Improving the Effectiveness of Time-Based Display Advertising (Extended Abstract)*

Daniel G. Goldstein
Microsoft Research
New York, NY, USA
dgg@microsoft.com

R. Preston McAfee
Google Research
Mountain View, CA, USA
mcafeer@google.com

Siddharth Suri
Microsoft Research
New York, NY, USA
suri@microsoft.com

Abstract

CPM or cost per thousand impressions is the prevalent metric used for selling online display ads. In previous work, we have shown that the exposure duration of an ad has strong effects on the likelihood of an ad being remembered [Goldstein *et al.*, 2011], with the first seconds of exposure having the greatest impact on memory. Because an ad pricing metric that is based on both time and impressions should be more exact than one based on impressions alone, the industry has good reasons to move towards time-based advertising. We address the following unanswered question: how should time-based ads be scheduled? We test and present one schedule that leads to greater total recollection, which advertisers want, and increased revenue, which publishers want. First, we find that presenting two short, successive ads results in more total recollection than presenting one longer ad of twice the duration. Second, we show that this effect disappears as the duration of these ads increases. Together, these findings suggest a form of time-based ad pricing that should appeal to advertisers and publishers alike.

1 Introduction

Online display advertisements are pictorial ads that advertisers pay publishers to run for sums that amount to billions of dollars annually [PricewaterhouseCoopers, 2011]. With display ads, advertisers seek to increase brand recognition, brand awareness [Drèze and Hussherr, 2003], and of course sales [Lewis and Reiley, 2011; Manchanda *et al.*, 2006]. For nearly a century, advertisers have measured the effectiveness of brand advertising using the proxy of memory metrics [Starch, 1923; Wells, 2000]. To date, display ads have tended to be sold on CPM (cost per thousand impression) terms, where an impression is simply one appearance of an ad.

In previous work [Goldstein *et al.*, 2011], we demonstrated a causal effect of display time (the time an ad is in view) on the probability that a user will remember it. In addition, we found steep increases in the probability of remembering an ad for exposure times of up to roughly 40 seconds, followed by a tapering off. Together, these findings suggest that exposure time combined with impressions may constitute a better metric for pricing advertising than impressions alone, since exposure time causally influences, and more exactly leads to the recognition and recall that display advertisers seek.

In this paper, we consider how a publisher might schedule time-based ads to increase the total effect on memory per unit of time. Specifically, when scheduling the ads that load with a page, we ask if it is better to show one ad with a long exposure time or two successive ads with half the exposure time each. Having one ad last a long time may increase the chance that a user notices it. However, having two shorter ads gives users more ads to notice. Without an experiment, it is not clear which would result in more overall recollection.

We conduct an experiment in which people are instructed to read an online news article, next to which ads are displayed according to one of four schedules. Imagine two impressions, shown to two users u_1 and u_2, which last for $2t$ seconds each. Suppose that, under impression-based advertising, u_1 sees ad A and u_2 sees ad B. A time-based alternative is for u_1 to see ad A for t seconds followed by ad B for t seconds, and u_2 to see ad B for t seconds followed by ad A for t seconds. We aim to uncover which of these two schedules would be better for advertisers A and B in terms of recollection.

2 Methods

Participants were 1,100 U.S. workers from Amazon's Mechanical Turk participant pool who had with approval ratings of 90% or more and were allowed to participate one time. We here describe the format of the experiment and the various treatments to which the participants were randomly assigned.

2.1 Experimental Design

Payment to participants consisted of a 50 cent flat rate to start the experiment plus a 10 cent bonus for each question answered. Participants decided to join on the basis of a preview page, which consisted of a brief consent form and a description of the task: reading a web page and answering questions about it.

After consenting, participants were shown an image of a news story from an actual Yahoo! website. Images were used

*The paper on which this extended abstract is based was the recipient of the best paper award of the 2012 ACM Conference on Electronic Commerce (EC'12) [Goldstein *et al.*, 2012].

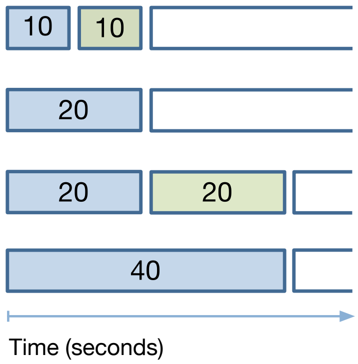

Figure 1: The four time treatments in which participants were randomly placed. Each colored rectangle represents an ad with the number of seconds it was in view. The white rectangles on the right side of the figure indicate the absence of an ad.

to prevent participants from clicking on links or ads and navigating away. The news story image comprised text and graphics with a display ad in the sidebar. The goal of this research is to compare the recollection of two short ads to that of one longer one. Thus, for each memory measure we investigate, we compared the sum of the metric for the two short ads to the metric measured on one long ad plus the false positive rate.

We define the following notational convenience to formally describe the experimental treatments.

Definition 1. *If X, Y, Z are (not necessarily different) ads and u is a user, let* $\Pr(X|YZ) =$
$\Pr(u$ *remembers* $X \mid Y$ *shown for* t *secs then* Z *for* t *secs*$)$.

After the duration of ad Z expired it was replaced by whitespace. If $Y = Z$ then the above definition describes a situation where the same ad is shown continuously for $2t$ seconds. Thus our experiment will measure and compare $\Pr(A|AA) + \Pr(A|BB)$ with $\Pr(A|AB) + \Pr(A|BA)$. The term $\Pr(A|BB)$ may be considered the false positive rate where one claims to remember ads that were not displayed perhaps due to having seen the ad elsewhere. Observe that this comparison holds the total time the two ads are in view constant. It also holds the within-impression timing constant. That is, in both schedules, each ad is shown from 0 to t seconds, and from t to $2t$ seconds exactly once. If $\Pr(A|AB) + \Pr(A|BA)$ exceeds $\Pr(A|AA) + \Pr(A|BB)$ then publishers should split time slots between two advertisers to improve recollection. To test this, we used one pair of time treatments where $t = 10$ and another pair of time treatments where $t = 20$ resulting in four time treatments overall. Figure 1 gives a pictorial representation of the time treatments.

The treatments with two short ads necessarily involve two advertisements, and of course different advertisements may be differentially memorable. To hold all of this constant, the simplest test involves two orderings of the short ads. Denote the treatment that shows ad A followed by ad B as AB. Then the simplest test is to compare the effectiveness of AB for one user and BA for a second user, to AA for a third user and BB for a fourth. In the interests of external validity we used four different ads as stimuli. As a result, the first ad treatment had a Netflix ad shown first and a Jeep ad shown second; the second ad treatment had the opposite order. The third ad treatment had an Avis ad shown first and an American Express ad shown second, and the fourth ad treatment had the opposite order. In "single ad" time treatments, in which only one ad was shown, the second ads in the order described above were left out. In all, the four time treatments and the four ad treatments yielded a 4×4, between subjects design. Subjects were randomly placed into one of these 16 treatments at the point of accepting the experiment to avoid any confound between dropping out of the experiment and the treatment assigned.

After participants finished reading the article at their own pace they clicked a link and were taken to a page where they played Tetris for a prescribed amount of time. The game time was chosen such that, on average, the amount of time between the first ad disappearing and the following questionnaire was the same across conditions. This ensures that on average, each participant experienced roughly the same amount of forgetting time between the initial ad exposure and test. After playing for the designated amount of time, participants were automatically directed to a questionnaire.

Once having arrived at the questionnaire, participants were unable to press the "back" button on their browser to return to the article. Participants were asked two multiple choice reading comprehension questions about the article on the previous page, after which they were asked three different types of ad memory questions, which have been used in other advertising studies [Drèze and Hussherr, 2003]: unaided recall, textual recognition and visual recognition. The first page after playing Tetris asked an unaided recall question: "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 next page then consisted of four separate recognition questions with textual cues of the form, "Did you see a ___ ad?" with Netflix, Jeep, Avis, and American Express being the advertisers filling in the blank. After answering these questions participants then went to a page which consisted of four separate recognition questions with cues of the form, "Did you see the following ad?" with a picture of the Netflix, Jeep, Avis, and American Express ads following each question. The ads were chosen such that each had a strong visual resemblance to another ad. The Avis ad is primarily red, much like the Netflix ad, and the American Express ad is primarily black, much like the Jeep ad. Thus when the Netflix ad was shown, the Avis ad acted as its "lure" ad and *vice versa*. Similarly when the Jeep ad was shown the American Express ad acted as its lure ad and *vice versa*. The lure ads were used to check, for example, if users were simply remembering that there was a red rectangle in the top right part of the screen or if they actually noticed the ad itself and the advertiser depicted in it.

3 Results

From the initial sample of 1,100, we excluded observations on the basis of the following *a priori* criteria: not completing the survey, incorrectly answering both reading questions, taking fewer than 40 seconds (and thus not getting the full time treatment), or taking more than 4 minutes. The remaining 916 participants make up the set we analyze.

The top three panels of Figure 2 graphically address the question of whether a greater total probability of remembering an ad is achieved with two ads of $t = 10$ seconds or one ad of length $2t = 20$ seconds. For each of the three memory metrics, the sum of the metric for the two 10 second ads ($\Pr(A|AB) + \Pr(A|BA)$, $t = 10$), indicated by the dotted green line, is significantly higher than the sum of the metric for the single 20 second ad and the false alarm rate ($\Pr(A|AA) + \Pr(A|BB)$, $t = 10$), indicated by the orange line. For example, when $t = 10$, the probability visually recognizing the ad in the AB and BA treatments sums to .58, while that of the AA and BB treatments is only .41. Text recognition and unaided recall show a similar pattern. Thus, when t is 10 seconds, the total amount of recollection in AB + BA treatments exceeds the total in the AA + BB treatments.

However, when t is 20 seconds, a very different picture emerges. As can be seen in the bottom panels of Figure 2, there is no significant difference between $\Pr(A|AB) + \Pr(A|BA)$ and $\Pr(A|AA) + \Pr(A|BB)$. Surprisingly, the sum of two ads ($\Pr(A|AB) + \Pr(A|BA)$) at $t = 10$ is comparable to the sum of two ads at $t = 20$, differing by at most four percentage points. Thus, not only are two short (10 second) ads better than one ad of twice the duration, they are also roughly equivalent to *two* ads of twice the duration.

The main conclusion of the above analysis is that if A and B are advertisers and ad slots are short (around 10 seconds), it seems that more total impact on memory is created when splitting an impression between two advertisers than giving each advertiser its own full slot. That is, in the terminology established earlier, the memory under AB + BA is greater than memory under AA + BB. However, it is in principle possible for $\Pr(A|AB)+\Pr(A|BA) > \Pr(A|AA)+\Pr(A|BB)$ to hold averaged over many different advertisers, but not for a specific advertiser. For example, advertiser A might benefit greatly from the split impressions, while advertiser B suffers slightly. To check whether this occurs in practice, we take advantage that the experimental design uses four unique advertisers, each of which can be used as a test to see whether two short ads lead to more recall than one ad of twice the duration plus the false alarm rate. Again, a difference between the $t = 10$ and the $t = 20$ condition emerges. In the former case, in 11 of 12 tests, $\Pr(A|AB) + \Pr(A|BA) > \Pr(A|AA) + \Pr(A|BB)$, while in the latter case, there is no clear pattern. This lends support to the conclusion that the benefits of shorter ads hold within advertisers.

3.1 Lure Ads

Across all conditions, the rate of incorrectly indicating memory for one of the visually similar lure ads were low and quite similar to those in [Goldstein *et al.*, 2011]: 0% for recall, 6.6% for text recognition and 7.5% for visual recognition questions. Thus if we asked a user if they remembered an ad that was not shown, whether or not that ad was visually similar to the ad that was shown did not have a substantial effect on the false recall rate. We can conclude that in the text and visual recognition numbers, i.e. $\Pr(A|AB)$, $\Pr(A|BA)$, and $\Pr(A|AA)$, reported in Figure 2, users remembered more than just the predominant color of the ad.

3.2 Effect of Onset Time

We define an ad's onset time to be the amount of time between the page loading and the ad appearing. In Figure 2 if one compares $\Pr(A|AB)$ with $\Pr(A|BA)$ one will see $\Pr(A|AB) > \Pr(A|BA)$ for all treatments. That is, the second ad presented is at a disadvantage compared to the first. Moreover, the difference $\Pr(A|AB)-\Pr(A|BA)$ is far larger when $t = 20$ seconds than when $t = 10$ seconds. Figure 2 shows the difference in heights is much greater in the bottom three panels ($t = 20$) than the top three panels ($t = 10$). This shows that the longer the onset time of an ad, the less likely it is to be remembered. Accordingly, advertisers should value ads with early onset times. Onset time can also explain why two short ads are roughly as effective as two longer ads. When the ads are longer, the second ad appears at a greater onset time reducing the chance of it being remembered. In addition, this suggests that slotting more than two ads into an impression is not likely to be effective. Given ads of even short durations, such as 10 seconds, ads beyond the second would have onset times so great as to diminish their memory rates.

4 Conclusion

Display ads are typically sold by impression, a scheme in which a two-second impression costs the same as a two-minute impression despite the latter having a larger effect on memory. We have shown that two, short-duration ads increase memory per slot compared to a single, longer duration ad when slot lengths are reasonably small (around 10 seconds), and that this effect disappears as slot lengths grow. Not only may advertisers be better off under a time-based policy, but publishers should as well, since having short slots increases the number of ads that can be displayed. Our results strengthen the case for moving from an impression based pricing scheme to one that is either partially or completely based on exposure time.

Acknowledgments

This research was conducted while the authors were members of Yahoo! Research. This work was invited by the IJCAI program committee to serve as a summary of Goldstein et al. [2012].

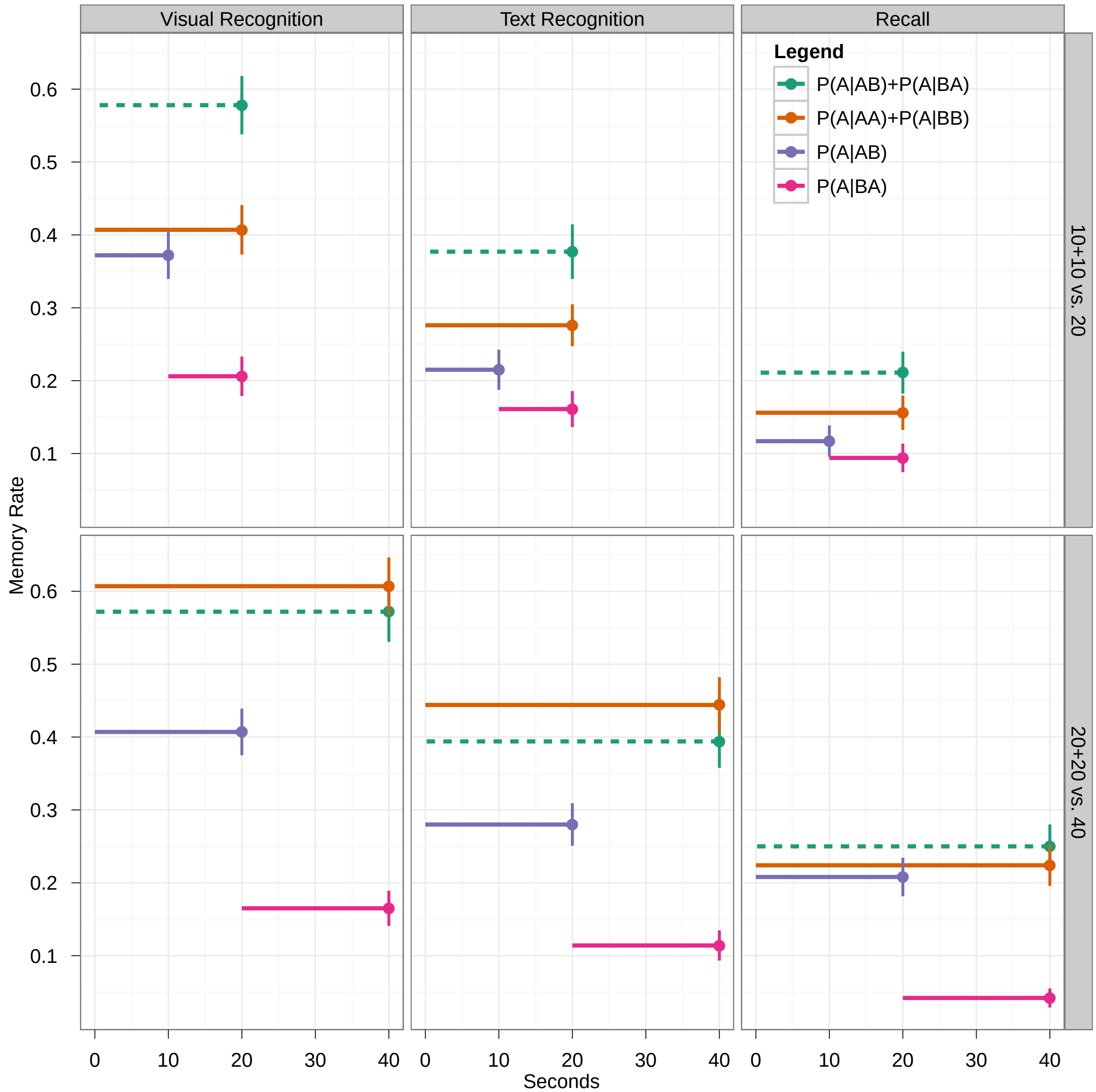

Figure 2: The x-coordinate of the leftmost endpoint of each horizontal line indicates the time at which the ad appeared (either at 0, 10, or 20 seconds). The length of each line indicates the duration of the ad. The y-axis indicates the probability (or sum of probabilities) of remembering according to the three memory metrics. Vertical line segments are confidence intervals of one standard error. For each memory metric, the dotted green line shows the sum of the metric for the two short ads. The top three panels, the $(t = 10)$ condition, compares the sum of two 10 second ads $(P(A|AB) + P(A|BA))$ to the sum of one 20 second ad and the false alarm rate $(P(A|AA) + P(A|BB))$. The bottom three panels, the $(t = 20)$ condition, are analogous except that the shorter ads had a duration of 20 seconds each and the longer ads had a duration of 40 seconds each.

References

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

[Goldstein *et al.*, 2011] Daniel G. Goldstein, R. Preston McAfee, and Siddharth Suri. The effects of exposure time on memory of display advertisements. In *Proceedings of the 12th ACM conference on Electronic commerce*, EC '11, pages 49–58, New York, NY, USA, 2011. ACM.

[Goldstein *et al.*, 2012] Daniel G. Goldstein, R. Preston McAfee, and Siddharth Suri. Improving the effectiveness of time-based display advertising. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 639–654, New York, NY, USA, 2012. ACM.

[Lewis and Reiley, 2011] Randall A. Lewis and David H. Reiley. Does retail advertising work? Measuring the effects of advertising on sales via a controlled experiment on Yahoo! Working Paper, June 2011.

[Manchanda *et al.*, 2006] Puneet Manchanda, Jean-Pierre Dubé, Khim Yong Goh, and Pradeep K. Chintagunta. The effect of banner advertising on internet purchasing. *Journal Of Marketing Research*, 43(1):98–108, 2006.

[PricewaterhouseCoopers, 2011] PricewaterhouseCoopers. IAB internet advertising revenue report: 2010 full year results, April 2011.

[Starch, 1923] D. Starch. *Principles of Advertising*. A. W. Shaw Company, Chicago, 1923.

[Wells, 2000] W. Wells. Recognition, recall, and rating scales. *Journal of Advertising Research*, 40(6):14–20, 2000.