an incentive system. *American Economic Review*, 84(4):972–991, September 1994.

[7] John J. Horton, David G. Rand, and Richard J. Zeckhauser. The online laboratory. *Experimental Economics*, 2011. To Appear.

[8] IAB internet advertising revenue report, 2009.

[9] John Jonides and Steve Yantis. Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, 43(3):346–354, 1988.

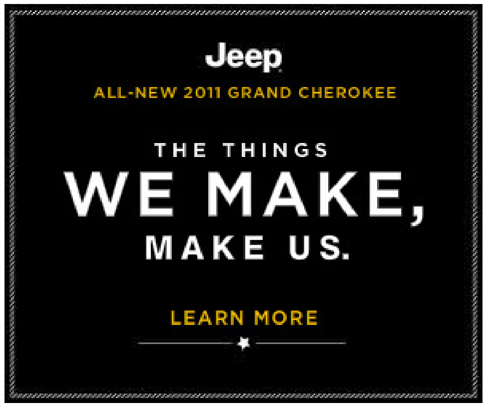
[10] Gabriele Paolacci, Jesse Chandler, and Panagiotis G. Ipeirotis. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5:411–419, 2010.

[11] Gordon L. Patzer. Multiple dimensionsof performance for 30-second and 15-second commercials. *Journal of Advertising Research*, 31(4):18–25, 1991.

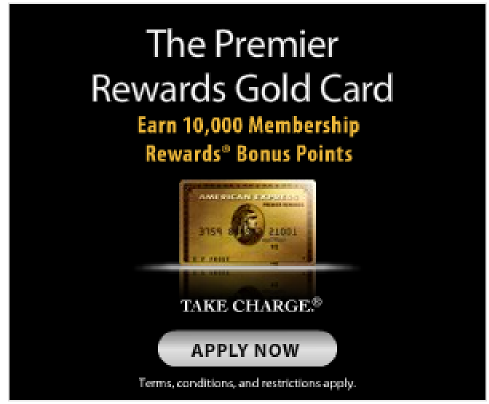
[12] Siddharth Suri and Duncan J. Watts. Cooperation and contagion in networked public goods experiments. *PLoS ONE*, 2011. To Appear.

APPENDIX

A. ADVERTISEMENTS AND LURES



(a)



(b)

(c)

(d)

Figure 8: Panels 8(a) and 8(c) are the ads used in the experiment with endogenous exposure time (Experiment 1). Panel 8(b) was the lure for 8(a), and Panel 8(d) was the lure for 8(c) in the experiment with exogenously imposed exposure time (Experiment 2).