

Effect of Temporal Spacing between Advertising Exposures: Evidence from an Online Field Experiment

Navdeep Sahni*

Booth School of Business

The University of Chicago

September, 2011
(Job-market Paper)

Abstract

This paper aims to understand the impact of temporal spacing between ad exposures on a consumer's decision of whether to purchase the advertised product. I create an individual-level dataset with exogenous variation in the spacing and intensity of ads by running online field experiments. In these experiments, exposure to ads is randomized across individuals and over time. The data show that at a purchase occasion, the likelihood of a product's purchase increases if its past ads are spread apart rather than bunched together, even if spreading apart of ads involves shifting some ads away from the purchase occasion. Because the traditional models of advertising do not allow for this effect, I build a new memory-based model of learning through ad exposure. Using a nested test, I reject a goodwill stock model based on the Nerlove and Arrow [1962] approach, in favor of the more general memory-based model. Counterfactual simulations using parameter estimates show that not accounting for the spacing effect of ads might lead to significantly lower profits for the advertisers.

*Email: nsahni@chicagobooth.edu. I thank the members of my dissertation committee: Pradeep Chintagunta, Günter Hitsch, Jean-Pierre Dubé, Ali Hortaçsu; members of the faculty including Ron Goettler, Reid Hastie, Aparna Labroo, Yesim Orhun, Oleg Urminsky, Ting Zhu and my colleagues at Chicago Booth, Valentin Haddad, Elisabeth Honka, Pranav Jindal, Devesh Raval, Daniel Wilhelm for having numerous useful discussions with me. I am also grateful to an anonymous website for partnering with me to run experiments and to Deepinder Goyal, Gunjan Patidar and Pankaj Chaddah for their inputs throughout the process.

1 Introduction

Firms spend significant resources on advertising. In the CPG industry, for instance, advertising investments are 7% of sales revenues on average (AdvertisingAge [2010]). Accordingly, marketing researchers have been interested in quantifying the impact of advertising at both the aggregate market and individual consumer levels. Additionally, a common view about advertising is that it has long-lived effects (Clarke [1976], Lodish et al. [1995a]), which has generated a rich literature that has attempted to quantify both the short and long-run impacts of advertising (see Sethuraman et al. [Forthcoming] for a review).

To quantify the long-term effects of advertising on consumers' decisions, researchers have traditionally used models of advertising carry-over based on the notion of a goodwill stock that accumulates due to advertising exposures (Nerlove and Arrow [1962]).¹ According to these models, the marginal effect of an ad depends only on its timing relative to the purchase occasion, and the total goodwill from other ads. Holding these two factors constant, an ad's marginal effect does not depend on the timing of other ads relative to it.

In the behavioral literature, on the other hand, one of the key mechanisms suggested for advertising carry-over effects is learning through advertising exposures (Sawyer and Ward [1979]). According to this mechanism, repeated exposure to advertising for a product strengthens the memories associated with it, thereby increasing the likelihood of their recall at a purchase occasion. Memory research, however, also suggests the efficacy of learning through repetition increases if the learning occasions are spaced out in time rather than massed together. Janiszewski et al. [2003] discuss the potential of the spacing effect in the context of advertising though corresponding empirical evidence is currently lacking. Hence the importance of the spacing effect in advertising is yet to be quantified.

With the above background, the research questions I focus on in this paper are the following. For an individual consumer, how does the marginal effect of an ad depend on that individual's past *schedule* of exposure to ads for that product? Specifically, during a purchase occasion, does just the recency of past ads influence consumer decisions, or is there any direct effect of time intervals or spacing between past advertising occasions? From an advertiser's perspective, is the spacing effect important?

A prerequisite to estimating the impact of carry-over and spacing of past ads is the ability to measure the causal effect of advertising on consumer decisions. This is a non-trivial task for a researcher using secondary data. First, to credibly show that advertising does influence consumer decisions, one needs to find supporting evidence at the individual level. Second, even if individual-level observational data on exposures and decisions are available, establishing causality

¹Some researchers have used more general distributed lag models, where coefficients on lagged ads may not be restricted. Clarke [1976] summarizes some of the other modeling approaches used to quantify long-term effects.

is a challenge because of the targeting of ads. Individuals exposed to ads are the ones the advertisers are specifically targeting and hence constitute a selected population. Comparing purchase behaviors across targeted and un-targeted segments may lead to problematic inference because of other differences between these segments. Third, one needs to be concerned with another selection problem. Typically, consumers exposed to more ads might be systematically different from the rest of the population. For example, in the case of TV ads, individuals who spend more time watching TV view more ads. TV viewing habits might in turn be correlated with unobserved consumer characteristics that can potentially influence purchase behavior.² Therefore, unless individual differences are properly accounted for, this selection problem might cause biased inferences regarding advertising effects.

For this study, I focus on the context of sponsored search advertisements. To address the above challenges and accurately estimate the impact of advertising and the effects of spacing, I ran 11 randomized field experiments at an online restaurant search website. By design, every session for any individual visiting the website was randomly allocated to either an ad or a no-ad condition. In the ad condition, every page browsed during the session had a random chance of displaying an experimental banner ad, as opposed to a “dummy” banner (unrelated to the category). On the other hand, in the no-ad condition, all the pages browsed during the session displayed the dummy ad.

This unique empirical set-up has several advantages. First, it provides individual-level data. I observe page-wise ad exposure and detailed information on browsing, including the restaurant pages viewed by all individuals. Moreover, I observe when an individual generates a “sales lead,” by viewing the restaurant’s phone number. Information on calls made to the experimental restaurants is also available. These aspects of my dataset facilitate inference of the economic value of advertising and the role of ad spacing. Second, the experimental design provides exogenous variation in the intensity of ads displayed during a session, thereby overcoming the selection problems highlighted above. Third, since some individuals visiting the website at subsequent occasions might not be exposed to the ads, I get exogenous variation in the time gaps between advertising occasions, even among individuals that visit the website with the same frequency. This variation identifies the impact of spacing between ad occasions. Fourth, all but one of the experiments are new advertising campaigns, created for the purposes of this study for the restaurants that never advertised on this website before. Therefore, I am able to separately identify the impact of current and past ads without making assumptions about the advertising effect carried over from before the experimental time period.

Using these data, I first show that in the short run (i.e., during a session), sponsored search advertising has a statistically significant impact on consumer choice. In my data, on average, allocation to the ad condition increases the consumer’s likelihood of visiting the advertising

²In their empirical context of online display advertising at Yahoo!, Lewis and Reiley [2009] find individual purchase behavior is correlated with number of ads viewed.

restaurant's page by 15% and the likelihood of generating a sales lead by 48%. Furthermore, I find that conditional on the browsing behavior, multiple exposures to advertising during a browsing session leads to a greater impact on consumer choice. These findings provide the basis for further analyses of carry-over and spacing effects.

Data on repeat visits by individuals show the presence of longer-term effects. I provide model-free evidence of carry-over to future purchase occasions. I find the marginal effect of repeating an ad display at a later occasion is smaller if the individual was exposed to ads for the same restaurant in the previous sessions. Another interesting finding is that at a particular purchase occasion, the total effect of past ads does depend on time intervals (spacing) between previous ad exposures. Specifically, when past advertising occasions occur less than two weeks apart, increasing the time intervals increases the impact of advertising on consumer choice on average, even if ads are shifted away from the purchase occasion. Therefore, consistent with predictions from memory research, increasing the spacing between past advertising exposures might be beneficial for the advertiser. To my knowledge, this study is the first to provide data that identifies both carry-over and spacing effects of advertising.

With *prima facie* evidence for the effects of advertising, carry-over and spacing, I turn to building a framework for gauging the importance of the spacing effect and evaluating its impact on advertising strategies. Since the traditional models of advertising such as the goodwill stock model do not allow for this effect, I propose a new memory-based model of learning through advertising exposure, grounded in findings from memory research in cognitive psychology. Specifically, my model is based on the memory module of ACT-R (Adaptive Control of Thought - Rational) model of mind (Anderson et al. [2004]) that has previously been applied in education research to design optimal practice schedules (Pavlik and Anderson [2008]). Unlike the traditional advertising models, the memory-based learning model allows for the spacing effect.

I specify an empirical version of the memory model for my context of sponsored search advertising and estimate it using the data generated by my field experiments. Results from the estimation show the data support the memory-based model and reject the Nerlove-Arrow model which is nested in a specification of the proposed model. The memory model is also better for predicting the behavior of individuals exposed to ads repeatedly after small time gaps. By simulating returns from various advertising strategies, I show that using the wrong model might lead to significantly lower returns for the advertisers.

This paper contributes in several ways to research focused on understanding the impact of advertisements. It adds to the literature on advertising carry-over effects, by providing experimental evidence that even simple banner ads that are not too informative about the product can lead to persistent effects of advertising. Further, it suggests and provides evidence for an explicit model of ad carry-over through learning by repeated exposure to ads that is testable and can be easily compared with the traditionally used models. The paper also contributes to past research fo-

cused on understanding the effects of sponsored search ads (Ghose and Yang [2009]) by providing evidence of their impact using randomized field experiments. It quantifies the impact of different levels of ad exposures during a search session, which has implications for advertising strategies on search engines such as Google. For researchers and managers interested in scheduling ads over time, my study can provide inputs to understanding the trade-offs involved in ad scheduling. Furthermore, the paper contributes to research focused on incorporating findings in psychology into economic models of consumer behavior by estimating a model based on psychology primitives using real market data.

In the following sections, I start with a discussion of some relevant papers in the existing literature on advertising. In section 3, I discuss the source of my data for the analysis, followed by a discussion on the experiment design in section 4. Next, in section 5 I present my analysis providing evidence for the short-term effects of advertising on user clicks, page visits, and conversion to sales leads. In section 6, I focus on the long term, showing evidence of carry-over and spacing effects of ads. In section 7, I introduce a model of learning through advertising and discuss its key features that lead to different advertising implications. Details of structural estimation are discussed in section 8. Lastly, I summarize the findings and conclude in section 9.

2 Relevant Literature

The empirical literature on advertising is vast, Sethuraman et al. [Forthcoming] review the literature quantifying advertising elasticities. There are few studies showing clear evidence of the causal effect of ads on sales in a field setting. Lodish et al. [1995b,a] run field experiments using split-cable technology to show that TV advertising can lead to increase in sales, and this effect may carry over to the future. Lewis and Reiley [2009] show online display advertising can be effective in driving offline sales.

Online sponsored search ads have been a topic of recent, active research in Marketing (Ghose and Yang [2009, 2010], Rutz and Bucklin [2007]). Most of the studies on this topic use observational data about an advertiser’s ad-campaigns, aggregated at the keyword level on major search engines such as Google. Therefore in making causal inferences, they face the econometric challenges listed above. I overcome these issues by obtaining individual-level data from a search engine (rather than from one advertiser) and exogenously manipulating ad exposure.

On the impact of repeated exposures of ads in general, there is no clear evidence due to the econometric issues researchers using aggregate field data face. Many of the early experimental studies in lab settings (summarized by Pechmann and Stewart [1988]) used brand/advertising recall as dependent measures that have been found to be inaccurate predictors of sales (Lodish et al. [1995b]). Studies using Single Source data have used individual-level exposure and purchase information to study advertising effects (e.g., Ackerberg [2001], Terui et al. [2010]). However,

these data do not have exogenous variation in the intensity of ad exposure across individuals and are therefore not suited to make causal arguments unless other differences among individuals are controlled for. Winer [1993] discusses the usage and limitations of Single Source data in detail.

Most studies on modeling advertising carry-over use distributed lag models (Clarke [1976]). Many use an exponentially decaying goodwill stock model based on the approach suggested by Nerlove and Arrow [1962]. According to these models, impact of ads on a purchase occasion depends only on the time elapsed between the ad and the purchase occasion. On the other hand, the behavioral literature highlights the role of persistent learning, that is retention of information in the memory in the carry-over of advertising effects (see Sawyer and Ward [1979]).³ Janiszewski et al. [2003] further discuss the implications of the spacing effect in verbal learning on the impact of repeated exposure to ads. However, to my knowledge, there is no clear evidence supporting the spacing effect for advertising in the field. Also, the traditionally used models of advertising carry-over for individual consumers don't explain this effect.⁴

In the field of psychology, the spacing effect has been discussed repeatedly over the years since Ebbinghaus discovered the phenomenon in 1885 (see Ebbinghaus [1913]). Researchers in cognitive psychology have built models explaining the process of memory storage and retrieval - allowing for the spacing effects in learning. I specifically build on the ACT-R model of mind (see Anderson et al. [2004]). Anderson and Milson [1989] show that this model of memory is a rational memory design; that is, it is an optimal information retrieval system where items are stored in such a way that the cost of retrieval is minimal. It has also been used to design optimal practice schedules in education research; for example, Pavlik and Anderson [2008] show that an optimal schedule of practice based on this model maximizes recall at minimal cost (time invested in learning).

3 The Empirical Context

For this study, I conducted field experiments at one of the largest restaurant search websites in India. The portal is designed to help consumers choose restaurants for ordering food or dining. It provides information such as phone numbers, scanned menus, ratings, reviews and so forth, for more than 10,000 restaurants in six major cities in India. The website claims to cover all restaurants in these markets. Indeed, for the markets considered in this study, the number of restaurants in this website's database is roughly 40% higher than those returned by searching on the next major competitor. Analysis of the search data reveals the majority of website visitors use it to order food for home delivery.

³The term "learning" here is used to describe retention of information as opposed to gaining new information/quality e.g. in Erdem et al. [2008].

⁴Naik et al. [1998] build an aggregate market-level model of advertising, accounting for wearin-wearout of ads. This model allows for firms to spread their aggregate spending over time, but does not describe an individual's response to ads.

On the website, users can have multiple search sessions over different purchase occasions. Many sessions start at the home page, which contains a search query box and a set of filters the user can apply (e.g., choosing the geographic area for search). This page also displays some of the latest reviews of restaurants but has no ads. Once the user enters a query, he reaches a search results page, that contains a list of restaurants satisfying the filters applied, and sorted by popularity. Each search results page has links to pages of up to 20 restaurants listed, along with basic information such as the cuisines served, approximate cost for two, address, rating, and so on. In search engine terminology, these links are called organic links. By clicking on the organic links (or through ads in the right-hand panel), users can reach a restaurant’s page. Restaurant pages provide additional information: scanned menus, detailed ratings, and editorial and user reviews. Figure 1 shows a snapshot example of a search result and a restaurant page. To view the phone number of the restaurant, the user has to click on a link on the restaurant page, which then flips and displays the number. This extra step, viewed as a “sales lead,” was added to all restaurant pages for these experiments, to enable tracking of the restaurants users chose. Each page also has a search query box and filters to restart or narrow the search process. For the current analysis, this set up is important since all the restaurant pages are on the same website, allowing me to track all the choices individuals viewed.

Targeting of Ads

Any restaurant or search results page has upto eight slots for sponsored banners on the right-hand panel.⁵ Over the course of the experiments, the website practiced plain geographic targeting of ads; all users searching for a restaurant in a geographic area were exposed to the same set of ads.⁶ For example, during a particular week, every user who visited a page for a restaurant in South-Delhi, or browsed for restaurants in this area, was exposed to the same set of ads in the right-hand panel of the page, regardless of past browsing behavior or any other characteristics of the search.

Targeting of ads has implications for measurement of both immediate and long-term advertising effects. In the context of this restaurant search website, in the absence of an experimental manipulation, individuals exposed to ads for a restaurant are different from those not exposed, in terms of the geographic area of search. Compared with the untargeted population, the targeted individuals might be located closer to the advertiser’s restaurant and therefore more likely to purchase from it even in the absence of ads. Hence, comparing choices made by targeted individuals who see the ads with un-targeted ones who do not would lead to problematic inferences.

Furthermore, individuals who browse more pages on the website during a session see more ads.

⁵If the advertiser is a chain, then clicking on the ad banners typically takes the user to a page with links to web pages for the chain’s outlets in the filtered geographic area. For independent restaurants, on the other hand, ad banners point directly to the restaurant page.

⁶Restaurants in the database, are grouped into *zones* and *sub-zones* based on their geographic locations. A typical zone is spread across an area of around 25 sq. mi. and a sub-zone around 1 sq. mi. The advertisements are based on the filtered zone, or the zone of the restaurant whose page was visited.

Those who see more ads are part of a self-selected subset of the population who search more and therefore may have lower search costs, and possibly systematically different preferences. Hence, in the absence of exogenous variation, comparing choices of individuals who see more ads with those who see fewer can lead to biased estimates of the impact of additional advertising.

4 Experiment Design

I address the challenges in measuring the impact of ads by designing experiments to create exogenous variation in advertising exposure. To achieve generalizable results, I ran 11 similar experiments with different advertised restaurants in non-overlapping geographic areas and time periods. In every experiment, one advertising slot was reserved for experimental manipulation and the rest continued to display paid ads as usual. By experimental design, every session by a user at the website was randomly assigned to one of the following conditions:

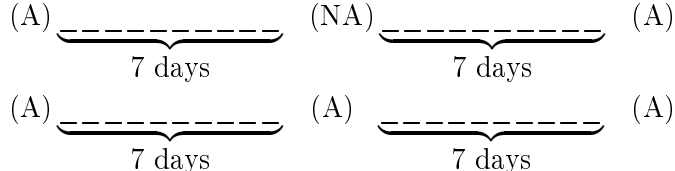
- *No-ad condition* (NA): A session allocated to the no-ad condition was not exposed to ads for the experimental restaurant in any of the pages browsed. The ad slot chosen for the experiment displayed a dummy banner instead.
- *Ad condition* (A): Every page request by a user session in this condition had a random chance of displaying an ad for the experimental restaurant. Since there is an element of chance even within this condition, two users in (A) that browsed the same number of pages could end up being exposed to the ad for the experimental restaurant a different number of times. Within (A) there were different randomized sub-conditions:

- (A1) Experimental banner was displayed in position 1.
- (A2) Experimental banner was displayed in position 3. There were multiple ad banner copies and the user session was exposed to one of them which was chosen randomly at the beginning of the session.
- (A3) Experimental banner was displayed in position 5.
- (A4) Experimental banner was displayed in position 7.
- (A5) Experimental banner was displayed in position 3 but a different restaurant was advertised in position 2 compared to (A2).

Figure 2 shows snapshots of all of the conditions for one of the experiments. Not all experiments had all sub-conditions because some geographical areas were expected to have less traffic than others. Appendix B summarizes details about the distribution of sessions across experiments and conditions.

Comparing the users in conditions (A) and (NA) identifies the main effect of treatment with ads. Because allocation to (A) versus (NA) condition is random, this manipulation helps overcome the measurement problems caused by the targeting of ads. Next, because the likelihood of displaying the experimental ad at every page is random in (A), individuals who browsed the same number of pages may be exposed to a different number of ads. Therefore, comparing the decisions of individuals with the same browsing behavior and a different number of ad exposures in (A) helps identify the impact of multiple ad exposures during a session. Differences across sub-conditions (A1) through (A4) would identify the impact of different ad characteristics, such as the location of the ad in the right-hand panel. For analysis in this paper, I pool data from all sub-conditions of (A). Therefore, the effects identified here should be interpreted as average effects across all experimental restaurants and all ad characteristics varying within (A).

Note that a user revisiting the website in the future started a new session and therefore could be allocated to a different experimental condition. This helps create variation in time intervals between advertising exposures for the experimental restaurants, even among individuals who visited the website with the same frequency. For illustration, consider two individuals who have the same browsing behavior and visited the website the same day of every week. Since every session had a random chance of being in (A), the past schedule of ads for the two individuals after their third sessions might look like this:



In this example, the spacing between ad occasions in the first schedule is 14 days whereas in the second one, it is seven, causing the needed variation in spacing. This variation, along with page-wise advertising randomization within (A) creates exogenous variation in the number of ads displayed in each of the sessions, which is necessary for identifying the carry-over effects.

Experimental restaurants and banners

Restaurants chosen for the experiments (detailed characteristics are shown in Table 1) are diverse, ranging from a 100-year-old restaurant with Indian cuisine to a two-year-old Chinese restaurant, to international pizza chains. They are fairly popular in their respective geographic areas; in the absence of ads, on average, around 1.5% of individuals browsing the website for the geographic area visit the restaurant's pages. All experimental restaurants also take food orders via phone for home delivery.

Throughout a browsing session, the ad banners displayed for the experimental restaurants remained the same. They typically displayed the brand name and did not provide much information

about the restaurants. Some ads also displayed a restaurant logo and a short three to four-word phrase that highlighted a cuisine and/or whether there was any discount.⁷

Restaurants chosen for 10 out of the 11 experiments had never advertised on the website before, and were not aware of the experiments. For these 10 restaurants, I created ad banners just for the purpose of the experiments. Therefore, the users were not exposed to these ads before their first visit to the website during the course of the experiments. This fact is important for separating the contemporaneous effects of advertising from carry-over effects, because in this setting, I can rule out the past carry-over effects before an individual’s first session. Moreover, five out of these 10 experimental restaurants were not even aware that this website was advertising their business. Therefore, for these experiments, I avoid picking up effects due to the interaction of advertising with other potential unobserved actions by the advertising restaurants.⁸

4.1 Data summary

I use data for 211,135 users (identified by cookies) who visited the website between August 11 and November 1, 2010, and a total of 256,690 sessions, with about 20% of the sessions in the no-ad condition.⁹ Just 11% of the individuals re-visited the website during this period; therefore, the rest do not help identify carry-over or spacing effects. However, the cross-sectional variation is important in identifying immediate effect of ads, which forms the basis for long-term effects. Throughout the search process, the users stayed on the domain of this website, enabling tracking of all the options viewed. Therefore, for each user, I observe the entire search process including the search queries, details about page-wise advertisement exposures (e.g., advertiser restaurant’s details, position on the right-hand panel), restaurant pages they decided to visit, the restaurants for which they generated a sales lead at the website (looked up the phone numbers). Furthermore, users in different experimental conditions were displayed different telephone numbers (owned by the search website) for each experimental condition. Whenever a call was made to one of these numbers, it was forwarded to the advertising restaurant and the conversation was recorded in audio files.¹⁰ This procedure will allow me to study the impact of different experimental conditions on calls and orders made to advertising restaurants.¹¹

The response variables I mainly focus on are (1) a consumer’s decision of whether to visit the experimental restaurant’s page and (2) whether the consumer generates a sales lead (viewed the

⁷Six of these 11 restaurants offered no discounts over the course of the experiments

⁸For example, if advertising restaurants also tend to run concurrent offline promotions, the ad effect an online experiment measures could be the special case of “impacts of ads during a promotion.”

⁹When a user first visits the website, a unique cookie is deployed in the browser and a session is initiated. If there is no activity for the user (no page requests) for more than 3 hours, the session expires. When the user visits next (could be after a few hours or days), a new session is started for the same cookie.

¹⁰Calls were tracked for 8 out of 11 experiments. The callers were notified of this recording before the call was forwarded.

¹¹However, matching the individual cookie with the caller is difficult if many individuals in an experimental condition call the same restaurant at the same time (I haven’t tried this matching yet).

experimental restaurant’s phone number).¹² I consider viewing the phone number as a stricter measure, which indicates specific interest in ordering food for delivery from the restaurant. The sales team of the website trusts the number of leads to be a good approximation of purchase. To check this approximation, I matched the number of leads generated with the number of calls made from unique phone numbers for 48 restaurant-week observations. The plot in Figure 3 shows these data. The points are close to the 45-degree line, indicating a nearly one-to-one match, supporting the belief that sales lead is a good measure of conversion of a page visit to an order.

To understand the heterogeneity in the browsing behavior, in Figure 4, I show a histogram of the distribution of the total number of pages browsed in the user sessions observed in the data. Note that it is positively skewed, and about 50% of all sessions ended in just one page visit. Typically, these are the individuals who reached a restaurant page through an external search engine such as Google. 97% of all sessions had less than 20 pages browsed, so for most of the analysis, I use data on these sessions to avoid the influence of outliers on the results. Figure 5 illustrates the variation in the number of ad exposures during a session for sessions that browsed five pages. In the absence of randomization, all the sessions with five pages would have had five ad exposures, but due to experimental design, the number of ad exposures varied from 0 to 5.

Table 2 summarizes the details of the users’ browsing behavior. Note that the median session had one restaurant page visit. Just a fraction (1.67%) of all users clicked on any advertisement during the time period of the data. About 17% of all sessions generated a sales lead, that is, viewed phone numbers for a restaurant.

Table 2 also shows the distribution of the repeat-visit behavior of users on the website. About 11% of all users visited the website more than once during the time period and 99.96% had less than 15 sessions in total. For analysis of carry-over and spacing effects across sessions, I focus on these users with less than 15 sessions. Figure 6 shows the distribution of time intervals between the first two sessions of those individuals who visited the website multiple times during the time period. The maximum gap between sessions is about 80 days but the majority of second visits happened within 10 days of the first.

5 Short-term Effect of Exposure to Ads

In this section, I explore whether the experimental ads, in the above empirical context of the restaurant search website, have an effect on the consumers’ decisions during the session in which the ads are shown. Previous findings (e.g., Lodish et al. [1995a]) indicate short-term effects are important for the carry-over of ad effects to future occasions.

¹²Nine out of the 11 experimental restaurants are chains. When the users click on ads, they are shown links to restaurants from the chain located around the geographic area of search. Hence, clicking on ads might not always lead to a visit to the restaurant page.

5.1 Effect of the ad condition

I begin the analysis by examining the main treatment effect of a session being allocated to the ad condition, on the individual's decisions during the browsing session. For measuring the short-term effects, I focus on all individuals' first sessions during the experimental time period (to avoid confound with carry-over). As dependent variables, I use dummy variables indicating whether the individual in the session chose to

- $visit_i$: visit the experimental restaurant's page and
- $view_num_i$: view the experimental restaurant's phone number.

These variables can be 0 or 1 in both ad and no-ad condition, since a user can reach any restaurant via searching and clicking on the organic links. Recall that an interested consumer can view the phone number of the restaurant only after visiting the restaurant's page.

I investigate the average treatment effect by pooling data from all the experiments. To estimate the effects, I regress the dependent measures on a dummy variable indicating whether the session was allocated to the ad condition. Estimates of OLS regressions are shown in Table 3. Columns I and III show the main treatment effect on $visit_i$ and $view_num_i$. The estimates show the probability of visiting the advertiser's page increases by 15.6%, from 1.6% to 1.85% when a session is in the ad condition, that is, when the consumer had a chance to see the ad for the experimental restaurant versus the dummy banner. The corresponding increase in likelihood of the user looking for the phone number is higher. On average, it increases from 0.23% to 0.34%, which is a 48% rise. Note that the treatment effects are statistically significant at a 1% confidence level. To show outliers did not drive my effect, in columns II and IV of the table, I restrict the data to just the sessions with fewer than 20 pages browsed. Note the estimates do not change much.

The results provide clear evidence in data suggesting the ads in this context do have a significant impact on the decisions consumers make. In the online advertising industry, effectiveness of ads is measured by click-through rates (CTR). The average CTR for ads in experiments studied here is 0.50%. This number is considerably higher than the average CTR of internet display ads, which is around 0.10% for various industry verticals and ad formats, according to DoubleClick [2010], an online advertising industry benchmark. This difference highlights the targeting power of search engines, which leads to more efficient advertising. Also, notice that CTR is twice the actual increase in visits to the page due to ads.¹³ Therefore, clicks may be a noisy measure of gauging returns to investing in advertising on the website.

¹³Recall that clicking on an ad on this website does not imply a restaurant page visit, if the ad is for a chain. Clicking on the ad takes the user to a page with links to the chain outlets in the area of search. The user may then choose to visit the restaurant page.

5.2 Impact of multiple exposure to ads in a session

In this section, I investigate whether the number of times a banner ad is displayed during a session is important for the session’s outcome. Measuring the impact of repeated exposures by itself is an important task advertisers face when making media-planning decisions in which they must choose between reach and frequency. While planning online campaigns, for example, advertisers must decide whether to place ads for websites/keywords that consumers are likely to visit during the same purchase occasion.

For a non-experimental ad, the number of exposures is higher for individuals that browse more pages on the website. However, due to this experiment’s design, the display of the experimental ad on any page is random during a session. This generates exogenous variation in ad exposure, but the average number of experimental ads seen during a session is still higher for individuals who browse more pages in the ad condition. To see this point, consider two sessions: one with one page browsed and the second with five pages browsed. If every page has a 50% chance of displaying the ad then we would expect a consumer in the first session to see the experimental banner 0.5 times, whereas a consumer in the second session would see it 2.5 times. Therefore, for identification of the impact of repeated exposures during a session, controlling for the number of pages browsed within a session is necessary.

I estimate the average consumer response to repeated exposure of experimental ads by regressing the decision to visit the experimental restaurant page on the number of times (pages) the individual was exposed to the experimental ads during the session ($nExp$). To avoid a bias in estimates, I control for the number of pages browsed in the session.

Columns I and II of Table 4 show results from a logistic regression of $visit_i$ on the number of exposures ($nExp_i$), with and without controlling for number of pages browsed. Notice the coefficient of $nExp_i$ is positive and statistically significant, but due to the positive correlation between exposure and pages, the magnitude goes down significantly when I control for the number of pages. Column III adds a quadratic term of number of exposures, which has a negative coefficient suggesting diminishing returns to exposure. To confirm an increase in ad intensity beyond the first exposure drives these coefficients, I add a dummy variable indicating positive exposure. If the coefficients are identified by positive exposure rather than repeated exposure, the dummy variable will pick up the variation changing the coefficient on $nExp$. Estimates in column IV show the coefficient of $nExp_i$ is still positive and the dummy variable indicating positive ad exposure is negative (not statistically significant). The regression results discussed up to this point are with all data pooled, that is, sessions with number of pages ranging from 1 to 20. In Appendix C, I extend this analysis by putting more controls for this heterogeneity and discussing the quantitative significance of repeated exposures.

The evidence discussed above suggests that in the data, increasing the number of times a banner ad is displayed to an individual during a session increases the consumer’s likelihood of visiting

the advertiser’s page. Therefore, for studying the long-term effects, it is important to account for the number of times a banner is displayed in a session.¹⁴

5.2.1 Discussion on measurement

Measuring the value of a single ad exposure in an interactive set up such as sponsored search advertising is challenging because the consumers can choose when to stop browsing and thereby control the number of ad exposures. In the above approach to measuring the impact of different levels of ad exposures, the experimental manipulation does not directly identify the effects. In all the regressions, I controlled for the number of pages browsed in the session, which is an endogenous variable chosen by the browsing individuals. The variable may depend on the sequence of ad exposures and also on preferences for the advertised restaurants, which might cause a bias in the estimated effects.¹⁵

To see this bias, suppose λ_i represents the preference for the advertised restaurant, for an individual i browsing the website. Also, let a higher value of λ_i represent a higher preference for the advertised restaurant for i . Now, if individuals with higher λ are also more sensitive to the ad then they are likely to respond after only a small number of ad exposures. Therefore, these individuals might quit the search process after a few ad exposures. Hence, under this assumption, the population that is selected into viewing a higher number of ads would be comprised of individuals with lower λ s. In other words, individuals viewing more ads (i.e., observations with higher $nExp$) might have lower-than-average sensitivity to the ad. This implies that the above regression approach to getting the ad response for various levels of ad exposures would yield estimates that are biased downward for higher levels. Arguments such as this imply that accurately measuring the impact of a particular ad exposure requires conditioning on the prior search activity and the sequence of advertising treatments individuals receive.

In Appendix D I discuss this issue in detail and under weak assumptions show that the potential bias, if any is small relative to the magnitude of the effect. For the rest of this paper, I proceed with incorporating linear controls for the number of pages, as in the previous section.¹⁶ Recall that by experiment design, every page viewed during the ad condition has a random chance of displaying the ad. This feature in the data potentially allows me to condition on the browsing behavior and the past sequence of ads to create “proper” control groups for measuring the impact of each ad. In Appendix E, I explore this issue further and provide a non-parametric test showing an additional effect of multiple exposures to ads.

¹⁴These regressions were also replicated with $view_num_i$ instead of $visit_i$ as the DV. Results remain the same qualitatively. Some cases, however, do experience a loss in precision.

¹⁵This issue is not discussed in previous marketing research on the impact of sponsored search advertising (e.g., Ghose and Yang [2009, 2010]).

¹⁶Alternatively, I can proceed by explicitly modeling the search process. This approach is left for future work.

6 Longer-term Effects of Ads

Analysis in the previous section shows an increase in the number of ad exposures during a session can lead to an increase in page visits and conversion. Next, to understand the intertemporal tradeoffs important for timing of advertising, I focus on understanding the effect of advertising carried over to future sessions. Therefore, I examine the data for individuals that have more than one session. Also, I keep individuals that have more than one day's gap between sessions.¹⁷ For these consumers, I look for how advertising in the first session affects the decisions consumers make during the second session.¹⁸ Due to experimental variation, similar individuals receive different intensity of ads at different occasions. Therefore, I regress the decision to visit the experimental advertisers' page in the second session on past exposure ($nExp_1$) and exposure in the current session ($nExp_2$), and control for the consumer browsing type by adding the number of pages browsed in the current session ($nPages_2$) to the regression. If the effects carry over to the second session, the coefficient on $nExp_1$ should be positive.

Column I of Table 5 presents results from this regression.¹⁹ Notice the coefficient for the exposure in session 1, $nExp_1$, is positive and statistically significant, suggesting the effect of ads carries over to the future. The coefficient for exposure in the second session, $nExp_2$, is also positive and the magnitude is similar to that of $nExp_1$. Column II shows results when a quadratic term for past exposure is included in the model. Note the squared term has a positive coefficient suggesting a carry-over effect is likely when the number of exposures in the past session is high.

Further, to investigate the impact of past advertising on the marginal effect of advertising in the current session, I include in the regression, the interaction term $nExp_1 \times nExp_2$. Estimates in column III show the coefficient on this interaction term is negative and statistically significant, suggesting high exposure in the past leads to a decrease in the marginal effect of advertising in the current session. Next, I include in the model the impact of the time gap between the two sessions ($days$) to understand how the impact of past advertising changes with time elapsed. Estimates in column IV show a negative coefficient for $nExp_1 \times days$ suggesting a statistically insignificant decrease in the marginal effect of past advertising with elapsed time.

To summarize, the data show a carry-over effect of ad exposure. When the influence of past advertising on the consumer is strong, which could be due to high exposure in the past or a small time gap, the marginal benefit from advertising is small. I explore the effects in more detail and perform more robustness checks in Appendix G.

The tradeoff the advertiser faces is also economically significant. To gauge the quantitative

¹⁷This is to make sure that the next session is actually a different purchase occasion. It also reduces the presence of variety-seeking behavior that goes against repeated visits to the same restaurant across sessions.

¹⁸An implicit assumption here is that consumer's decisions to revisit the website are not correlated with past advertising exposure. In the data I can check for these correlations, more details are presented in Appendix F.

¹⁹For analyzing carry-over, I don't include data from an experiment for which the advertiser was advertising before the experiments started.

impact of the above estimates, I simulate the likelihood of visiting the advertiser’s page as a function of current and past ad exposure for an individual, using the model in column III. Table 6 shows the estimates. Change of exposure in the second session, $nExp_2$, from 0 to 5 has an impact of 1.5% when $nExp_1$ is 0. This impact, however, is just 0.5% when $nExp_1$ is 5. The marginal impact of $nExp_1$ also varies with $nExp_2$ in a similar way and the changes are of the same magnitude.

6.1 Spacing effect

One of the key mechanisms driving carry-over of advertising effects is learning (Sawyer and Ward [1979]), where repeated exposure to the ads for the same product leads to easier recall of the product at purchase occasions. Memory research has found that spreading out past stimuli over time (spacing) can lead to more efficient learning and hence better recall of the associated memories at later occasions (Janiszewski et al. [2003]). In this section, I show evidence of this spacing effect in my data; that is, total impact from past advertising at a particular session may be higher if the previous ads are spread further apart.

To find model-free evidence for the impact of spacing, I consider individuals with more than two sessions in the data and focus on the impact of spacing between the first two sessions on the likelihood of the individual choosing to visit the advertiser’s page in the third session ($visit_3$):

$$1 \underbrace{\hspace{1.5cm}}_{days_{1-2}} 2 \underbrace{\hspace{1.5cm}}_{days_{2-3}} 3$$

An increase in the spacing between sessions 1 and 2 can have two effects. First, when $days_{1-2}$ increases, session 1 gets further away from the purchase occasion, which is session 3. Since the impact of ads decreases with time, increase in $days_{1-2}$ reduces the effect of exposures in session 1, thereby reducing the probability of visiting the advertiser’s page in session 3. Second, according to the spacing effect, when $days_{1-2}$ increases, the effectiveness of learning during session 2 increases. This increased efficacy of learning due to ads in session 2 increases the ad effect carried over to session 3, thereby increasing the likelihood of visit in session 3.

To see the presence of the spacing effect, I first search for moments in the data that would identify it. By definition, the detectable benefit from spacing is more likely to occur when there is high ad exposure in session 2, that is, when there is more opportunity to learn. Therefore, I focus on individuals exposed to more than two ads in session 2 ($nExp_2 > 2$). Figure 7 shows the mean chance of visiting the advertiser’s page in session 3 for this subset of individuals, split by the time gap between the first two sessions ($days_{1-2}$). It shows the likelihood of visiting the advertiser’s page in session 3 is about 2.5% on average if the second session occurred within a week of the first one. On the other hand, if the second session occurs after a larger time interval, the chance is higher, 4.4% on average, and the difference is statistically significant. This increase

might be driven by better learning in session 2 due to spacing. However, other differences among these individuals might also be the cause; for example, more frequent visitors may have different tastes or seek variety if they visited the advertiser earlier.

I control for these aspects by looking at partial correlations. Specifically, I regress $visit_3$ on the number of ad exposures in the current and the past sessions ($nExp_1$, $nExp_2$ and $nExp_3$), spacing between sessions ($days_{1-2}$, $days_{2-3}$) and the interaction terms. The coefficient of the interaction term $nExp_1 \times days_{1-2}$ represents the change in the marginal effect of exposures in session 1 due to an increase in spacing. Controlling for $days_{2-3}$, an increase in $days_{1-2}$ implies an increase in the time gap between session 1 and the purchase occasion, which is session 3. Therefore, the coefficient of $nExp_1 \times days_{1-2}$ is expected to be negative. On the other hand, the coefficient of the interaction term $nExp_2 \times days_{1-2}$ should be positive according to the spacing effect, which implies the efficacy of ad exposures in session 2 increases with $days_{1-2}$.

Table 7 shows results from various specifications of this logistic regression. Column I shows estimates of a model without interaction terms. As expected, $nExp_3$ and $nPages_3$ have a positive impact on the probability of a visit in the third session. Column II shows estimates when I include interaction terms in the regression. The coefficient of the first interaction term, $nExp_1 \times days_{1-2}$ is negative, suggesting that as the time between the first and second sessions increases, the marginal impact of $nExp_1$ decreases. The positive coefficient on $nExp_2 \times days_{1-2}$ shows that controlling for $days_{2-3}$, an increase in $days_{1-2}$ increases the marginal effect of $nExp_2$, but this coefficient is not statistically significant. Next, I focus on the subset of data, where $days_{1-2} \leq 15$ to see the impact of change in spacing at smaller levels. Column III shows results for this regression. The interaction term $nExp_2 \times days_{1-2}$ is now positive and statistically significant, showing that impact from exposure in the second session comes only when spacing between the first and second sessions is large. Next, to confirm variety-seeking behavior does not drive the effect, I run the same regression on the subset of individuals who did *not* visit the advertiser's page or that of any other major substitute during the past two visits.²⁰ The subset of individuals remaining in the data are less likely to exhibit variety-seeking behavior. Estimates of this regression in column IV show that although the standard errors are higher for this regression, the coefficient for $nExp_2 \times days_{1-2}$ remains the same in magnitude compared to column III. This regression indicates the effect identified earlier is not driven by the fact that relatively frequent website visitors are more variety seeking and hence, less likely to have $visit_3 = 1$.

To appreciate the total effect of an increase in spacing on $visit_3$ as predicted by the regressions, in Figure 8 I plot the predicted probabilities using the model in column V for multiple scenarios. I fix the number of exposures in the second and third sessions to four and vary the intensity of exposure in the first session. The plot shows that for a small number of exposures in the first session ($nExp_1$), the probability of $visit_3$ increases with the spacing between sessions 1 and 2.

²⁰I define substitutes as restaurants that are of the same cuisine as the advertiser, are in the same geographic area, and are frequently co-visited in the data.

That is, when $nExp_1$ is small, the loss due to reduced recency of $nExp_1$ is smaller than the gains from spacing. On the other hand, when $nExp_1$ is high, the loss in recency from shifting session 1 away from the purchase occasion dominates the gains from better learning through spacing. Therefore, an increase in spacing decreases the likelihood of choice in session 3 when $nExp_1$ is high.

This analysis shows evidence in the data, supporting the spacing effect suggested by memory research; the spacing of advertising across time has an economically and statistically significant impact on the effect of ads on consumer choice.

7 Framework for Modeling Advertising Carry-over

In the previous sections, I discussed the evidence in the data showing the presence of short-term advertising effects and carry-over to future purchase occasions. Also, I provided evidence indicating that spacing of advertising occasions plays a role in effectiveness of ads. Now, I focus on building a framework incorporating these aspects into a consumer demand model, to enable counterfactual analysis for comparing consequences of different advertising strategies or policy regulations.

In most of the existing research on advertising, carry-over is formulated as a distributed lag model. Current and past advertising exposures contribute to current sales for the advertised product and each lag is assigned a weight. A common approach is based on the model suggested by Nerlove and Arrow [1962], where overall awareness or goodwill for the product for an individual i at time period t is given by

$$G_{it} = \rho G_{it-1} + A_{it} \tag{1}$$

$$= \sum_{k=1}^t \rho^{t-k} A_{ik} \tag{2}$$

where A_{it} is the advertisement exposure for the individual, and ρ is the carry-over rate of past goodwill. According to this model, the effect of past advertising decays over time and is replenished by more advertising. Note that the goodwill model assigns strictly smaller weights to the advertising exposures further away from the purchase occasion. Therefore, for a particular purchase occasion, any schedule with ads closer to it is strictly better than one in which ads are spread out. Thus a model based on the Nerlove-Arrow approach does not allow for spacing effects.²¹ This might have implications in designing optimal advertising strategies; an optimal ad schedule based on this model might place ads too close together, possibly leading to suboptimal outcomes.

²¹This argument is shown more formally treated later in this section.

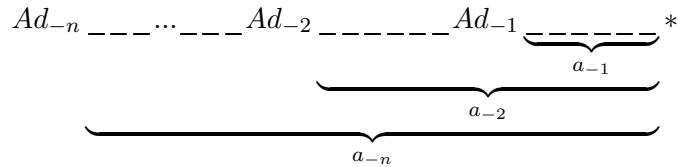
In the remainder of this section, I first describe an alternative memory-based model of carry-over of ad effects for a general context, which allows for the spacing effect and is based on foundations in research on memory. Next, I discuss the implications of the assumptions the model makes and compare them with the predominantly used exponentially decaying ad-stock model, both analytically and using simulations.

7.1 Memory-based model

The basic working of the model is as follows. Exposure to ads activates memories associated with the advertised product that in turn help individual consumers recall their experiences. Over time, this memory activation decays and so does the probability of recalling the advertised product at the purchase occasion. Future exposure to ads reinforces these memories and increases their likelihood of being recalled at the later occasions. The model is based on the following assumptions:

1. At a purchase occasion, the probability of the recall of experiences associated with the product depends on the strength of the memory trace, which I call the activation level.
2. The total strength of memory is the sum of strengths from each of the past exposures occasions.
3. The strength of memory due to an ad exposure decays as a power function of time.
4. The decay rate of strength due to an ad exposure is *specific* to the particular exposure occasion. When the memory for a product is fresh – for example due to recent advertising activity – an additional ad exposure will not have persistent effects. On the other hand, when the consumer has no prior memory about the product, the ad will have a longer-lasting impact.

Suppose that before a purchase occasion, an individual i is exposed to ads for the product on n occasions. Let the ages of these ads at the time of purchase be a_{-1}, \dots, a_{-n} , where the k^{th} last ad has age a_{-k} and a_{-n} is the age of oldest ad.²² The scenario can be shown as



²²The age of an ad at a particular point of time is the time elapsed since the ad exposure.

where $(*)$ represents the time the decision is being made. The contribution of the k^{th} last ad to the strength of memory is given by (ignoring individual subscripts i for now)

$$s_k = (a_{-k})^{-d_k}$$

where $d_k (> 0)$ is the decay rate of impact of an ad aged a_{-k} and is specific to the ad exposure at that time. As the exposure gets older, that is, as a_{-k} gets bigger, its contribution to memory strength s_k declines. The total memory activation due to all n ads is given by the log sum

$$m(a_{-n}, \dots, a_{-1}) = \beta + \ln \left(\sum_{k=1}^n s_k \right) \quad (3)$$

where β represents individual-specific ease of memory activation. Now the decay rate d_k for activation due to the k^{th} last exposure Ad_{-k} depends on the activation at the time Ad_{-k} took place:

$$d_k(m_k) = b + ce^{m_k} \quad (4)$$

where b and c ($c > 0$) are parameters and m_k is the activation due to ad-exposures $Ad_{-n}, \dots, Ad_{-(k+1)}$ that happened *before* Ad_{-k} happened. m_k would therefore depend on the time between Ad_{-k} and the previous ads as

$$m_k = m(a_{-(k+1)} - a_{-k}, \dots, a_{-n} - a_{-k}) \quad (5)$$

Equations (4) and (5) imply that when the memory of the product prior to an ad occasion is strong, the impact of advertising at that point has low persistence because its decay rate is large. I investigate this prediction of the model more formally later in this section, but to see it intuitively, suppose $a_{-(k+1)}$ is reduced, that is, $Ad_{-(k+1)}$ gets closer to Ad_{-k} . Therefore, strength of memory just before Ad_{-k} occurs, m_k increases leading to an increase in its decay rate d_k when $c > 0$. This is a key feature of the model that allows for different implications from the traditional goodwill stock model.

Memory activation before the first exposure is $m_n = -\infty$ (assuming no other source of awareness). Therefore, parameter b in (4) determines the decay rate when there is no past memory for the product. c is the decay-scale parameter that determines the increase in decay rate due to higher memory activation in the past.

7.1.1 Discussion

Two points are worth noting here. First, according to the model, the memory strength at a point of time very close to an ad is infinite. That is, in the above scenario, if $a_{-1} \rightarrow 0$ then

$$\begin{aligned} \lim_{a_{-1} \rightarrow 0} m(a_{-n}, \dots, a_{-1}) &= \lim_{a_{-1} \rightarrow 0} \left(\beta + \ln \left(\sum_{k=1}^n a_{-k}^{-d_k} \right) \right) \\ &= \lim_{a_{-1} \rightarrow 0} \left(\beta + \ln \left(\frac{1}{a_{-1}^{d_1}} + \sum_{k=2}^n a_{-k}^{-d_k} \right) \right) \\ &= \infty \end{aligned}$$

where the second step follows because age a_{-1} , is non-negative. So if the occasion of recall is close to an ad, then the probability of recall would be higher due to higher strength of the memory. Second, if two ads occur in close succession, the decay rate of the second will be high. If the last and the second-to-last ads in the above scenario are close, that is, $a_{-1} - a_{-2} \rightarrow 0$, then from equation (5), $m_1 = m(a_{-n} - a_{-1}, \dots, a_{-2} - a_{-1}) \rightarrow \infty$ following the same argument as before. So

$$\lim_{a_{-1} \rightarrow a_{-2}} d_1(m_1) = \lim_{a_{-1} \rightarrow a_{-2}} b + ce^{m_1} = \infty$$

Therefore, if two ads are close together then, according to the model, the second ad quickly loses its impact.

7.2 Implications of the model

The memory model implies the *spacing effect* mentioned earlier. Note that equation (4) implies the decay rate of activation due to exposure to an ad at one occasion depends on the activation level before *that* occasion. For illustration, consider two schedules of two ads each:

$$\begin{array}{c} \begin{array}{ccc} 1 & \underbrace{\quad\quad}_1 & 2 \underbrace{\quad\quad\quad\quad}_x \quad * \quad (S1) \end{array} \\ \begin{array}{ccc} 1 \underbrace{\quad\quad\quad\quad}_5 & 2 \underbrace{\quad\quad\quad\quad}_x & * \quad (S2) \end{array} \end{array}$$

S1 Before the decision occasion (*), there are two advertising occasions: one age x days and another age $x+1$.

S2 The two advertising occasions are now further apart: one age x days and another age $x+5$.

In this example, both schedules are the same except the older ad in schedule 1 is more recent than the older ad in schedule 2. Because of this difference, the memory strength contributed by the older ad is higher in schedule 1. On the other hand, since the two ads are closer to each other,

according to the memory model, the decay rate for the recent ad will be higher for schedule 1 compared to 2. Hence, the ad 2 will be more effective in schedule 2. Therefore, switching from schedule 1 to 2, that is, shifting the ad 1 away from the decision occasion has two opposite effects. The net effect can be positive if the gains from an increased impact of the second ad is greater than the loss due to reduced effectiveness of ad 1.

7.2.1 Comparison with an individual-level exponentially decaying ad-stock model

I formally examine the intuition of the above example by comparing the memory activation model (M) with an exponentially decaying ad-stock model (E). For illustration, consider a simple case where an individual i decides whether to purchase a product. Assume this decision depends solely on the individual's exposure to ads before the purchase occasion through a goodwill metric G_i^J , which is defined differently below for the two models $J \in \{M, E\}$. I assume the utility i gets (ignoring the subscript i for now)

$$\begin{aligned} (\text{purchase})u^1 &= \alpha + \gamma^J G^J + \epsilon^1 \\ (\text{no purchase})u^0 &= \epsilon^0 \end{aligned}$$

where α represents preference for the product, γ^J is the sensitivity of the decision to goodwill accumulated due to advertising, and ϵ^1, ϵ^0 are idiosyncratic time-varying shocks assumed to be i.i.d. according to type 1 extreme value distribution. The probability of purchase, using goodwill definition from one of the models $J \in \{E, M\}$ is given by

$$p^J = \frac{\exp(\alpha + \gamma^J G^J)}{1 + \exp(\alpha + \gamma^J G^J)} \quad (6)$$

Exponentially decaying ad-stock model (E): As pointed out earlier, according to this model, every ad contributes to the total ad stock and its contribution decays exponentially over time. The accumulated ad stock affects the utility obtained by the individual from choosing the advertised product. If the individual i is exposed to n ads with ages a_{-1}, \dots, a_{-n} days before this purchase occasion, the goodwill stock due to ads is

$$G = \sum_k^n \rho^{a_{-k}}$$

where ρ ($0 \leq \rho \leq 1$) is the decay parameter, which is the same for each ad occasion and a_{-k} is

the age of the k^{th} last exposure for individual i .²³ Substituting this in equation (6),

$$\begin{aligned} p^E &= \frac{\exp(\alpha + \gamma^E (\sum_k \rho^{a-k}))}{1 + \exp(\alpha + \gamma^E (\sum_k \rho^{a-k}))} \\ \implies L_E(a_{-1}, \dots, a_{-n}, \rho | \alpha, \gamma^E) &= \ln\left(\frac{p^E}{1 - p^E}\right) = \alpha + \gamma^E \left(\sum_k \rho^{a-k}\right) \end{aligned} \quad (7)$$

where L_E is the log odds ratio of purchase for individual i according to model E .

Memory Activation model (M): Here I assume that the goodwill due to advertising is the memory activation level (m), which influences the purchase decision.²⁴ Therefore, from the same schedule of n ads, according to this model,

$$G^M = m = \beta + \ln\left(\sum_{k=1}^n a_{-k}^{-d_k}\right)$$

where d_k is the decay rate for the k^{th} last exposure for individual i . Substituting in equation (6),

$$\begin{aligned} p &= \frac{\exp\left(\alpha + \gamma^M \left(\beta + \ln\left(\sum_{k=1}^n a_{-k}^{-d_k}\right)\right)\right)}{1 + \exp\left(\alpha + \gamma^M \left(\beta + \ln\left(\sum_{k=1}^n a_{-k}^{-d_k}\right)\right)\right)} \\ \implies \ln\left(\frac{p^M}{1 - p^M}\right) &= \alpha + \gamma^M \beta + \gamma^M \ln\left(\sum_{k=1}^n a_{-k}^{-d_k}\right) \end{aligned}$$

Substituting $\chi = \alpha + \gamma^M \beta$,

$$L_M(a_{-1}, \dots, a_{-n}, d_1, \dots, d_n | \chi, \gamma^M) = \ln\left(\frac{p^M}{1 - p^M}\right) = \chi + \gamma^M \ln\left(\sum_{k=1}^n a_{-k}^{-d_k}\right) \quad (8)$$

where L_M represents log odds ratio for purchase for individual i according to the memory activation model.

In equations (7) and (8), advertising affects the log odds ratios L_M and L_E in different ways.

Different functional forms

First, note the ages of past ad occasions enter the log-likelihood expressions for the two models in different functional forms. To understand the differences in the model, apart from those that

²³Specifying different decay rates for different ad occasions (non-systematically) will not change the main argument.

²⁴Another way to think about this is a two stage decision process (where recall is necessary for purchase) and a higher level of activation leads to higher recall rate that influences the decision. In this case one can write the purchase probability $p = Pr(recall) \times Pr(purch | recall)$. Here, instead, I assume a direct impact of the activation on the decision to ease the comparison with standard exponentially decaying ad-stock model.

arise due to *past exposure dependent* decay rates ($d_k(t_1, \dots, t_{k-1})$) in M , I take partial derivatives of log-likelihood expressions (in equations (7) and (8)) with respect to the age of the k^{th} last exposure, a_{-k} :

$$\frac{\partial L_E}{\partial a_{-k}} = \gamma^E \rho^{a_{-k} \ln \rho} \quad (9)$$

$$\frac{\partial L_M}{\partial a_{-k}} = \gamma^M \frac{-(d_k + 1) a_{-k}^{-(d_k + 1)}}{\left(\sum_{k=1}^n a_{-k}^{-d_k} \right)} = \gamma^M \frac{-(d_k + 1) a_{-k}^{-(d_k + 1)}}{\exp(m - \beta)} \quad (10)$$

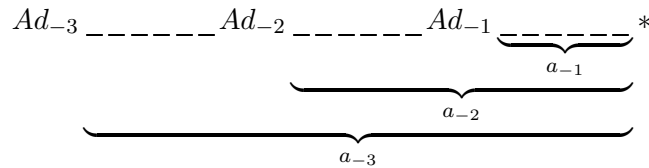
As age of the ad exposure increases, log odds of purchase decrease according to both models (since $\ln \rho$ is negative). One difference is that according to model M , if the memory activation m due to all advertising is high, then the slope magnitude decreases. This implies the change in log odds due to the change in age of an ad exposure might become insignificant if there is high enough activation due to total advertising. On the other hand, according to the exponential decay model E , the impact on the slope is independent of other ad exposures.²⁵ A more flexible rather than linear function of G in the model specification, however, would attenuate this difference.

Ad occasion specific decay rate in M

Now, to examine the different spacing implications of these models, I focus on total change in log odds due to change in age of ads. Note that for L^E , the total derivative is the same as the partial in (9). On the other hand, total change in L_M due to change in a_{-k} is the sum of the direct effect of change in age of the k^{th} last ad (10) and the indirect effect of changes in decay rates (d_{k-1}, \dots, d_1) of effects of ads occurring after the k^{th} last ad:

$$\frac{dL_M}{da_{-k}} = \frac{\partial L_M}{\partial a_{-k}} + \frac{\partial L_M}{\partial d_{k-1}} \cdot \frac{dd_{k-1}}{da_{-k}} + \dots + \frac{\partial L_M}{\partial d_1} \cdot \frac{dd_1}{da_{-k}}$$

For illustration, I consider a case of three ads Ad_{-1} , Ad_{-2} and Ad_{-3} of ages a_{-1} , a_{-2} and a_{-3} respectively ($a_{-3} > a_{-2} > a_{-1}$):



L_M are the corresponding log odds for purchase at purchase occasion according to (8).

²⁵If I specify goodwill for M as an exponential of m ($G^M = e^m$) instead of the linear form, this difference goes away.

The effect of change da_{-3} on L_M is

$$\frac{dL_M}{da_{-3}} = \frac{\partial L_M}{\partial a_{-3}} + \frac{\partial L_M}{\partial d_2} \cdot \frac{dd_2}{da_{-3}} + \frac{\partial L_M}{\partial d_1} \cdot \frac{dd_1}{da_{-3}}$$

The total effect has three components (assume $da_{-3} > 0$):

1. *Direct effect* due to change in a_{-3} according to (10). This effect will be negative since all else being equal, the effectiveness of an ad decreases with age.
2. *Effect through change in d_2* : As a_{-3} increases, the decay rate for Ad_{-2} decreases, leading to an unambiguous increase in the log-likelihood at the purchase occasion. To see how, note that since $da_{-3} > 0$, the relative time between Ad_{-3} and Ad_{-2} , $(a_{-3} - a_{-2})$ increases, so m_2 (activation at the time of Ad_{-2} , due to Ad_{-3}) decreases, thereby decreasing d_2 according to (5). This term can be expressed as

$$\frac{\partial L_M}{\partial d_2} \cdot \frac{dd_2}{da_{-3}} = \frac{\partial L_M}{\partial d_2} \cdot \left(\frac{\partial d_2}{\partial m_2} \cdot \frac{dm_2}{da_{-3}} \right)$$

and by (5)

$$m_2 = \beta + \ln \left((a_{-3} - a_{-2})^{-d_3} \right)$$

An increase in d_2 decreases the effectiveness of Ad_{-2} at the purchase occasion, so $\frac{\partial L_M}{\partial d_2} < 0$. The decay rate of Ad_{-2} increases with activation at that time, so $\frac{\partial d_2}{\partial m_2} > 0$. Activation due to Ad_{-3} at the time of Ad_{-2} decreases in age a_{-3} , so $\frac{dm_2}{da_{-3}} < 0$. These arguments imply the overall effect is positive. Also note that $\frac{dd_2}{da_{-3}} < 0$, that is, the decay rate of Ad_{-2} decreases with spacing between Ad_{-2} and Ad_{-3} .

3. *Effect through change in d_1* : This effect is caused by change in activation at the time of $Ad_{-1}(m_1)$ due to changes in a_{-3} and d_2 . Change in log-likelihood can be written as

$$\begin{aligned} \frac{\partial L_M}{\partial d_1} \cdot \frac{dd_1}{da_{-3}} &= \frac{\partial L_M}{\partial d_1} \cdot \frac{\partial d_1}{\partial m_1} \cdot \frac{dm_1}{da_{-3}} \\ &= \frac{\partial L_M}{\partial d_1} \cdot \frac{\partial d_1}{\partial m_1} \left(\frac{\partial m_1}{\partial a_{-3}} + \frac{\partial m_1}{\partial d_2} \cdot \frac{dd_2}{da_{-3}} \right) \\ &= \frac{\partial L_M}{\partial d_1} \cdot \frac{\partial d_1}{\partial m_1} \frac{\partial m_1}{\partial a_{-3}} + \frac{\partial L_M}{\partial d_1} \cdot \frac{\partial d_1}{\partial m_1} \cdot \frac{\partial m_1}{\partial d_2} \cdot \frac{dd_2}{da_{-3}} \end{aligned} \quad (11)$$

where by (5),

$$m_1 = \beta + \ln \left((a_{-3} - a_{-1})^{-d_3} + (a_{-2} - a_{-1})^{-d_2} \right)$$

The first term of RHS of equation (11) represents an increase in L_M due to a smaller decay rate d_1 because of a decrease in activation by Ad_{-3} . This term is positive using an explanation similar to that in step 2.

The second term represents the impact on L_M due to a change in d_1 resulting from a decreasing decay rate of Ad_{-2} . Mathematically, it is a product of $\frac{\partial L_M}{\partial d_1} < 0$ (same argument as before), $\frac{\partial d_1}{\partial m_1} > 0$ (same argument as before), $\frac{\partial m_1}{\partial d_2} < 0$ (activation due to Ad_{-2} decreases with its decay rate), and $\frac{dd_2}{da_{-3}} < 0$ (shown in step 2). So this term is negative. The overall effect due to change in d_1 can be positive or negative depending on the relative magnitudes of the two terms in (11).

Hence the total effect of change in age of Ad_{-3} can be expressed as

$$\begin{aligned} \frac{dL_M}{da_{-3}} &= \frac{\partial L_M}{\partial a_{-3}} + \frac{\partial L_M}{\partial d_2} \cdot \frac{dd_2}{da_{-3}} + \frac{\partial L_M}{\partial d_1} \cdot \frac{dd_1}{da_{-3}} \\ &= \underbrace{\frac{\partial L_M}{\partial a_{-3}}}_{\text{Due to } Ad_{-3}} + \underbrace{\frac{\partial L_M}{\partial d_2} \cdot \frac{dd_2}{da_{-3}}}_{\text{Due to } Ad_{-2}} + \underbrace{\frac{\partial L_M}{\partial d_1} \cdot \frac{\partial d_1}{\partial m_1} \left(\frac{\partial m_1}{\partial a_{-3}} + \frac{\partial m_1}{\partial d_2} \cdot \frac{dd_2}{da_{-3}} \right)}_{\text{Due to } Ad_{-1}} \end{aligned} \quad (12)$$

$$\begin{aligned} &= \underbrace{\frac{\partial L_M}{\partial a_{-3}}}_{<0} + \underbrace{\frac{\partial L_M}{\partial d_2} \cdot \frac{\partial d_2}{\partial m_2} \cdot \frac{dm_2}{da_{-3}}}_{>0} + \underbrace{\frac{\partial L_M}{\partial d_1} \cdot \frac{\partial d_1}{\partial m_1} \cdot \frac{\partial m_1}{\partial a_{-3}}}_{>0} + \underbrace{\frac{\partial L_M}{\partial d_1} \cdot \frac{\partial d_1}{\partial m_1} \cdot \frac{\partial m_1}{\partial d_2} \cdot \frac{dd_2}{da_{-3}}}_{<0} \end{aligned} \quad (13)$$

In other words, if the age of an ad increases, the ad's impact on the purchase occasion goes down unambiguously, which causes the effectiveness of the immediate next ad exposure (Ad_{-2}) to increase. These two changes play a role in the change of effectiveness of the subsequent ads. If the increase in activation due to Ad_{-2} outweighs the decrease due to a weaker effect of Ad_{-3} then the effectiveness of Ad_{-1} decreases. Otherwise, it increases.

This analysis shows that according to the proposed memory activation model M , by increasing the spacing between the past ad occasions some schedules of ads can be improved for the advertiser. This implication is driven by the assumption that the decay rate of the effect due to a particular ad (say Ad_{-k}) depends on the ads occurring prior to Ad_{-k} ($\frac{\partial d_k}{\partial m_k} > 0$). In terms of the notation used to define the model M , this phenomenon is caused by the parameter c (equation (4)); if $c = 0$, the last three terms of equation (13) vanish since $c = 0 \implies \frac{\partial d_k}{\partial m_k} = 0$, $k \in \{1, 2\}$.

The simple version of the distributed lag model considered here does not predict this behavior. Intuitively, specifying a different set of weights to lagged ad exposures will still not predict spacing behavior unless the effectiveness of ads in some way depends on the spacing of the past schedule of ads (which drives the positive terms in equation (13)).

7.3 Simulations

I simulate different scenarios to examine the shape of the log-odds function L^M , comparing it with L^E , and the impact of the spacing parameter c (in equation (4)) on it. Again, for illustration, I use the scenario where a consumer is exposed to three ads, Ad_{-1} , Ad_{-2} , and Ad_{-3} that are of ages a_{-1} , a_{-2} , and a_{-3} respectively ($a_{-3} > a_{-2} > a_{-1}$), at the purchase occasion.

Effect of changing a_{-3}

Figure 9 shows how log of purchase odds at the current occasion vary for the two models with age of the oldest exposure, for the set of specified parameters. Here I fix $a_{-2} = 10$ and $a_{-1} = 5$ and vary a_{-3} from 11 to 30. Note that

- L_E , the log odds according to the exponentially decaying ad stock model decreases as age of Ad_{-3} increases.
- L_M , according to the memory activation model, is nonmonotonic. It increases initially and then decreases because of the two forces noted above: a decrease in direct impact of Ad_{-3} due to increasing age, and a positive effect due to an increase in the effectiveness of Ad_{-2} and Ad_{-1} as Ad_{-3} moves away. The latter effect is stronger when $a_{-3} - a_{-2}$ is small and decreases as a_{-3} increases and eventually gets dominated by the former around $a_{-3} = 15$.

Effect of changing the parameter c

As c increases, the impact of contemporaneous activation levels on decay rates increases. Therefore, one would expect a more prominent spacing effect when c is high and no effect when $c = 0$. Figure 10 shows simulation results for different values of c when other parameters are held fixed. As expected, when c is zero, an increase in age of Ad_{-3} causes a monotonic decrease in log odds. When $c > 0$, pushing Ad_{-3} further back provides some benefit for the advertiser. The range for which $\frac{dL_M}{da_{-3}} > 0$ is larger for larger values of c .²⁶

8 Structural estimation

In this section, I apply the above framework to the online sponsored search advertising setting described earlier. While estimating the model parameters for counterfactual analysis, I also test for the key implications of the memory model M .

8.1 Empirical specification

I start with the specification of an exponential decay goodwill stock model (E) that has a constant decay rate for an individual; that is, ad stock built by any ad exposure has the same decay rate. Using this model as a benchmark, I generalize it to a model GE , where an extra parameter enables the decay rate of ad stock contributed by any ad to depend on contemporaneous ad stock from the past. As explained earlier, this is the key feature of the memory activation model that raises the possibility of spacing effects. Since the exponential decay goodwill model

²⁶Note that since a is kept constant, an increase in c leads to a lower average carry-over for Ad_{-2} and hence lower L_M levels for high c .

is nested within this generalized model, estimating GE will provide a test in the data for this key assumption. Eventually, I specify the empirical version of the memory activation learning model (M) discussed above.

8.1.1 Choice model

I assume the utility an individual i ($i \in \{1, \dots, I\}$) gets from visiting the advertiser's web page at purchase occasion (session) t ($t \in \{1, \dots, T_i\}$) depends on the number of ad exposures ($nExp_{it}$) in the session t and the advertising effect carried over from the past sessions, G_{it}^ψ for $\psi \in \{E, GE, M\}$. It also depends on the inherent search propensity, which includes search cost, familiarity with the web, and so on. This factor is unobserved in the data and may be correlated with $nExp_{it}$ since $nExp_{it}$ is high for individuals who search more. As in earlier analysis (section 5.2), to avoid omitted variables bias in an estimate of the marginal effect of $nExp_{it}$, I include in the model the number of pages i decides to browse in session t . Other than these factors, the likelihood of visiting the web page will also depend on the individual's inherent propensity to click on the experimental advertiser's page (α_i^ψ). Therefore,

$$\begin{aligned} u_{it}^{1\psi} &= \alpha_i^\psi + \delta_i^\psi f(nExp_{it}) + \gamma_i^\psi h(G_{it}^\psi) + \phi_i^\psi nPages_{it} + \sum_{j=1}^J \tau_j^\psi x_{ij} + \epsilon_{it}^{1\psi} \\ u_{it}^{0\psi} &= \epsilon_{it}^{0\psi} \end{aligned}$$

where $u_i^{1\psi}$ represents the utility gain from visiting the advertiser's web page and $u_i^{0\psi}$ is the utility from choosing the outside option. Since different experiments used different advertising restaurants, I control for differences in preferences across experiments by including fixed effects in the model. x_{ij} is a dummy variable that indicates individual i lies in experiment j .²⁷ Functions $f(\cdot)$ and $h(\cdot)$ allow flexibility in how $nExp_{it}$ and G_{it}^ψ affect utility.²⁸ $\epsilon_{it}^{1\psi}$ and $\epsilon_{it}^{0\psi}$ are idiosyncratic error shocks that are assumed to be i.i.d. according to type 1 extreme value distribution. $\theta_i^\psi = \{\alpha_i^\psi, \delta_i^\psi, \gamma_i^\psi, \phi_i^\psi, (\tau_j^\psi)_{j=1}^J\}$ are model ψ specific parameters to be estimated using data.

8.1.2 Model for carry-over

Carry-over from past occasions, G_{it}^ψ for model $\psi \in \{E, GE, M\}$, is a function of the number of prior ad exposures, the time intervals between them, and model-specific parameters ξ_i^ψ .

Exponential decay goodwill stock model (E)

²⁷Which in many cases is a different geographic area.

²⁸I start with an identity function and then allow for more flexible polynomial transformations.

G_{it}^E is the discounted sum of effects due to ad exposures in sessions prior to t . I define it as

$$\begin{aligned} G_{it}^E &= \rho_i^{days_{(t-1),t}} (nExp_{it-1} + G_{it-1}^E) \\ &= \sum_{k=1}^{t-1} \rho_i^{days_{kt}} nExp_{ik} \end{aligned}$$

where $\rho_i \in (0, 1)$ is the discount rate parameter and $days_{kt}$ is the time interval between sessions k and t in days.

Generalized exponential decay goodwill model (GE)

G_{it}^{GE} is again a discounted sum of past effects of ad exposures in sessions prior to t , defined as

$$G_{it}^{GE} = \sum_{k=1}^{t-1} \rho_{ik}^{days_{kt}} nExp_{ik}$$

where $\rho_{ik} \in (0, 1)$ is the carry-over rate for ad stock accumulated through ads in session k and $days_{kt}$ is the time gap between sessions k and t in days. I parameterize the discount rate, $\rho_{ik} (\xi_i^{GE} = \{b_i^{GE}, c_i^{GE}\})$ as

$$\rho_{ik} = \frac{\exp(b_i^{GE} + c_i^{GE} G_{ik}^{GE})}{1 + \exp(b_i^{GE} + c_i^{GE} G_{ik}^{GE})} \quad (14)$$

According to equation (14), the discount rate for a past ad occasion k depends on the total goodwill accumulated from ads before session k . Note that this model reduces to E if $c_i^{GE} = 0$. If parameter $c_i^{GE} < 0$, carry-over ρ_{ik} will decrease with increase in contemporaneous goodwill G_{ik}^{GE} .

Memory-activation model (M)

To apply the model in this empirical context, I use a modified, discrete time version of the memory activation model discussed earlier. I specify a browsing session as a learning occasion with time intervals in days. The number of ad exposures in the session represents the degree of learning.²⁹

Following equation (3), I define the strength of memory trace for the advertised restaurant for individual i before a session t as given by

$$m_{it}(\{nExp_{ikt}\}_{k=1}^{t-1}, \{days_{ikt}\}_{k=1}^{t-1}) = \beta_i + \ln \left(\sum_{k=1}^{t-1} (nExp_{ik} \times days_{ikt}^{-d_{ik}}) \right) \quad (15)$$

where β_i represents individual-specific ease of memory activation, $days_{ikt}$ is the time interval

²⁹In general, advertising context is different from situations where agents learn at every occasion, for example, verbal learning (Pavlik and Anderson [2008]) or forced exposures to TV ads in a lab. In a real-world setting, an ad exposure may not mean that consumer “saw” the ad. So for my case of online advertising, one ad exposure may not be the correct learning occasion for the model.

between sessions t and k for individual i , $nExp_{ik}$ is the number of exposures in session k , and d_{ik} is the decay rate for activation due to ads in session k .³⁰ d_{ik} depends on parameters $\xi_i^M = \{b_i^M, c_i^M\}$ and the activation just before session k started. I specify it as

$$d_{ik}(m_{ik}, \xi_i^M) = \exp(b_i^M + c_i^M e^{m_{ik}}) \quad (16)$$

where I use an exponential function so the decay rate is always positive. I define goodwill as a function of the memory activation, $G_{ik}^M = g(m_{ik})$, and compare estimates for different non-decreasing functions $g(\cdot)$.

8.2 Distributional assumptions and identification of parameters

Because the number of sessions browsed by individuals in data is small, I am unable to identify individual-specific parameter vectors θ_i^ψ and ξ_i^ψ . So I assume a distribution of these parameters in the population and identify the location and spread of the distribution. Specifically, I assume a Normal distribution

$$\begin{pmatrix} \theta_i^\psi \\ \xi_i^\psi \end{pmatrix} \sim N \left(\begin{pmatrix} \bar{\theta}_\psi \\ \bar{\xi}_\psi \end{pmatrix}, \Sigma_\psi \right) \quad (17)$$

where $\begin{pmatrix} \bar{\theta}_\psi \\ \bar{\xi}_\psi \end{pmatrix}$ is the population mean parameter vector and Σ_ψ is the variance covariance matrix of the parameters in the population.

In the data, I observe variation in covariates across individuals and time, and the change in choice behavior of the agents at these occasions. This variation identifies all components in the mean vector of choice parameters $\bar{\theta}_\psi$, except $\bar{\gamma}_\psi$ (the coefficient of G^ψ). Given the carry-over parameters ξ_ψ , variation in the intensity of ad exposure in the past causes different levels of ad carry-over from the past (G^ψ), which identifies the mean $\bar{\gamma}_\psi$.

The first component of $\bar{\xi}_\psi$, that comprises of the mean carry-over parameters $\bar{\rho}$, \bar{b}_{GE} and \bar{b}_M , determines the degree of carry-over effect in the absence of prior ads. Because there was no advertising before the first session, these parameters are identified from systematic differences in choices made in the first and second sessions, conditional on ads and other session characteristics. Similarly, the systematic change in choice behavior from the second to the third sessions in the ad condition is attributed to the changed decay rates of ad stock accumulated during the second session. This variation identifies parameters \bar{c}_{GE} and \bar{c}_M .

The time-series aspect of the data helps identify the variance covariance matrix Σ_ψ . For example, the presence of individuals in the data with different responses to ads in previous sessions

³⁰Since ease of learning and the impact of G_{it}^M on choice affect the decision in the same way, identifying them is difficult. So I fix $\beta_i=0$.

identifies the variation in the parameters ρ_i , b_i^{GE} and b_i^M .

8.3 Estimation methodology

I estimate the model parameters using the simulated maximum likelihood method. Given the parameters and data ($X_{it} = \{nPages_{it}, \{nExp_{ik}\}_{k=1}^t, \{days_{ik}\}_{k=1}^t, \{x_{ij}\}_{j=1}^J\}$), the probability of an individual i visiting the advertiser's page in session t ($y_{it}=1$) according to model $\psi \in \{E, GE, M\}$ is given by

$$\begin{aligned} Pr_{it}^\psi &= Pr(y_{it} = 1 | X_{it}, \theta_i^\psi, \xi_i^\psi) \\ &= \frac{\exp(\alpha_i^\psi + \delta_i^\psi f(nExp_{it}) + \gamma_i^\psi G_{it}^\psi + \phi_i^\psi nPages_{it} + \sum_{j=1}^J \tau_j^\psi x_{ij})}{1 + \exp(\alpha_i^\psi + \delta_i^\psi f(nExp_{it}) + \gamma_i^\psi G_{it}^\psi + \phi_i^\psi nPages_{it} + \sum_{j=1}^J \tau_j^\psi x_{ij})} \end{aligned}$$

The likelihood of the sequence of observed decisions $y_i = \{y_{it}\}_{t=1}^{T_i}$ made by i is then

$$\begin{aligned} \pi_i^\psi(y_i | X_i, \theta_i^\psi, \xi_i^\psi) &= \prod_{t=1}^{T_i} (Pr_{it}^\psi \times y_{it} \\ &\quad + (1 - Pr_{it}^\psi) \times (1 - y_{it})) \end{aligned}$$

where T_i is the total number of observed sessions for individual i . Next, because the parameters $\{\theta_i^\psi, \xi_i^\psi\}$ are random, I compute the expected likelihood by integrating the individual's likelihood over the distribution of random parameters:

$$\bar{\pi}_i^\psi(y_i | X_i, \bar{\theta}_\psi, \bar{\xi}_\psi, \Sigma_\psi) = \int \pi_i^\psi d\Omega(\theta_i^\psi, \xi_i^\psi)$$

where $\Omega(\cdot)$ is the density function for the distribution defined in (17).³¹ So the log-simulated likelihood of observed data conditioned on parameters is given by

$$LL^\psi(\{y_i, X_i\}_{i=1}^I, \bar{\theta}_\psi, \bar{\xi}_\psi, \Sigma_\psi) = \sum_{i=1}^I \log(\bar{\pi}_i^\psi)$$

The estimated point estimates of parameters for the model are the ones that maximize the log-likelihood function.

8.4 Results

I estimated multiple specifications for the models discussed above. For estimation, I removed data for the restaurant that had advertised before the time period of the experiments. Also, I removed

³¹The integration is implemented using a Monte Carlo simulation method; I take draws of θ_i^ψ and ξ_i^ψ from the distribution and compute the average value of π_i^ψ for these draws.

individuals that had consecutive sessions with time intervals of a day or less.³² Therefore, the estimation sample comprises of 9,081 individuals and 20,959 sessions.

First, I discuss the results from estimation of homogenous model specifications with no unobserved heterogeneity in parameters across the population (where $\Sigma_\psi = 0$). Column I of Table 8 shows results for baseline specification of E , where carry-over from the past, G enters the utility linearly. Estimates show the number of ad exposures in the current session ($nExp_t$) has a positive impact on the likelihood of the individual visiting the advertiser’s page. Also, more pages browsed in a session lead to a higher chance of visit to the page. The estimates also suggest a high carry-over of ad-stock across time with a decay rate of 0.97 per day or ~ 0.80 per week. However, the coefficient on the goodwill stock (G_{it}^E) carried over from the previous period is positive but not statistically significant. Motivated by earlier reduced form data analysis (Table 5, column II), I allow for quadratic and cubic terms of goodwill G_{it}^E . Estimates of these model specifications in columns II and III, respectively, indicate behavior similar to the first specification, except the carry-over effect is now more precise. The cubic specification in column III suggests goodwill affects utility only at higher levels.

Column IV shows estimates for the baseline specification of the generalized exponential decay ad-stock model (GE). As predicted by theory discussed in the last section, the parameter c^{GE} is negative but statistically not significantly different from 0. Allowing for this change in decay rate across sessions, however, leads to an increase in size of the coefficient on goodwill carry-over from past sessions (compared to column I), which is now statistically significant. Other estimates are comparable to the corresponding model E in column I. In column VI, I allow for the marginal effect of past goodwill to change with its level by including a quadratic and cubic terms of G_{it}^{GE} . The positive coefficient of this quadratic term indicates the impact of previous goodwill stock is higher at higher levels.³³ Importantly, now the negative parameter c^{GE} is estimated precisely, statistically different from zero, providing support for the assumption (from equation 14) that goodwill stock accumulated in a session decays faster if contemporaneous goodwill carry-over from the past is high. Other parameter estimates are similar to the model E . The log-likelihood estimates for the cubic specifications of GE and E show the improvement in data fit by adding an extra parameter is marginal (1862.1 - 1861.92). This finding is not surprising given that about three quarters of the observations in the data are for individuals with just two sessions during the time period. These observations are not affected by previous spacing, hence their fit cannot improve by the added flexibility in GE . However, for the individuals that are affected by past spacing, I find a significant increase in log-likelihood; for 763 session observations, the log-likelihood changes from -105.7 to -103.9.³⁴ Moreover, GE is better at predicting the behavior for the sample held out from estimation, that is, individuals that have at least 1 instance of sessions

³²This is done to avoid the impact of variety-seeking behavior.

³³For small levels, the total effect is not statistically different from zero.

³⁴These are the individuals included earlier in Figure 7; ones with more than two sessions in the data and more than two ad exposures in the second session.

with time gap of one day. The predicted log-likelihood for 2,407 sessions is -1001.7 for E and -961.3 for GE .

Columns VII to IX of Table 8 show estimates of similar specifications for the memory activation model M .³⁵ As expected, the estimates are qualitatively similar to GE . In the models where quadratic and cubic terms for past goodwill are allowed, parameter $c^M > 0$ and is statistically significant at the 10% confidence level. This estimate suggests the decay rate of memory activation due to an ad increases if the contemporaneous activation level is high.

To understand the economic significance of these estimates, I compare the long-term effects of advertising at a particular occasion predicted by models E and GE . Consider a situation where the individual visiting the website in the current time period saw the ad banner 15 days ago. Figure 11 shows the predicted impact over time, of displaying the ad again in the current session. Using the model GE , the first plot shows the predicted probability of visiting the advertiser's web page in two conditions: when the ad banner is displayed (I assume 3 exposures) and when it is not. This simulation is done using point estimates in column VI of Table 8. The second plot shows the same scenario simulated for model E using estimates from column III. Note that in the ad condition, both models predict a similar increase in probability of visiting the web page in the immediate future. This increase, however, decays faster according to the model GE due to the increased decay rate because of prior ad exposure. This example illustrates that in the presence of prior awareness due to advertising, GE predicts the benefit from advertising will vanish sooner (in this case, within 2 weeks), whereas model E predicts a longer-term benefit.

Table 9 shows the estimates when unobserved heterogeneity is accounted for.³⁶ Column I shows results for E . The mean values of the random parameters in the population are similar to the homogenous model. The standard deviations estimated are not statistically significantly different from zero except for the intercept, suggesting a sizable spread in the preferences for the advertised restaurants. Column II shows estimates for the baseline specification for GE . The population mean of coefficient c_{GE} is statistically significant and negative, supporting the assumption that high advertising exposure in the past can lead to a decrease in carry-over of the current ad exposure. The population mean for the coefficient for G is not statistically significantly different from zero, but the estimate for standard deviation is precise and large relative to point estimates seen in Table 8. To examine this heterogeneity further, I allow the impact of carry-over to vary with observable factors by adding a quadratic term and the interaction of G with other variables such as experiment indicators x_1, x_2, x_3 and $nExp$.³⁷ Column III shows estimates for this model. Note that the standard deviation for the random coefficient of G is now small and insignificant and the coefficient on the quadratic term is significantly positive. The interaction term $nExp \times G$ is negative but statistically insignificant.

³⁵The $h(\cdot)$ function used here is exponential; that is, the memory activation enters the utility function via an exponential transformation: $G_{it}^M = e^{m_{it}}$

³⁶For now, I restrict Σ_ψ to be a diagonal matrix.

³⁷to allow for the marginal effect of advertising to change with awareness carried over from past

To summarize, the data support the hypothesis that persistence of advertising decreases in the presence of contemporaneous awareness carried over from past occasions. The additional parameter in the generalized exponentially decaying goodwill stock model, which nests the predominantly used model E within it, is statistically significant and in the direction predicted by memory research. Estimates from the structural models suggest that in presence of past advertising, E may significantly over-predict the long-term effect of additional advertising. This tradeoff may provide incentives for the advertiser to spread the advertising activity over a wider time period. To explore this aspect further, in the next section, I use the demand parameter estimates to compare various advertising strategies.

8.5 Implications of the models for advertising strategies

In general, online media plans specify the number of impressions to be bought from an ad network during a particular time period. One of the ways advertisers control repeated exposure for their online ads on networks such as Google AdWords is through the use of frequency capping. The term "frequency capping" refers to restricting (capping) the number of times (frequency) a specific visitor is shown a particular advertisement within a period of time. For example, a frequency cap of "10 per week" for an ad means that after exposing the user to the same ad 10 times, the visitor will not be shown that ad for the rest of the week.

In this section, I simulate outcomes for an advertiser from applying various advertising strategies in a simple scenario, to choose the best strategies according to the models E and GE . Consider a scenario where a consumer visits the website in every week for a month. Suppose the advertiser knows the browsing behavior of the consumer and has the option of choosing a frequency cap and the weeks of the month when advertising is switched on. Also, for now I assume the advertiser cares about the gains from advertising during this one month only. If the consumer browses ten pages in each of the four sessions in the four weeks of the month, the number of possible choices of a freq. cap is 11, from 0 to 10, and there are 2^4 possible combinations of switching the advertising on or off in any of the weeks. Therefore, the total number of possible strategies for the advertiser in this situation are

$$\underbrace{\# \text{ possible freq. caps}}_{11} \times \underbrace{\text{possible combination of ads switched on or off for 4 weeks}}_{2^4} = 176$$

For each of the strategies, I simulate the total profits for the month in both the cases, when the advertiser uses the model E or GE .³⁸ The best strategies, giving highest profits for the month according to the two models are shown in Figure 12. If the advertiser uses the traditional exponential decay model, the highest profits are attained by choosing a frequency cap of 7, and

³⁸For these calculations, I assume the profits resulting from the restaurant page visit to be \$10 and the marginal cost of an ad impression to be \$0.1.

advertising is employed in three out of the four weeks. After advertising in the first two weeks, the goodwill level is high and the marginal benefit from more advertising goes down, so during the third week, no ads are shown. On the other hand, an advertiser using GE would choose a higher frequency cap of 10 and advertise in just one of the four weeks. According to this model, after advertising in the first week the goodwill level is high during the later weeks leading to lower returns from advertising at the later occasions because of reduced persistence of additional advertising. Therefore, compared to the traditional model, in this case the advertiser chooses a higher level of advertising, but advertises in the first week only. In this example, if the true model is GE (as I find in my data) and an advertiser wrongly uses E to choose her advertising strategy, then she gets 70% lower profits compared to another advertiser who selects ad schedule based on GE . This happens because the gains from advertising in the later weeks resulting from a strategy based on E are too low to justify the costs and therefore lead to significantly lower profits.

9 Conclusion

This paper contributes to the empirical literature on advertising in several ways. To ensure the identification of various advertising effects, I ran field experiments to create a novel individual-level dataset that provides detailed information on consumer search leading to purchase and contains exogenous variation in ad exposure. Using these data, I first provide model-free evidence on the impact of advertising, and carry-over of ad effects to future purchase occasions. I find that spacing or the temporal proximity between ad exposures plays an important role in the impact of ads. I show the impact of allowing for the spacing effect on advertising strategy by estimating a memory-based model of learning through advertising. Using counterfactual simulations I demonstrate that the new model has different implications for advertisers compared to the traditionally used models.

An important direction for future work is the investigation of the impact of the memory-based model on dynamically optimal advertising policies. Advertiser’s gains from spacing of ads, allowed by the memory-based model, might provide an additional explanation for the observed use of pulsing advertising strategies.³⁹ The website engaged in the current experiments also provides a potential avenue for eventually implementing the optimal strategies according to different models and comparing the actual returns, and further validating the use of the new model. However, a challenge in solving for the dynamically optimal policies under the memory model comes from the increased complexity of the state space of the advertiser’s problem. Unlike the traditional models where the carry-over to the future depends only on the goodwill level at a point of time,

³⁹Pulsing ad schedules are characterized by weeks with high levels of advertising followed by periods with no ads (Mahajan and Muller [1986], Dube et al. [2005]). In the past, researchers have explained pulsing behavior by deriving optimal advertising schedules in the presence of a threshold in the consumer’s ad response curve.

the future carry-over in the memory model depends on the entire past ad schedule leading to a potentially infinite dimensional state space.

This paper also sets up a “platform” to answer more questions about whether and how advertising characteristics affect consumer decisions.⁴⁰ One direction for further research is the impact of a firm’s advertising on its competitors’ demand. This spillover of ad effects might play an important role in understanding advertising dynamics in the economy. The experimental variation (condition A5) also allows me to study the impact of the clutter caused by competitor’s advertising - the change in advertising effects when more competitors advertise in the same session.

In summary, this paper takes a first step towards understanding the role of temporal spacing between ad exposures in consumer response to advertising. Future work can build on these findings to investigate their impact on firm’s strategy and related issues such as impact on competitive outcome. Data generated from the experiments in the current paper provide a good basis for exploring these issues.

⁴⁰Characteristics such as the location of the ad in the right-hand panel, presence of discounts and identity of the competitors.

Appendix

A Data-cleaning details

In this section, I describe the criteria used to clean data for analysis.

- *Bots / web-crawlers*: I use the fact that most of internet web crawlers originate outside India and browsers' i.p. addresses can be matched to their countries with good accuracy. To avoid noise in data due to automated web crawlers, I remove users with i.p addresses outside India. This process is reasonable for filtering bots since the website is designed to cater to individuals within the country.
- *Users browsing multiple zones*: I also remove users that browse zones across experiments. Since I'm interested in understanding the effect of multiple exposures to ads, this step is needed to avoid a possible selection problem. For example, consider two individuals A and B who both browse 10 pages on the website. The pages A browses are relevant to geographic area X, and she is exposed to 5 targeted ads. B browses 10 pages, but just 5 of them are relevant to zone X, and she is exposed to 2 ads targeted to this zone. In my data, B will be an individual with fewer ad exposures but also different preferences than A. By removing individuals such as B, I avoid a bias that can arise from comparing people with not just different ad exposure but also potentially different preferences.

B Experiment Details

Table 10 summarizes the number of sessions in different experiments and conditions. Figure 2 shows an example of how the ads in the right-hand panel change with different conditions described in section 4.

C Effect of Multiple Exposure to Ads within a session

In this section, I extend the earlier regression analysis in section 5.2. To better control for heterogeneity in browsing type, I put in more controls for number of pages ($nPages$). Column I of Table 11 shows results when the regression is run including only a subset of sessions with number of pages between 6 and 10. For flexibility, I use indicators of different levels of exposures as independent variables and control for number of pages by adding fixed effects for each level. Estimates show the effect of exposure on the likelihood of a visit is higher when the number of exposures is higher. In column II, I show estimates when I restrict the data to sessions

with exactly seven pages. Again, the coefficient on the number of exposures remains positive and statistically significant, suggesting more exposure to ads might lead to a higher impact on consumer decisions.

To appreciate the quantitative significance of multiple exposures, I simulate likelihood probabilities for different levels of exposures from the quadratic model estimated in column III of Table 4. Figure 13 shows the simulated probabilities for a session with 10 pages browsed ($nPages = 10$), plotted against the number of ad exposures in the session. The figure also shows the 25th and 75th percentiles of the distribution arising as a result of sampling errors. First note that the predicted likelihood at four exposures is significantly greater than the prediction for one exposure. For the first three exposures, there is an increase of about 0.4% for every exposure. After three, the marginal effect goes down, and after five exposures, additional ads result in little gain.

D More on the measurement issue

In this section I discuss in detail, the potential measurement problem mentioned earlier in section 5.2.1 and argue that the assumptions made for the current analysis are weak.

Recall that the experiment design has two levels of randomization.

- First, every new session is allocated to the Ad or the No-Ad condition. In the No-Ad condition, none of the pages browsed in the session show the experimental ad.
- Second, within the Ad condition there is a 60% chance that any page shows the experimental ad banner.⁴¹

Let the response variable of interest be Y (visiting the experimental restaurant’s page or generation of a sales lead). Difference of the mean of Y in the Ad and the No-Ad condition gives an unbiased estimate of the average treatment effect in the population (*treatment* is a 60% chance of seeing the ad). This average effect is the parameter γ in the regression

$$Y_i = \text{intercept} + \gamma \text{Ad}_i + \epsilon_i \quad (18)$$

where Ad is a dummy indicator of the treatment or the session being in the Ad condition.

Measuring the effect of an ad exposure

Now within the ad condition, there is variation in the number of times the ad is seen, conditional on the number of pages browsed. This variation allows me to get an estimate of the population

⁴¹For more than half of the duration of the experiments, the chance of getting an ad display at a page in the ad condition was set to 60%. For later time periods it was changed. For the purpose of this section, I focus on the time period when the chance was fixed to 60%.

average effect of an ad exposure controlling for the browsing type, by estimating a model M such as

$$Y_i = c + \alpha nExp_i + \beta nPages_i + \eta_i \quad (19)$$

where $nExp_i$ is the number of times the experimental ad was displayed in the session and $nPages_i$ is the number of pages browsed in the session.

Note that the true model of the effect of ads on individual's decision may be

$$Y_i = c + \alpha_i nExp_i + \beta nPages_i + \eta_i$$

where the decision of any individual i depends on the browsing type $nPages_i$ and has his own sensitivity toward number of ads shown, α_i . Therefore, a problem with estimating model M is that the independent variables, $nExp_i$ and $nPages_i$ are chosen by the consumer i and may depend on the preferences of the individual; $\alpha_i \rightarrow nPages_i, nExp_i$. Estimation of the regression (19) assumes this away, which may lead to a bias.

- Example: Conditioned on the number of pages browsed, sessions that end up seeing the ads larger number of times may be for people that did not respond for smaller levels of $nExp$ and therefore did not quit searching. These individuals are likely to be ones who are less sensitive to the ad. Because of this selection, the model will underestimate the effect of large $nExp$.

So in the regression (19), the estimated parameter $\hat{\alpha}$ may under-estimate the true $\alpha = E(\alpha_i)$. Let $\hat{\alpha} = \alpha + a$, where a is the magnitude of the bias.

Impact of Ad condition on distribution of $nPages$

The key reason behind this measurement problem is that the number of pages browsed may change in the ad condition, which may change the number of ad exposures too. In order to gauge this problem, I compare the empirical distributions of the number of pages browsed by sessions in the Ad and the No-Ad conditions. I use the two-sample Kolmogorov-Smirnov non-parametric test for this purpose and *fail* to reject the hypothesis (p-val=0.99) that the distributions are same. Therefore in these data, the assumption that the number of pages browsed in the ad condition is same as no-ad condition is not very strong.

Now the question - can the magnitude of the bias, 'a' be gauged?

The idea is to use the estimates of M to compute the predicted average population treatment effect $\hat{\gamma}_M$, and compare it with the unbiased estimate of the average population treatment effect $\hat{\gamma}$ from the regression (18).

$$\hat{\gamma}_M = E\left(\hat{Y}_{\text{Treatment}} - \hat{Y}_{\text{No-Treatment}}\right)$$

Where

- $\hat{Y}_{\text{Treatment}}$ is the predicted Y by the model estimates when a person is treated with a 60% chance of seeing the ad
- $\hat{Y}_{\text{No-Treatment}}$ is the predicted Y by the model estimates when there is a 0% chance of seeing the ad

The expectation is taken over the population.

What will comparing $\hat{\gamma}$ and $\hat{\gamma}_M$ show?

$\hat{\gamma}_M$ is a function of a .

$$\begin{aligned}
\hat{\gamma}_M &= E \left(\hat{Y}_{60\% \text{ chance of ads}} - \hat{Y}_{0\% \text{ chance of ads}} \right) \\
&= E \left(E \left(\hat{Y}_{60\% \text{ chance of ads}} - \hat{Y}_{0\% \text{ chance of ads}} | nPages \right) \right) \\
&= E \left(E \left(\hat{Y}(nExp, nPages) - \hat{Y}(0, nPages) | nPages \right) \right) \\
&= E \left(E \left(\hat{c} + \hat{\alpha}nExp + \hat{\beta}nPages - \left(\hat{c} + \hat{\beta}nPages \right) | nPages \right) \right) \\
&= E \left(E \left(\hat{\alpha}nExp | nPages \right) \right) \\
&= \hat{\alpha} \times E \left(E \left(nExp | nPages \right) \right) \\
&= (\alpha + a) \times E \left(E \left(nExp | nPages \right) \right) \\
&= (\alpha + a) \times E \left(nExp \right) \\
&= \alpha \times E \left(nExp \right) + a \times E \left(nExp \right) \\
&= \gamma + a \times E \left(nExp \right)
\end{aligned}$$

Using law of iterated expectations and the last step follows because $E(\alpha \times nExp)$ is the true average treatment effect γ . This analysis shows if a is large, $\hat{\gamma}_M$ is more likely to be off the true value γ . If $\hat{\gamma}$ and $\hat{\gamma}_M$ are close, i.e. the model predicts the average treatment effect to be close to the unbiased estimator, the bias size may not be too big.

Estimating $\hat{\gamma}_M$

The model assumes the number of pages browsed and the number of ads seen are independent of the preferences. Therefore I proceed as follows:

- From the population in the No-Ad condition, I get the empirical distribution of $nPages$.
- For individual $i \in \{1, \dots, I\}$, I draw $nPages_i$ from this empirical distribution.
- As in the experiment design, i draws its treatment $nExp_i \sim \text{Binomial}(nPages_i, 0.60)$

For this simulated population of I individuals, average treatment effect is

$$\hat{\gamma}_M = \frac{1}{I} \sum_i \left(\hat{Y}_i(nExp_i, nPages_i) - \hat{Y}_i(0, nPages_i) \right)$$

The comparison

When M is the linear regression model exactly same as (19), I get

Y	$\hat{\gamma}$		$\hat{\gamma}_M^*$	
	Estimate	95% Conf. interval	5 th percentile	95 th percentile
<i>Visit</i>	0.238%	(0.093%, 0.38%)	0.31%	0.33%
<i>View_num</i>	0.074%	(0.018%, 0.13%)	0.077%	0.080%

* Simulated 1000 samples of $I = 10000$ each. Reporting the percentiles from the distribution of estimated $\hat{\gamma}_M$ from all samples

When I use a logistic regression model for M , I get the following

Y	$\hat{\gamma}$		$\hat{\gamma}_M$	
	Estimate	95% Conf. interval	5 th percentile	95 th percentile
<i>Visit</i>	0.238%	(0.093%, 0.38%)	0.21%	0.23%
<i>View_num</i>	0.074%	(0.018%, 0.13%)	0.045%	0.050%

Note that estimated $\hat{\gamma}_M$ is within the 95% confidence intervals for $\hat{\gamma}$ for all the above cases. Therefore, this analysis suggests the potential bias a is not large relative to the absolute value of the population mean α .

E Non-parametric Evidence for the Impact of Ad Repetition

To provide evidence for an incremental impact of multiple exposures to ads, I use the fact that by experiment design, experimental ad display at any page is random during a session. Consider the set of individuals that browsed n or more pages in their first sessions. I focus on the marginal effect of an ad exposure at the n^{th} page conditional on treatment in the first $n - 1$ pages, for the subset of individuals that saw the experimental banner at least once in the first $n - 1$ pages. Specifically, I look for the effect of an additional ad at the n^{th} page on $Visit_{n+1,n+2}$, a dummy variable indicating a visit to the advertiser's page in the next two pages.⁴² This effect can be

⁴²I chose to focus on 2 pages ahead because after clicking on the ad users may take 2 pages to reach the restaurant page. Recall that if the ad is that of a chain restaurant, clicking on it takes the user to a page with links to pages for outlets in the geographic area of search.

identified through the regression

$$Visit_{n+1,n+2} = \beta_0 + \beta_1^n Ad_n + \beta_2 nExp_{1,n-1} + \epsilon \quad (20)$$

conditional on $nExp_{1,n-1} \geq 1$, where $nExp_{1,n-1}$ is the number of ad exposures in the first $n - 1$ pages and Ad_n is a dummy variable indicating ad display at the n^{th} page. Since, by experiment design, Ad_n is random, the coefficient β_1^n is a consistent estimator of the additional effect of advertising conditional on past treatment $nExp_{1,n-1} \geq 1$. If there is no additional effect of multiple advertising, β_1^n would not be positive for any n . Table 12 shows estimates from the regressions on the subset of sessions with $nPages \geq n$ and $nExp_{1,n-1} \geq 1$, for $2 \leq n \leq 9$. Note the coefficient β_1^n is positive in most cases and statistically significant in some, indicating the positive effect of additional advertising at the n^{th} position. Also note the magnitude of impact of an extra exposure is about 0.20% when it is significant.

The above test is conservative since it averages across all past treatment patterns. However, additional advertising might be more effective when $nExp_{1,n-1}$ is small. Also, the sequence of past advertising exposure might be important. To further explore these aspects, I focus on the two cases that provided the weakest evidence in the Table 12: $n = 8$ and $n = 4$. Column I of Table 13 shows results from the regression in equation (20) when the past exposures before the 8^{th} page, $nExp_{1,7}$, is positive but small. The coefficient is now positive but statistically not different from zero. However, if the immediate previous page displayed the ad banner ($Ad_7 = 1$), Ad_8 has a significant effect as shown in column II. A similar case for $n=4$ is shown in columns III and IV which again shows an effect of similar magnitude when ads are displayed in quick succession.

F Impact of ads on future revisit

An implicit assumption made in the analysis is that the future sessions on the website and the time intervals between sessions are not dependent on advertising exposure. In this section, I check for the presence of such correlations in the data. In column I of table 14 I regress a dummy variable indicating that the user just had one session over the course of the experiments, on allocation of the user's first session to the ad condition. I find that the coefficient for the dummy variable is statistically not different from zero indicating no correlation between being in the ad condition and future revisit decisions. In columns II and III I include the past decisions of whether to visit the advertise restaurant's page or generate a sales lead as explanatory variable. Again, I find no significant impact of ads on future revisit decisions. Next, for individuals that do visit more than once I regress the time intervals between first two sessions on the allocation to ad condition in the first session, in table 15. Again, I find no significant correlation between allocation to the ad condition and website re-visit frequency.

This evidence shows that in my data the assumption about the future revisit frequency being independent of advertising exposure is weak.

G More on the Impact of Ad Exposures Across Sessions

The model for long-term effects of advertising proposed in this paper is driven by learning through advertising exposure. Another mechanism driving carry-over could be state-dependence or experience. According to this explanation, high advertising exposure leads to purchase at the previous occasion. This purchase in turn increases the chance of repurchase in the future time period due to consumer inertia (Dubé et al. [2010]). If carry-over occurs solely through this mechanism versus learning, it would have different implications for the advertiser; the advertiser would have an incentive to focus more on the subset of consumers that have not recently purchased the product. To investigate the presence of this mechanism, I focus on a subset of individuals with repeat visits (considered in section 6) but who do not visit the advertiser’s page in their first sessions. Columns IV and V of Table 16 show the results from the logit regression of visits in session 2 on current ($nExp_1$) and past ($nExp_2$) ad exposures for this subset of individuals. The coefficient on $nExp_1$ is positive and statistically significant. This finding rules out the explanation that carry-over of ad effects are due solely to past experience induced by past advertising.

References

- Daniel A. Akerberg. Empirically distinguishing informative and prestige effects of advertising. *RAND Journal of Economics*, 32(2):316–333, 2001.
- AdvertisingAge. Advertising to sales ratio by industry. *Data center: Advertising spending*, 2010.
- John R. Anderson and Robert Milson. Human memory: An adaptive perspective. *Psychological Review*, 96(4):703–719, 1989.
- John R. Anderson, Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. An integrated theory of the mind. *Psychological Review*, 111(4):1036–1060, 2004.
- Darral G. Clarke. Econometric measurement of the duration of advertising effect on sales. *Journal of Marketing Research*, 13(4):345–357, 1976.
- DoubleClick. 2009 year-in-review benchmarks. *doubleclick.com*, 2010.
- Jean-Pierre Dubé, Günter J. Hitsch, and Peter E. Rossi. State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics*, 41(3):417–445, 2010.
- Jean-Pierre H. Dube, Guenter J. Hitsch, and Puneet Manchanda. An empirical model of advertising dynamics. *Quantitative Marketing and Economics*, 2005.
- H. Ebbinghaus. Memory: A contribution to experimental psychology (translated by Henry A. Ruger and Clara E. Bussenius; original German work published 1885). *New York: Teachers College, Columbia University.*, 1913.
- Tülin Erdem, Michael P. Keane, and Baohong Sun. A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science*, 27(6):1111–1129, 2008.
- Anindya Ghose and Sha Yang. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science*, 55(10):1605–1622, 2009.
- Anindya Ghose and Sha Yang. Modeling cross-category purchases in sponsored search advertising. *SSRN eLibrary*, 2010.
- Chris Janiszewski, Hayden Noel, and Alan G Sawyer. A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research*, 30(1):138–149, 2003.
- Randall Lewis and David Reiley. Retail advertising works! measuring the effects of advertising on sales via a controlled experiment on yahoo! *Working Paper*, 2009.

- Leonard M. Lodish, Magid Abraham, , Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, and Mary Ellen Stevens. A summary of fifty-five in-market experimental estimates of the long-term effect of tv advertising. *Marketing Science*, 14(3):G133–G140, 1995a.
- Leonard M. Lodish, Magid Abraham, Stuart Kalmenson, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, and Mary Ellen Stevens. How t.v. advertising works: A meta-analysis of 389 real world split cable t.v. advertising experiments. *Journal of Marketing Research*, 32(2): 125–139, 1995b. ISSN 00222437.
- Vijay Mahajan and Eitan Muller. Advertising pulsing policies for generating awareness for new products. *Marketing Science*, 5(2):89–106, 1986.
- Prasad A. Naik, Murali K. Mantrala, and Alan G. Sawyer. Planning media schedules in the presence of dynamic advertising quality. *Marketing Science*, 17:214–235, March 1998. ISSN 1526-548X.
- Marc Nerlove and Kenneth J. Arrow. Optimal advertising policy under dynamic conditions. *Economica*, 29(114):129–142, 1962.
- Philip I. Pavlik and John R. Anderson. Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology*, 14(2):101–117, 2008.
- Cornelia Pechmann and David W. Stewart. Advertising repetition: A critical review of wearin and wearout. *Current Issues and Research in Advertising*, 11(2):285–330, 1988.
- Oliver J. Rutz and Randolph E. Bucklin. A model of individual keyword performance in paid search advertising. 2007.
- Alan G. Sawyer and Scott Ward. Carry-over effects in advertising communication. *Research in Marketing*, 2:259–314, 1979.
- Raj Sethuraman, Gerard J. Tellis, and Richard A. Briesch. How well does advertising work? generalizations from meta-analysis of brand advertising elasticities. *Journal of Marketing Research*, Forthcoming.
- Nobuhiko Terui, Masataka Ban, and Greg M. Allenby. The effect of media advertising on brand consideration and choice. *Working Paper*, 2010.
- Russell S. Winer. Using single-source data as a natural experiment for evaluating advertising effects. *Journal of Marketing Science*, 2(2):15 – 31, 1993.

Experiment	Restaurant Cuisine	% of visitors in the area that visit the restaurant's page in the absence of ads	Some description
1	Indian/Mughlai	1.1%	100 years old small chain with less than 10 outlets
2	Chinese	1.5%	Relatively new restaurant (about 2 years old)
3	Italian	2.5%	International pizza chain
4	Chinese	2%	Local chain
5	Chinese	1.4%	Local chain
6	Italian	0.7%	Local chain
7	Indian	2%	Small local chain
8	Indian/Bengali	5.7%	Local chain
9	Italian	3.1%	International pizza chain
10	Indian	2%	Local chain
11	Chinese	1.4%	International chain

Table 1: Characteristics of advertised restaurants used for the experiments

	Median	5 th percentile	95 th percentile	Number of sessions in data ($NSess$)	No. of users	Percent
Number of unique restaurant pages visited	1	0	6	$NSess=1$	193,712	89%
% of users with at least one ad click (On any ad)	1.67%			$NSess = 2$	15,394	7%
% of users that viewed a phone number (for any restaurant)	17%			$3 \leq NSess \leq 5$	6,813	3%
				$6 \leq NSess \leq 15$	1,399	1%

Table 2: Summary statistics about the browsing behavior of the website users in data

Search Page

Restaurants in South Delhi

Change your location:
Zone: South Delhi
Subzone: All Go

Search within results:
Search for restaurant / locality / cuisine
 Go

Advanced Filters | Filter results by Cuisines, Cost, Ratings, Veg, Bar etc...

Search Results

Sort by: Popularity (best) Go

Displaying results 1 - 20 of 821

« Previous
1
2
3
4
5
...
Next »

DISCOUNTS

Asia 7

ASIAN, CHINESE, JAPANESE

The Qutub (Clanor Collection), Qutub Institutional Area...

★★★★☆

Qutub Institutional Area

Cost for two: 1000

DISCOUNTS

The Royal Kitchen

CHINESE, NORTH INDIAN, MUG LAI

9, Sant Nagar, East of Kailashi, New Delhi

★★★★☆

East of Kailash

Cost for two: 450

DISCOUNTS

Brown Sahib

BENGALI, OTHERS

2nd Floor, MG Metroopolitan Mall, Saket, New Delhi

★★★★☆

Saket

Cost for two: 1600

Spice Market

NORTH INDIAN, MUGHALI, RAJAS HANI

GFI8, D2, Southern Park, Behind Select Citywalk, D...

★★★★☆

Saket

Cost for two: 1200


Lodi - The Garden Restaurant

CONTINENTAL, MEDITERRANEAN

★★★★☆

Lodhi Road


Cost for two: 1200



abia seven

THE ABIA GROUP


BOOK NOW FOR A 10% DISCOUNT



ADVERTISE WITH US

email us at restaurants@zomato.com

15% OFF



YULETIDE MENU AT


Brown Sahib

METROPOLITAN MALL SAKET


DEC 15 - JAN 3

birthday cakes @

BAGELS & BROWNIES



ORDER NOW!




Great offers for private parties

Book now!

spicemarket

INDIAN RESTAURANT



ROYAL KITCHEN

Upto 20 % off

On Dine-in

CELEBRATING 10 YEARS

Restaurant Page

Karims

NORTH INDIAN, MUGHLAI

Address: 168/2, Jha House Basti, Nizamuddin West, New Delhi

Phone: [Click to View Phone Number](#)

Timings: Cost for 2: Rs. 700


No user reviews yet for Karims. **Write one now!**

Editor Rating: ★★★★★ [**Food:** ★★★★★ **Service:** ★★★★★ **Ambience:** ★★★★★]

[Menu](#)
[Map](#)


[« Previous](#)
[1](#)
[2](#)
[3](#)
[Next »](#)

1. Murgh ka Shorba	16. Karahi Chicken
2. Mutton Yakhni	17. Chicken Noorjahani
3. Sabzi Ka Garam Ras	18. Butter Chicken Boneless
4. Tomato Soup	19. Butter Chicken
5. Tandoori Murgh	20. Shahjahani Murgh
6. Tandoori Raan	21. Shahjahani Murgh (Boneless)
7. Tandoori Fish (Winter Season)	22. Jahangiri Qorma
8. Tandoori Burr	



asia seven
THE ASIAN KITCHEN

BOOK NOW FOR A 10% DISCOUNT



ADVERTISE WITH US
email us at restaurants@zomato.com

15% OFF

YULETIDE MENU AT
Brown Sahib
METROPOLITAN MALL
SAREET
DEC 15 - JAN 3

birthday cakes @ **BAGELS & BROWNIES**

ORDER NOW!

spicemarket
AUTHENTIC INDIAN RESTAURANT

Great offers for private parties
Book now!

The ROYAL KITCHEN

Organize parties starting
₹ 275/- per head

CELEBRATING 10 YEARS

Figure 1: Snapshot of a search page and a restaurant page at the restaurant search website.



Figure 2: Example snapshot of experiment conditions. The experimental banner is circled

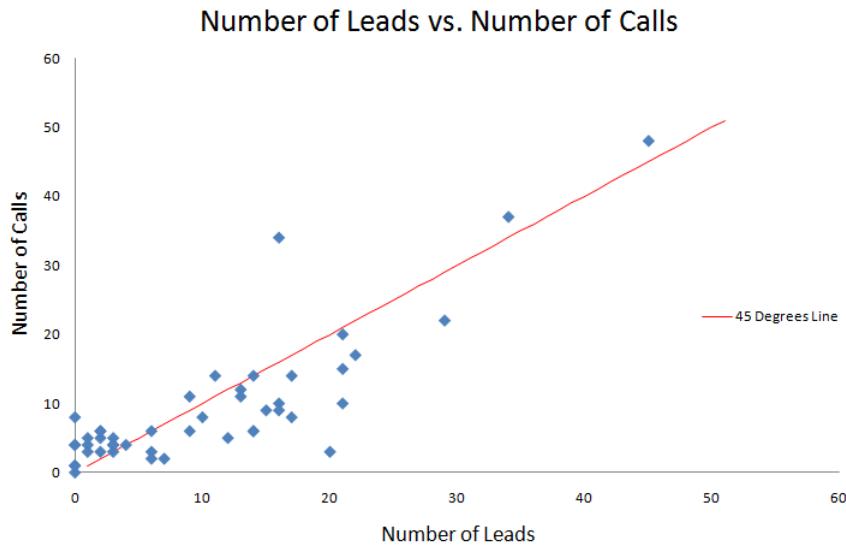


Figure 3: A plot of number of times the phone number was viewed (aggregated for a week) v.s. number of calls from unique phone numbers received for the restaurant during the week, for 48 restaurant-week observations

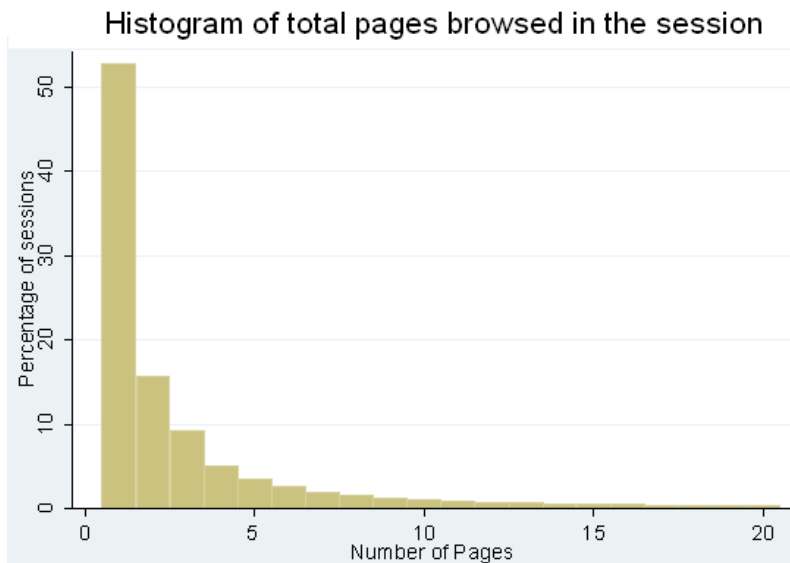


Figure 4: Distribution of number of pages browsed across sessions

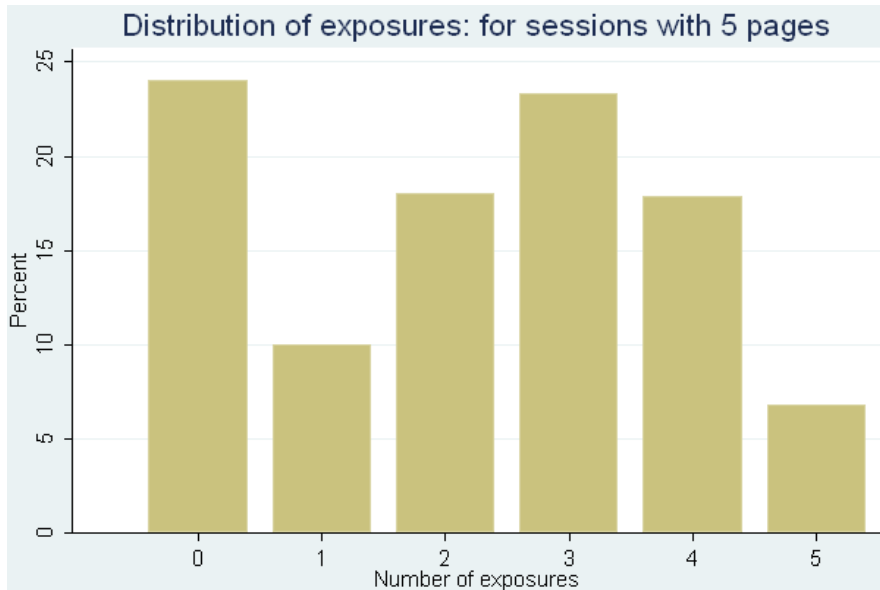


Figure 5: Histogram showing variation in the number of ad exposures, controlling for the number of pages browsed

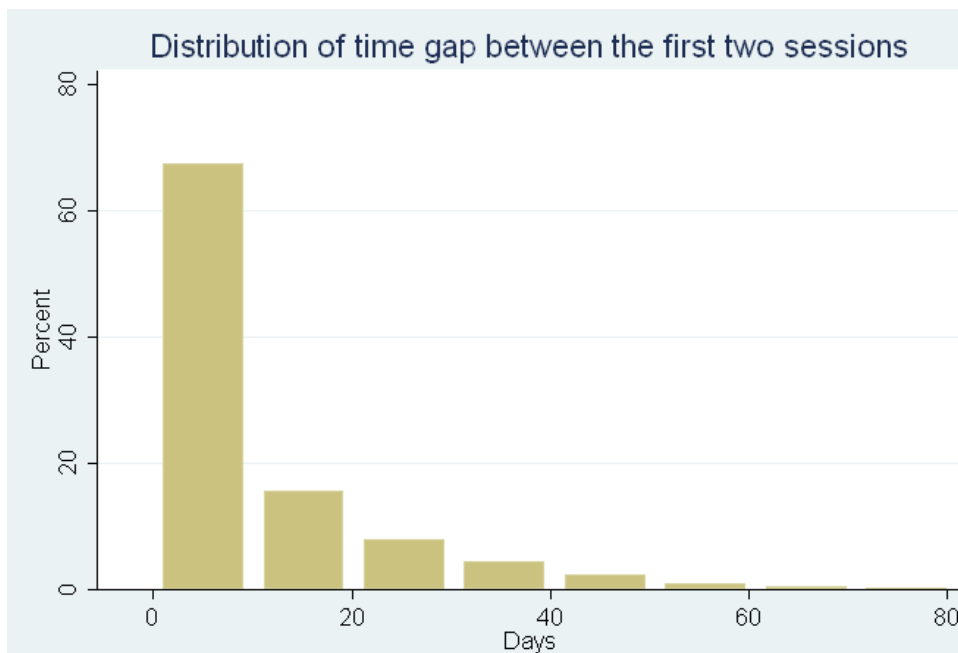


Figure 6: Histogram showing time intervals in days, between sessions for returning users in the data

	DV: Visit to advertiser's page		DV: View advertiser's phone num.	
	I	II (Num. Pages ≤ 20)	III	IV (Num. Pages ≤ 20)
Dummy indicating the Ad condition	0.0025*** (0.001)	0.0023*** (0.001)	0.0011*** (0.0003)	0.0011*** (0.0003)
Intercept	0.016*** (0.001)		0.0023*** (0.0003)	
Num. Observations	217,401	211,283	217,401	211,283

Standard errors in parentheses. Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3: Effect of the experimental manipulation: allocation of the session to the ad condition

Logistic Regression DV: Visit to advertiser's page in the session				
	I	II	III	IV
<i>nExp</i> :				
Num. of Exposures	0.18*** (0.005)	0.04*** (0.008)	0.146*** (0.017)	0.17*** (0.011)
<i>nExp</i> ²			-0.01*** (0.001)	-0.01*** (0.002)
<i>nExp</i> ≥ 1				-0.078 (0.058)
Num. Pages browsed		0.10*** (0.005)	0.104*** (0.005)	0.108*** (0.005)
Experiment Fixed effects	✓	✓	✓	✓
Intercept	-4.13*** (0.09)	-4.29*** (0.09)	-4.37*** (0.1)	-4.35*** (0.28)
Num. Observations	211,283	211,283	211,283	211,283

Table 4: Impact of repeated exposure to the banner ads in a browsing session.

Logistic Regression DV: Visit to advertiser's page in the second session				
	I	II	III	IV
$nExp_1$	0.053** (0.024)	-0.046 (0.064)	0.104*** (0.031)	0.122*** (0.04)
$nExp_1^2$		0.009* (0.0056)		
$nExp_2$	0.058* (0.031)	0.059* (0.031)	0.103*** (0.036)	0.103*** (0.036)
$nExp_2 \times nExp_1$			-0.015** (0.007)	-0.015** (0.007)
Time gap (<i>days</i>)				-0.001 (0.007)
$nExp_1 \times days$				-0.001 (0.002)
Num. Pages browsed in Session 2	0.100*** (0.019)	0.100*** (0.019)	0.095*** (0.019)	0.095*** (0.019)
Intercept	-4.170*** (0.323)	-4.096*** (0.326)	-4.275*** (0.326)	-4.272*** (0.334)
Experiment Fixed effects	✓	✓	✓	✓
Num. Observations	10,739	10,739	10,739	10,739

Table 5: Impact of number of exposure to ads in the first ($nExp_1$) and second ($nExp_2$) sessions on probability of visiting advertiser's page in the second session.

	$nExp_2$	0	1	2	3	4	5
$nExp_1$							
0		3.6%	3.9%	4.3%	4.8%	5.3%	5.8%
1		3.9%	4.3%	4.7%	5.1%	5.5%	6.0%
2		4.3%	4.7%	5.0%	5.4%	5.7%	6.1%
3		4.8%	5.1%	5.4%	5.7%	6.0%	6.3%
4		5.3%	5.5%	5.7%	6.0%	6.2%	6.5%
5		5.8%	6.0%	6.2%	6.3%	6.5%	6.7%

Table 6: Likelihood of visit to the advertiser's page when $nExp_1$ and $nExp_2$ vary. This is simulated for a session with 10 pages in the second session.

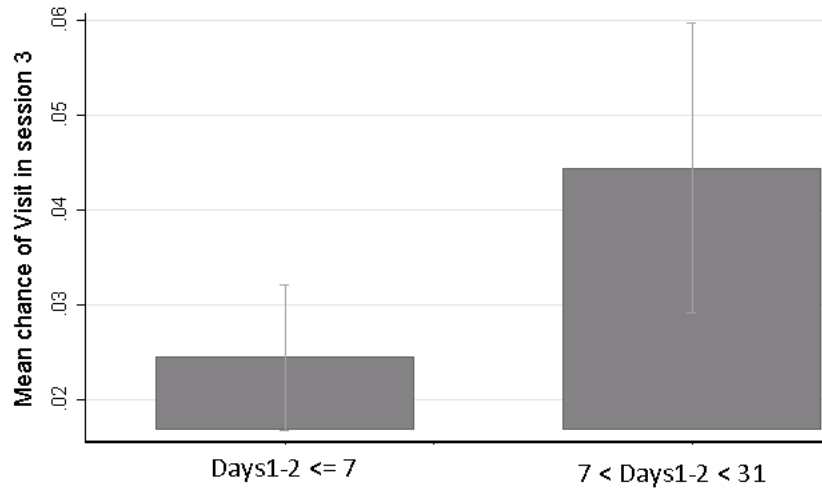


Figure 7: Mean chance of visiting the advertised restaurant's page in the third session, split by time gap between the first two sessions ($days_{1-2}$). The error bars show 90% confidence intervals for the means.

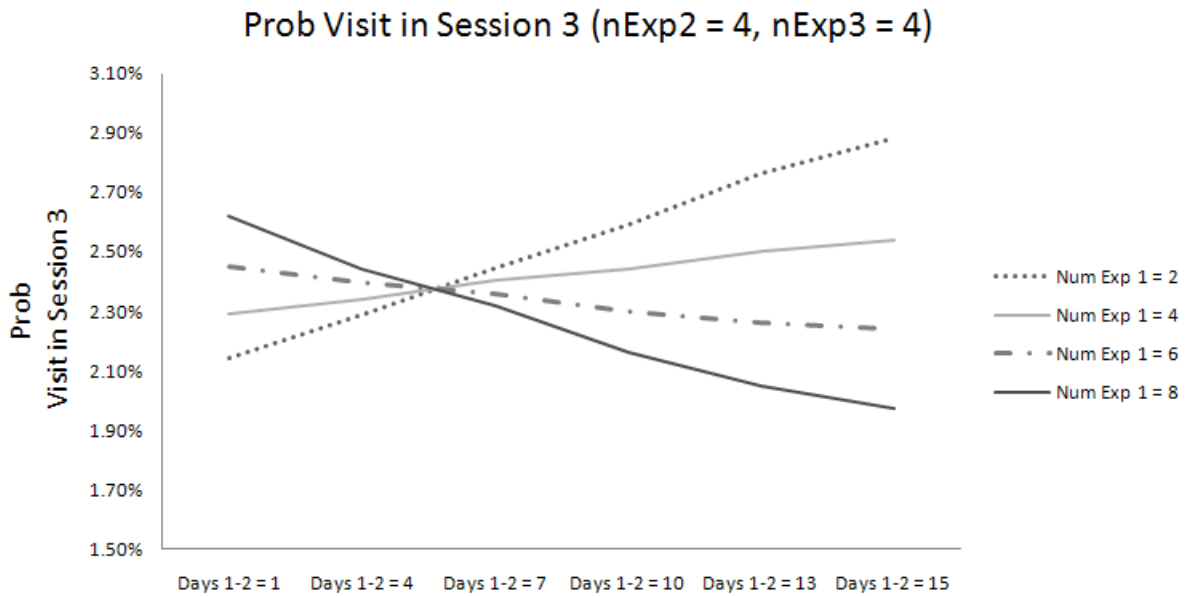


Figure 8: Quantitative significance of spacing: Change in the predicted probability of visiting the advertiser's page in session 3, with change in spacing between 1 and 2.

DV: $Visit_3$					
	I	II	III $days_{1-2} \leq 15$	IV $days_{1-2} \leq 15$ (No prior visit to the advertised restaurant's or its substitute's page)	V $days_{1-2} \leq 15$
$nExp_1$	0.027 (0.031)	0.074* (0.039)	0.089** (0.045)	0.12 (0.089)	0.12*** (0.048)
$nExp_2$	-0.031 (0.035)	-0.058 (0.046)	-0.141** (0.064)	-0.13 (0.11)	-0.06 (0.071)
$nExp_3$	0.111*** (0.041)	0.109** (0.045)	0.070 (0.051)	-0.18** (0.085)	0.072 (0.051)
$nExp_1 \times nExp_2$					-0.028** (0.014)
$nExp_1 \times days_{1-2}$		-0.007* (0.004)	-0.01 (0.008)	-0.02 (0.016)	-0.007 (0.008)
$nExp_2 \times days_{1-2}$		0.003 (0.003)	0.020** (0.008)	0.020 (0.014)	0.017** (0.008)
$nExp_3 \times days_{1-2}$		0.000 (0.002)	0.007 (0.006)	0.001 (0.01)	0.007 (0.006)
$days_{1-2}$	-0.009 (0.009)	-0.004 (0.015)	-0.040 (0.040)	0.018 (0.06)	-0.042 (0.04)
$days_{2-3}$	0.008 (0.008)	0.008 (0.008)	0.003 (0.009)	-0.004 (0.016)	0.0026 (0.009)
$nPages_3$	0.073*** (0.027)	0.073*** (0.027)	0.083*** (0.029)	0.08 (0.051)	0.079*** (0.029)
Intercept	-4.84*** (0.231)	-4.89*** (0.250)	-4.77*** (0.297)	-5.68*** (0.49)	-4.85*** (0.30)
Experiment Fixed effects	✓	✓	✓	✓	✓
Number of observations	6,047	6,047	4,908	3,081	4,908

Table 7: Impact of the spacing between sessions 1 and 2 ($days_{1-2}$), on the effect of ads on the decision of visiting the advertiser's page in session 3

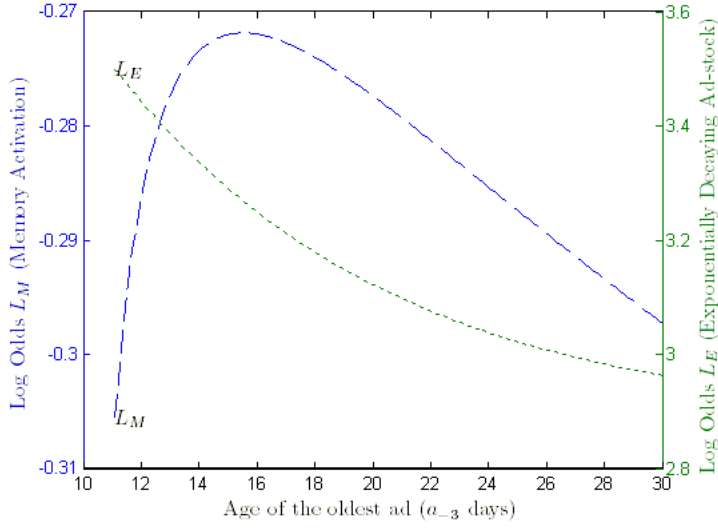


Figure 9: Simulation of the log odds for purchase for the memory-based model and the goodwill model (L_M & L_E). Impact of changing age of the earliest exposure on odds of purchase. Ages of other exposures are fixed to ($a_{-2} = 10, a_{-1} = 5$). Parameters ($c = 0.7, b = 0.2, \alpha_i=1, \gamma_i^{E,M} = 2, \rho = 0.9, \beta = -1$)

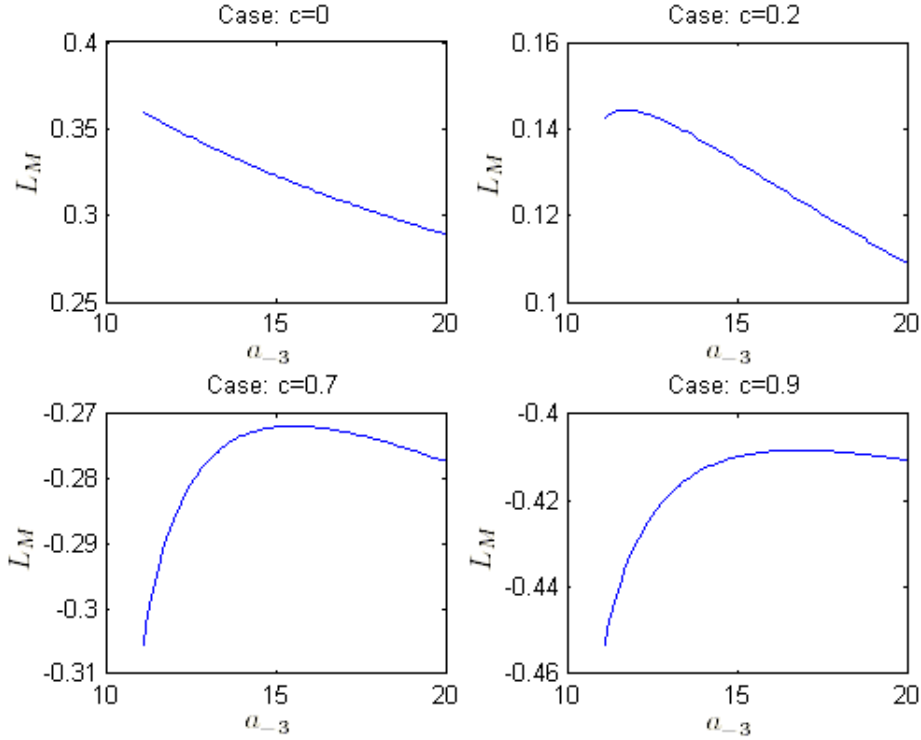


Figure 10: Varying parameter c to see the impact of changing a_{-3} on log odds of purchase L_M according to the memory-based model. ($a_{-2} = 10, a_{-1} = 5$). Parameters ($b = 0.2, \alpha_i=1, \gamma_i^M = 2, \beta = -1$)

	Exponential Decay Ad-Stock Model (E)			Generalized Exponential Decay Ad-Stock Model (GE)			Memory Activation Model (M)		
	I	II	III	IV	V	VI	VII	VIII	IX
ρ	0.97** (0.04)	0.98** (0.05)	0.99** (0.04)						
b^{GE}				3.79** (1.25)	5.72** (1.45)	4.87** (1.11)			
c^{GE}				-1.01 (0.88)	-0.94** (0.42)	-0.86** (0.29)			
b^M							-2.50 (2.36)	-3.79** (1.77)	-1.67** (0.89)
c^M							0.83 (0.72)	0.82* (0.44)	0.61** (0.27)
$nExp_t$	0.07** (0.02)	0.07** (0.02)	0.069** (0.02)	0.07** (0.02)	0.07** (0.02)	0.069** (0.02)	0.07** (0.02)	0.07** (0.02)	0.069** (0.02)
G_t	0.04 (0.03)	0.03 (0.05)	-0.28** (0.10)	0.06** (0.03)	-0.08* (0.05)	-0.27** (0.11)	0.05 (0.03)	-0.08* (0.05)	-0.38** (0.18)
G_t^2		0.0005 (0.004)	0.069** (0.025)		0.01** (0.00)	0.068** (0.03)		0.01** (0.01)	0.13 (0.11)
G_t^3			-0.003** (0.0015)			-0.003* (0.002)		-0.01 (0.01)	-0.01 (0.01)
$nPages_t$	0.10** (0.01)	0.10** (0.01)	0.10** (0.01)	0.10** (0.01)	0.10** (0.01)	0.10** (0.01)	0.10** (0.01)	0.10** (0.01)	0.10** (0.01)
Intercept	-4.42** (0.18)	-4.42** (0.18)	-4.34** (0.18)	-4.43** (0.18)	-4.38** (0.18)	-4.36** (0.18)	-4.43** (0.18)	-4.38** (0.18)	-4.34** (0.18)
- Log Likelihood	1869.07	1869.07	1862.01	1867.66	1864.02	1861.92	1868.12	1864.47	1862.29

Table 8: Estimation results for the homogenous model specifications for the models E , GE and M

Experiment fixed effects are not shown here.

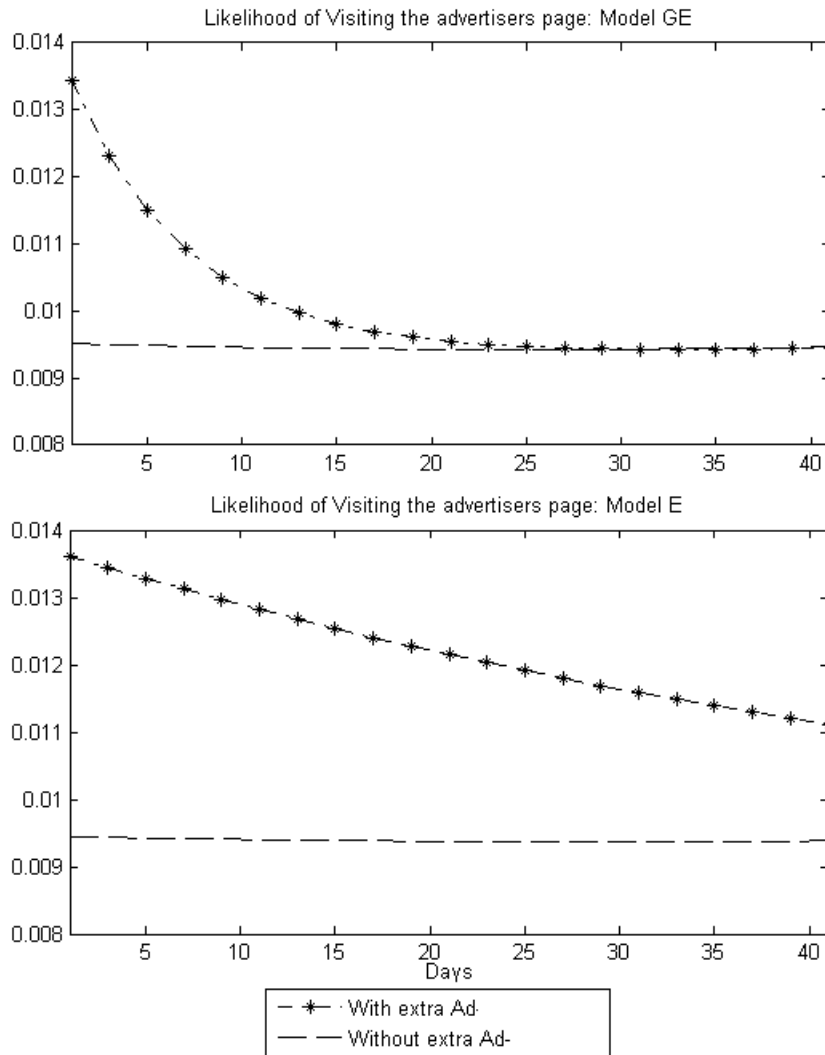


Figure 11: Comparing the impact of ad-exposures predicted by models GE and E . X axis is the time (days) in future from the current period (when the ad exposure takes place). On Y axis is the probability of the individual choosing to visit the advertiser's page if the purchase occasion arises at time x .

	Model <i>E</i>		Model <i>GE</i>			
	Column I		Column II		Column III	
	Mean	Standard Dev.	Mean	Standard Dev.	Mean	Standard Dev.
<i>b</i>	4.32** (1.25)	0.09 (1.36)	5.16** (1.52)	0.21 (1.41)	5.89** (1.61)	0.069 (1.61)
<i>c</i>			-0.81* (0.42)	0.03 (0.34)	-0.77** (0.39)	0 (0.46)
<i>nExp</i>	0.09** (0.03)	0.00 (0.06)	0.10** (0.03)	0.02 (0.06)	0.104** (0.030)	0.007 (0.058)
<i>G</i>	-0.03 (0.08)	0.18 (0.11)	-0.09 (0.10)	0.27** (0.11)	-0.07 (0.07)	0 (0.20)
Intercept	-6.42** (0.37)	2.11** (0.18)	-6.37** (0.36)	2.07** (0.17)	-6.30** (0.341)	2.03** (0.155)
G^2					0.018** (0.006)	
$G \times nExp$					-0.006 (0.006)	
$G \times x_1$					-0.020 (0.054)	
$G \times x_2$					-0.041 (0.062)	
$G \times x_3$					-0.039 (0.067)	
<i>nPages</i>	0.13** (0.02)		0.13** (0.02)		0.12** (0.018)	
-Log likelihood	1807.42		1805.63		1802.775	

Table 9: Estimates for models *E* and *GE* when unobserved heterogeneity is allowed
Note that for model *E*, $\rho = \frac{\exp(b)}{1+\exp(b)}$. Experiment fixed effects are not shown here.

Experiment No.	Conditions						Total
	NA	A1	A2	A3	A4	A5	
1	10,752	5,246	14,722	10,119	5,468	6,344	52,651
2	7,143		30,517				37,660
3	3,229	1,553	3,563	1,659	1,669	1,692	13,365
4	709	408	4,391	412	439	1,149	7,508
5	1,678	949	3,694	992	961	1,450	9,724
6	9,217	4,480	9,547	4,844	4,793	5,316	38,197
7	2,888	2,718	4,300	1,659	1,624	2,302	15,491
8	2,697	1,577	3,457	1,673	1,747	1,820	12,971
9	508		1,819			354	2,681
10	1,124		3,764			820	5,708
11	3,311		14,905			3,229	21,445
Total	43,256	16,931	94,679	21,358	16,701	24,476	217,401

Table 10: Number of sessions in different experiments and conditions



Figure 12: Advertising strategies yielding highest profits for the month. The advertiser is allowed to choose a frequency cap and the weeks when advertising is switched on.

Logistic Regression DV: Visit to advertiser's page in the session		
	I ($6 \leq nPages \leq 10$)	II ($nPages = 7$)
$nExp$:		
Num. of Exposures		0.099* (0.059)
$nExp \geq 1$		-0.328 (0.321)
$1 \leq nExp \leq 3$	0.084 (0.125)	
$4 \leq nExp \leq 7$	0.28** (0.11)	
$8 \leq nExp \leq 10$	0.30 (0.24)	
Num Pages	✓	
Fixed effects		
Experiment	✓	
Fixed effects		✓
Intercept	-4.56*** (0.264)	-2.47*** (0.49)
Num. Observations	18,340	4,124

Table 11: Impact of repeated exposure to the banner ads in a browsing session.

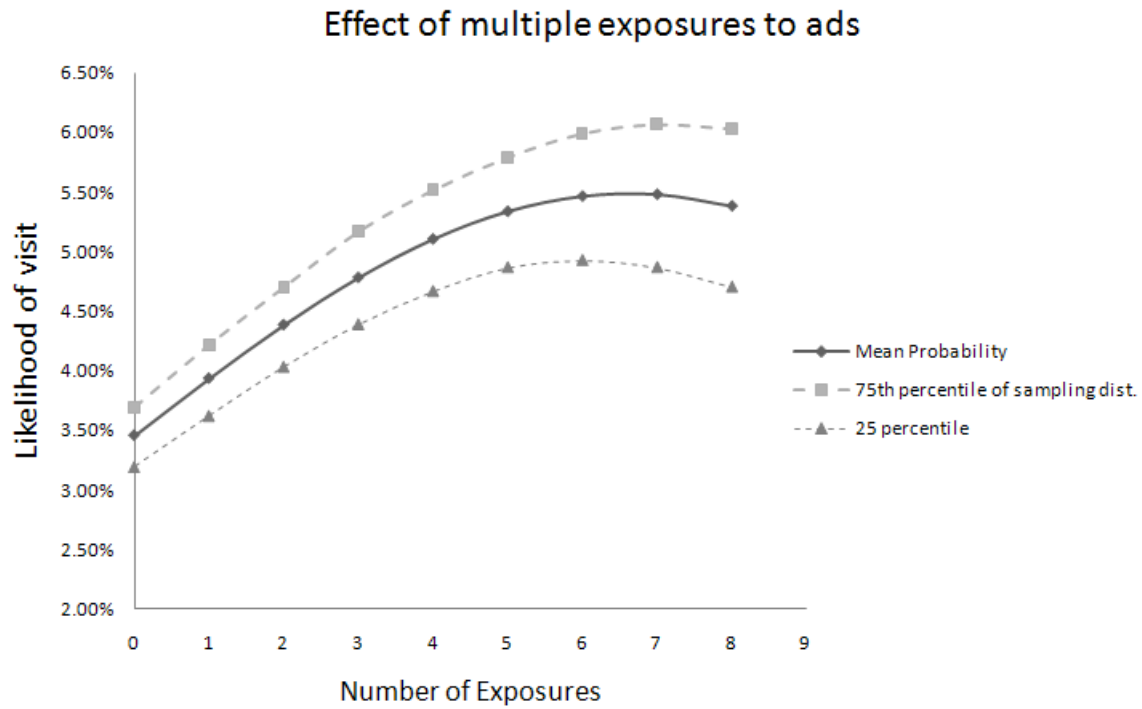


Figure 13: Quantitative significance of the impact of multiple exposures to ads within a session. Simulation for a session with 10 pages browsed.

Conditional on the subset of sessions with $nExp_{1,n-1} \geq 1$ and more than or equal to n pages				
	n=2	n=3	n=4	n=5
β_1	0.000 (0.001)	0.000 (0.001)	-0.000 (0.001)	0.002*** (0.001)
β_2	(dropped)	0.001 (0.001)	0.000 (0.001)	0.000 (0.000)
β_0	0.006*** (0.001)	0.005*** (0.001)	0.006*** (0.001)	0.005*** (0.001)
Num. Obs	47,063	45,749	37,800	31,893
	n=6	n=7	n=8	n=9
β_1	0.001 (0.001)	0.002** (0.001)	-0.000 (0.001)	0.002 (0.001)
β_2	0.001 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
β_0	0.004*** (0.001)	0.004*** (0.001)	0.005*** (0.001)	0.005*** (0.002)
Num. Obs	26,949	22,950	19,923	17,325

Table 12: Non-parametric evidence of presence of incremental effect of multiple exposure to ads

	n=8		n=4	
	Column I $1 \leq nExp_{1,7} \leq 4$	Column II $nExp_{1,7} \leq 4$ and $Ad_7 = 1$	Column III $nExp_{1,3} = 1$	Column IV $nExp_{1,3} = 1$ and $Ad_3 = 1$
β_1	0.000 (0.001)	0.0037* (0.002)	0.0011 (0.0014)	0.0039 (0.0026)
β_2	0.00 (0.00)	0.00 (0.001)	(dropped)	(dropped)
β_0	0.005** (0.002)	0.003 (0.004)	0.006*** (0.001)	0.005*** (0.002)
Num. Obs	12,260	5,268	12,881	4,227

Table 13: Evidence showing that ads may be more effective when displayed consecutively

OLS DV: Dummy indicating one session only			
	I	II	III
Ad Condition	-0.00	0.001	0.00
(Ad)	(0.002)	(0.002)	(0.002)
<i>Visit</i>		-0.06*** (0.01)	
<i>Visit</i> ×Ad		-0.013 (0.015)	
<i>View_num</i>			-0.10*** (0.04)
<i>View_num</i> ×Ad			0.006 (0.04)
Intercept	0.88*** (0.002)	0.94*** (0.01)	0.94*** (0.01)

Table 14: Impact of ads on consumer's decision to revisit the website

DV: days between first two sessions		
	I	II
Ad condition	0.24	0.21
in session 1	(0.24)	(0.24)
(Ad ₁)		
<i>Visit</i> ₁		-3.72*** (1.37)
<i>Visit</i> ₁ × Ad ₁		1.84 (1.58)
Intercept	6.89*** (1.38)	7.0*** (1.38)

Table 15: Impact of ads on time interval between sessions

Logit DV: Visit to advertiser's page in the second session					
	I	II	III	IV	V
			$Ad_2 = 0$	$Visit_1 = 0$	$Visit_1 = 0$
$nExp_1$	0.051** (0.024)	0.087** (0.041)	0.111* (0.057)	0.033 (0.029)	0.079** (0.039)
$nExp_2$	0.056* (0.031)	0.055* (0.031)		0.056* (0.034)	0.087** (0.039)
$nExp_2 \times nExp_1$					-0.012 (0.008)
Num. Pages browsed in Session 2	0.189*** (0.070)	0.193*** (0.070)	0.221 (0.181)	0.177** (0.076)	0.172** (0.077)
Num. Pages browsed in Session 1		-0.029 (0.027)			
$1 \leq nPages_2 \leq 3$	✓	✓	✓	✓	✓
$4 \leq nPages_2 \leq 7$	✓	✓	✓	✓	✓
$8 \leq nPages_2 \leq 10$	✓	✓	✓	✓	✓
$11 \leq nPages_2 \leq 14$	✓	✓	✓	✓	✓
Intercept	-5.919*** (1.263)	-5.906*** (1.264)	-22.607	-5.937*** (1.389)	-5.965*** (1.388)
Experiment Fixed effects	✓	✓	✓	✓	✓
Num. Observations	10,739	10,739	2,111	10,494	10,494

Table 16: Impact of exposure to ads on probability of visiting advertiser's page in the next session