

# **Modeling the Clickstream: Implications for Web-Based Advertising Efforts**

Patrali Chatterjee

Department of Marketing  
Rutgers University  
Newark, NJ 07102-1897  
email: *patrali@newark.rutgers.edu*

Donna L. Hoffman  
Thomas P. Novak

Owen Graduate School of Management  
Vanderbilt University  
Nashville, TN 37203  
email: *donna.l.hoffman@vanderbilt.edu*  
email: *tom.novak@vanderbilt.edu*

May1998

The first author thanks Project 2000, the Owen Graduate School of Management, the National Science Foundation Economic, Decision and Management Science Grant Number SBR 9422780, and an anonymous corporate sponsor for their generous support of this project.

# **Modeling the Clickstream: Implications for Web-Based Advertising Efforts**

## **Abstract**

Advertising revenues have become a critical element in the business plans of most commercial Web sites. Despite extensive research on advertising in traditional media, managers and researchers face considerable uncertainty about its role in the online environment. The benefits offered by the medium notwithstanding, the lack of models to measure and predict advertising performance is a major deterrent to acceptance of the Web by mainstream advertisers. Presently, existing media models based on aggregate vehicle and advertising exposure are being adapted which underutilize the unique characteristics of the medium. What is required are methods that measure how consumers interact with advertising stimuli in ad-supported Web sites, going beyond mere counts of individual "eyeballs" attending to media. Given the enormous potential of this dynamic medium, academic investigation is sorely needed before cost implications and skepticism endanger the ability of the medium to generate and maintain advertiser support.

This paper addresses advertiser and publisher need to understand and predict how consumers interact with advertising stimuli placed at Web sites. We do so by developing a framework to formally model the commercial "clickstream" at an advertiser supported Web site with mandatory visitor registration. Consumers visiting such Web sites are conceptualized as deriving utility from navigating through editorial and advertising content subject to time constraints. The clickstream represents a new source of consumer response data detailing the content and banner ads that consumers click on during the online navigation process. To our knowledge, this paper is the first to model the clickstream from an actual commercial Web site.

Clickstream data allow us to investigate how consumers respond to advertising over time at an individual level. Such modeling is not possible in broadcast media because the data do not exist. Our results contrast dramatically from those typically found in traditional broadcast media. First,

the effect of repeated exposures to banner ads is U-shaped. This is in contrast with the inverted U-shaped response found in broadcast media. Second, the differential effects of each successive ad exposure is initially negative, but non-linear, and becomes positive later at higher levels of passive ad exposures. Third, the negative response to repeated banner ad exposures increases for consumers who visit the site more frequently. Fourth, in contrast to findings in traditional media the effect of exposure to competing ads is either insignificant or positive. However, carryover effects of past advertising exposures are similar to those proposed in broadcast media. Finally, heterogeneity in cumulative effects of advertising exposure and involvement across consumers captured by cumulative click behavior across visits and click behavior during the visit was found to significantly contribute to differences in click response. This has implications for dynamic ad placement based on past history of consumer exposure and interaction with advertising at the Web site. Response parameters of consumer visitor segments can be used by media buyers and advertisers to understand the benefits consumers seek from the advertising vehicle, and thus guide advertising media placement decisions. In sum, our modeling effort offers an important first look into how advertisers can use the Web medium to maximize their desired advertising outcomes.

Keywords: advertising and media research, buyer behavior, choice models, clickstream data, computer-mediated environments, Internet, World Wide Web

## 1. Introduction

Advertising sponsorship currently dominates online sales and subscriptions as a revenue generating mechanism on the World Wide Web (IAB 1998)<sup>1</sup>. Web advertising revenues are doubling each quarter and are predicted to reach \$4.3 billion by 2000; even as they represent only a fraction of the \$54.9 billion in advertising revenues across all media (IAB 1998). Yet, as the Web moves toward acceptance as a mainstream advertising medium, the lack of advertising performance measurement and predictability is hindering its adoption and growth by mainstream advertisers (Meeker 1997).

Though the Internet facilitates the capture of consumer response data unprecedented in the history of mass media, current metrics that can be used as a basis for measuring online advertising performance lack standardization. Debates regarding appropriate advertising pricing models (e.g. CPMs versus ad banner clickthrough rates) and confusion over the appropriate unit of analysis (i.e. hits, visits, unique visits, page views, and the like), mire industry efforts to grow the medium as a business. As a consequence, the resulting ambiguity in media decisions, the exponential growth in the number of publishers competing for limited advertiser budgets and the increasing costs of Web site maintenance serve to adversely affect the bottom lines of commercial Web publishers (Clark 1997).

Two broad categories of Web sites define the Web advertising problem: 1) the advertiser's Web site, and 2) the publisher's Web site. The advertiser's Web site contains the advertiser's marketing communications program and ranges from a simple one-page advertisement, to a multimedia multi-page description of one or more products or services, all the way to an entire electronic commerce enabled online storefront such as that developed by L.L. Bean (<http://www.llbean.com/>). The publisher's Web site, for example Salon Magazine

---

<sup>1</sup> Only a handful of high traffic Web sites, notably the Wall Street Journal Interactive, Slate, Business Week, and Playboy currently follow the paid subscription model (Fitzgerald 1998).

(<http://www.salon1999.com/>), contains information and/or entertainment-oriented editorial content.

Consumers visit publishers' Web sites according to their individual interests and tastes. Once there, consumers have the potential to be exposed to advertisements in the form of banner ads placed throughout the site. The initial goal of such Web advertising efforts is to attract consumers to the advertiser's Web site. Tracking consumer response to advertising at commercial Web sites occurs by analyzing the "log file." In some cases it may be advantageous for at least a portion of the advertiser's Web site to be physically located on the publisher's Web server.

Consumer interaction with and enjoyment of content on Web sites appears to generate positive images for brands (Consumer Experience Probe 1996). However, these outcomes are best achieved when consumers are actually aware of and visit the advertiser's Web site. Because consumer Web traffic is fragmented across millions of Web sites, commercial sites must compete intensely for even small share of consumer visits (Meeker 1997). This is despite the exponential growth in number of Web users and time spent online (GVU WWW User Surveys 1998). Publisher's ad-supported Web sites aggregate consumers and help advertisers attract some of the publisher site's traffic. Web advertisers want consumers visiting the publisher's site to view their ads, and click on them to interact with advertising information. However they are reluctant to commit advertising budgets because of the unpredictability and the lack of measurability of consumer response to their ads at publisher's Web sites.

In this paper, we suggest an approach to modeling consumer response to ads at an advertiser-supported publisher Web site. In the next section we discuss advertising vehicles and forms on the Web, and current practices in Web advertising measurement. In section 3 we introduce the problem and conceptualize the hypotheses under investigation. Then we develop an econometric model to examine and predict consumer response to ads at a Web site. We use clickstream data of consumer navigation activity to estimate the model. We conclude with a

discussion of managerial implications.

## 2. Concepts and Terminology

### 2.1 Advertising Vehicles and Forms on the Web

Just as with advertising in the print and broadcast media, publisher Web sites that function as ad-supported media vehicles, must serve the dual demands of two groups: Web consumers and Web advertisers. The role of the publisher's site includes enabling and encouraging the consumer to interact with advertising embedded within content at its site. Web advertising has two components - passive and active. These passive and active components are differentiated by the amount of control exercised by the consumer over their exposure.

Consumers are automatically exposed to **passive ads**, typically in the form of "banner ads", when they visit a publisher's Web site. Passive ads act as a gateway to active ads, since they provide the hyperlink that the consumer clicks to view the active ads. Each time a consumer is exposed to an editorial page with an embedded passive ad (or ads) on it, a **passive ad exposure** (or "impression" or "page view") is recorded for the advertising sponsor, representing an opportunity for the consumer to click on the ad. As in broadcast media, the number of passive ad exposures generated depends on exposure to the surrounding editorial content, and is thus in part under the publisher's control. Passive ad exposure is judged to be effective if it induces a consumer to click on the passive ad and access the active ad hyperlinked to it, resulting in "clickthrough" (Novak and Hoffman 1997). Inline ads (banners, icons or interstitial ads) are currently the most commonly used passive ads, they accounted for 54% of Web ad revenue in 1997 (IAB 1998).

If the consumer clicks on a passive ad, and the active ad is located on the publisher's Web server, an **active ad exposure** or (clickthrough) is recorded for the sponsor in the server access log and the consumer is then exposed to the sponsor's ad message. If the active ad is not located on

the publisher's Web server, then recording becomes more difficult. **Active ads** are displayed only if the consumer clicks on the corresponding passive ad and accesses the active ad hyperlinked to it. Unlike passive banner ads, consumers control their exposure to active ads.

The objective of an active ad is to encourage the consumer to and interact with the active ad content, possibly jumping directly to the advertiser's Web site. Like direct response print ads, direct mail pieces, or infomercials on broadcast television or cable, active ads on the Web may be directed towards lead generation or an immediate sale, or simply provide real time satisfaction consumer's need for more information about the sponsor's offering.

Active ads are appropriate for higher-level communication objectives, including facilitating comprehension and elaboration of the advertising message. Advertisers and publishers desire predictability of this consumer response since it has, at a minimum, implications for passive ad placement and pricing of ad space. This is the response we seek to model and understand.

## **2.2 Current Approaches To Web Advertising Measurement**

Novak and Hoffman (1997) provide an overview of current approaches to Web advertising measurement. Presently, most Web publisher's rate cards imitate measures similar to reach and gross rating points (GRPs) in broadcast media vehicles and are based on cost per thousand hits or visitors at site, cost per thousand (CPM) impressions for passive ad exposures or clickthroughs for active ad exposures. For details refer to Chatterjee (1998).

Over 86% of Web publishers (IAB 1997) use CPM impressions (passive ad exposures or page views) as the basis for pricing Web advertising. However, gross reach measures do not indicate actual perception of the ad by the consumer or the frequency of exposure. Some researchers (Briggs and Hollis 1997) suggest that mere exposure to passive ads improve brand awareness and image regardless of click response. However, the implicit assumption that "more (passive impressions) is better" may not hold true in generating advertiser-desired outcomes in the Web medium. Aggregate counts of clickthroughs or click rates (i.e., the ratio of passive ad clicks to

passive ad impressions) are actually preferable, since they measure actual exposure to passive ads and a commitment by the consumer to view the active advertising content.

### **3. The Problem and Hypotheses**

Our examination of how consumers respond to banner advertising in a commercial Web site and how this response varies across different segments of consumers is specifically directed at assessing the effects of:

- 1) Repeated passive advertising exposures on click probabilities;
- 2) Competing advertising exposures on click probabilities;
- 3) Repeated active advertising exposures on click probabilities;
- 4) Site interaction on click probabilities; and
- 5) Visitor segments and advertising response.

#### **3.1 Response to Repeated Passive Ad Exposure**

The concept of effective frequency is well-established in broadcast media (Sawyer 1981). The effects of repetitive advertising in terms of "wearin" and "wearout" have been confined to studies of television advertising (Sawyer 1981; Tellis 1988) and differ depending on whether advertising exposure is distributed over time or massed in laboratory or natural viewing situations. Unlike television advertising, the length of exposure to passive advertising on a Web page is under the consumer's control, an important factor in ad wearin. If an ad succeeds in attracting the consumer's attention on its first exposure, the consumer has the opportunity to attend and elaborate on the ad message. Hence it is possible for the online ad to wear in at the first exposure itself in the Web environment. An ad is said to have worn in at a particular level of repetition if, when consumers are exposed to it, it has a significant positive effect on them (Pechmann and Stewart 1989). An ad that has worn in is said to have worn out at a particular level of repetition if,

when consumers are exposed to it, it no longer has any significant effect on them or even has a significant negative effect. Exposures that are massed (i.e., repeated within the course of a few minutes or an hour during a visit) have been shown to accelerate wear out due to tedium effects if there is no incremental information for the consumer to process (Pechmann and Stewart 1989). Since the consumer can process all the information in a banner ad at the first exposure itself, it is possible for a passive Web ad to wear out at the second exposure.

Practitioners have charted decreasing returns to repeated passive ad exposures, suggesting that after the third exposure to the passive ad, the probability that a consumer will click on the passive ad is close to zero, referred to as “banner burnout” (DoubleClick 1996). This parallels findings in the print media that wear in and wear out of ads is faster since consumers control the rate of interaction (Schumann et al. 1988). Hence:

H1a. Click response to repeated exposures to passive ads within the same visit will be negative, and the mean click probability will decrease with each successive passive ad exposure during the same visit.

Conversely, the mere exposure hypothesis suggests repeated exposures distributed over time even in the absence of perception lead to awareness, familiarity, and positive affect towards the ad stimulus under low involvement conditions (Hoyer and Brown 1991). Briggs and Hollis (1997) documented increases in advertising awareness and brand perceptions to Web banner ads even if consumers do not click on them. Hence in the short-term i.e., during a visit, the negative effect of repeated passive ad exposures can be expected to level off after a few exposures.

H1b. There will be a non-linear effect of repeated passive ad exposures during the same visit.

Research on advertising carryover effects is concerned with residual or cumulative effects of prior advertising exposures at a subsequent point in time. In the long-term (i.e., across visits) familiarity due to repeated exposure to the ad stimulus have been shown to lead to positive affect.

However, if click behavior is driven by curiosity, we expect a negative relationship between cumulative effect of passive ad exposures and probability of click. The extent to which the positive affect and negative effects due to boredom interact to induce a consumer to click on the banner ad needs to be examined. We expect the positive effects to dominate across the history of consumers' visits to the site. Hence,

H1c. There will be a positive relationship between cumulative effect of passive ad exposure and probability of clicking on a passive ad.

### **3.2. Response to Competitive Advertising Exposure**

Distracting effects of ad clutter and presence of competing ads in media environment have been shown to affect consumer motivation to attend to an ad (MacInnis et al. 1991; Stewart et al. 1986). The negative effect of ad clutter on ad readership in the print media has been examined empirically (Houston and Scott 1984) indicating that the higher the ad clutter level in a magazine, the lower the readership of the ads.

In the Web environment it can be hypothesized that exposure to passive ads of other sponsors has distracting effects and reduces the probability of attending to the focal sponsor passive ad. The larger the number of passive ads on the same page (i.e., clutter effects due to proximity), the greater the competition for consumer share of attention and clicks, hence the lower the propensity of the consumer to click on a passive ad. Research on the quantity dimension of advertising clutter indicates that passive and active advertising exposures of other sponsors during network navigation may inhibit consumer's propensity to click on passive ad for the focal sponsor. Hence:

H2a. The greater the number of exposures to competing sponsors' passive ads, the lower the probability that the consumer will click on the focal sponsor's passive ad.

H2b. The greater the number of exposures to competing sponsors' active ads, the lower the probability that the consumer will click on the focal sponsor's passive ad.

### **3.3. The Effect of Active Ad Exposure**

As active ads represent sources of product information at ad-supported Web sites, passive ad click decisions allow us to examine actual product information search behavior by individual consumers over time. MacInnis et al. (1991) show that involvement influences motivation to attend and process advertising messages. Hence consumer involvement with the product/product class/ad/sponsor can be considered to drive his/her click decision. While clickstream data captures navigation behavior and not attitudinal data, heterogeneity in involvement across consumers at a Web site will account for differences in long-term click response to passive ads. A consumer's click behavior on exposure to passive ads across prior visits and within visits can be addressed from various perspectives. If the consumer clicks on a passive ad more often across visits one can infer that consumer interest and involvement is more enduring. This is in accordance with the literature, which argues that enduring involvement with the product/product class/ad/sponsor is a long-term effect. Thus:

H3a. The cumulative effect of active ad exposures during prior visits will have a significant positive effect on consumer's probability of clicking on a passive ad.

Situational involvement is temporary and can be considered to be a short term within visit effect captured by consumer's click behavior during the visit. Hence,

H3b. Active ad exposures during the present visit will have a significant positive effect on consumer's probability of clicking on the passive ad on subsequent exposure.

The relative importance of enduring involvement, (i.e., cumulative click behavior) and situational involvement (i.e., click behavior during the visit) in explaining consumer's click response is of interest. The effects of commercial messages have been shown to differ substantially depending on the use a particular consumer is making of a given media vehicle (Stewart and Ward 1994). If the site attracts consumers whose motivations for visiting the site are primarily goal-directed, present visit click response indicating situational involvement will have a bigger effect in

predicting click response. Conversely, cumulative click behavior (proxy for enduring involvement) will play a dominant role in predicting click response.

### **3.4. The Role of Consumer Interaction with the Web Site**

Stewart and Ward (1994) recommend that research in advertising requires a change in focus from analyses of stimulus and media to analysis of ways in which individuals interact with and act upon media. One of the unique characteristics of the Web medium is its interactivity. Consumer visit duration and number of pages browsed at the Web site capture consumer interaction with content at the site. In the print media, Carter (1968) found that print advertisement readership is higher in issues with fewer pages; hence larger issues have lower average ad readership. In the Web medium, visit duration represents opportunity for the consumer to respond to passive ads, while editorial pages represent competing activities. Hence:

H4a: The longer the visit duration, the higher the probability that the consumer will click on a passive ad on exposure.

H4b: The higher the number of editorial pages browsed by the consumer during the visit, the lower the probability that the consumer will click on a passive ad on exposure.

### **3.5. Segmentation and Response to Advertising**

Publishers of ad-supported Web sites seek to retain their consumers by catering to their needs for content and using different tools (e.g., regular e-mail of headlines, special contests and offers, special news features) to induce consumers to revisit their Web sites. Publishers of ad-supported Web sites desire a stable base of consumers to assure advertisers a return on their ad budgets since advertising exposures delivered will depend on consumer-initiated visit to the site. Intuition based on the concept of market segmentation suggests that targeted placement of ads will have a higher likelihood of eliciting consumer response resulting in higher efficiency in ad spending. Clickstream data allow measurement of actual response to advertising; it is possible to go beyond demographic (or host domain-specific) and social variables to identify consumer segments

responsive to a firm's online advertising. Moreover, the difficulties of collecting such detailed individual-level data in light of privacy concerns of Web consumers<sup>2</sup> and skewed demographics mask any insights that can be drawn. Stewart and Ward (1994) note that to understand how individuals interact with media we need to characterize individuals in terms of their media use and draw connections to their response to advertising embedded in it.

There is an extensive body of literature on segmenting consumers' use of broadcast and direct response media (Bawa and Shoemaker 1987). The logic for segmenting on the basis of frequency of readership, viewership or patronage of media vehicles can be found in media research conducted in the early 1970's. Urban (1976) suggested that heavy and light magazine readers might respond differently to ads with different creative appeals. Ehrenberg (1966) advocated the use of readership panel since consumer's ad readership scores change over time. Segmenting consumers on the basis of their usage frequency of the media vehicle will yield insights on whether the publisher site attracts and retains consumers that are more (or less) responsive to an advertiser's communication, an important input in evaluating the media vehicle's efficiency. Further, differences in visitor frequency may lead to differences in response to repeated passive ad exposures, competing ads of other sponsors and prior visit ad exposures.

## **4. Data and Model**

### **4.1. Clickstream Data**

Behavioral "clickstream" data of consumer navigation from Web server access logs will be used to investigate the hypotheses proposed in the last section. The main strengths of using server access log data are (i) they recreate behavior in the actual media environment, (ii) they are collected unobtrusively and based on observed behavior, rather than self reports, (iii) they are free from

---

<sup>2</sup> Most research on Web-based consumer attitudes indicate that privacy concerns outrank security and navigation challenges as the most important issue facing the Internet today (c.f., GVU WWW User Survey 1998; Hoffman,

confounds of researcher interaction, (iv) time pattern and order of activity is recorded, and (v) longitudinal data on census of consumers (and not just a sample) is obtained.

Access logs record consumer navigation at a Web site in terms of (i) the domain or IP address of the client (the name of network servers that connect an individual computer to the Internet, note no individual user identification can be automatically recorded), (ii) the date and time the document access took place, (iii) the HTTP method and protocol used for data transfer, (iv) the virtual path to the document transferred, (v) the status of the transfer, and (vi) how many bytes of information were transferred. There are ethical issues concerning consumer piracy and trust that impose serious restrictions on the use of server access log data for marketing purposes. Hoffman, Novak and Peralta (1997) suggest that online consumers are reluctant to provide information to Web providers in exchange for information offered onsite because of the fundamental lack of trust that currently exists between most businesses and consumers on the Web. The technical problems in using clickstream data for analyzing consumer behavior in online environments restrict the modeling scope.

#### **4.2. Clickstream Data Used in Study**

This study uses consumer clickstream data from a high-traffic sponsored content (or “e-zine”) Web site<sup>3</sup> from January 1, 1995 to December 31, 1995. From January 1, 1995 to August 13, 1995 the Web site required mandatory registration to enter the site. Starting August 14, 1995, registration was made voluntary. Demographic information was collected the first time a consumer registered at the site, and s/he was assigned a user name/id and a password to be used for future visits to the site. Once registered, the visitor can execute a login procedure for future visits. Demographic information that could be merged with the respondents’ clickstream data was not available. We consider ad exposure data for consumers during the mandatory registration period,

---

Novak, and Peralta 1998).

<sup>3</sup> Name undisclosed to respect sponsor confidentiality.

since as with many sites, it is not possible to track and measure consumer exposure to advertising across visits accurately in the absence of an identification procedure.

*Selection of Sponsors:* There were a total of 3810 Web pages at the Web site; 3046 (79.95%) were editorial pages with no ads (or pure editorial pages), 307 (8.06%) were editorial pages with banner ads, and 48 (1.26%) were active ad pages. While this site had 42 advertisers, complete data on passive banner exposures and consumer click decisions is available for only the 4 advertisers (identified as sponsors #15, #34, #35, #39) that had both banner and active ad pages on the publisher's Web server. The details of ad placements and exposures generated for these sponsors are provided in Table 1. Since a sufficient number of clicks were not available for advertisers #35 and #39 (only 8 and 12 respectively 39 because they were placed towards the end of the mandatory registration period) they were excluded from the analysis. Advertisers #15 and #34 are both high-technology firms, and their banner ads did not undergo any major execution changes or run promotional contests during the period of study. Despite less banner ads (2 for sponsor #15 vs. 6 for sponsor #34), banner exposures for sponsor #15 (i.e., 1,208,707) were significantly higher than that for sponsor #34 (86,251), most likely due to banner ad placement for sponsor #15 on entry or gateway pages which had relatively higher traffic.

*Selection of Consumers:* A total of 30,816 registered users visited the Web site under study from January 1, 1995 to December 30, 1995. For a distribution of Internet domains refer to Chatterjee (1998). The daily total of registered visitors ranged from 42,942 (on 7.27.95) to 57 (11.05.95), (probably because the publisher server was down), with a daily average of 21,850 (mode=28,664).

Table 1 Advertisements for Sponsors with Fixed Banner Ads

Advertiser ⇒	#15	#34	#35	#39
Number of banner pages	2	6	4	14
Number of active ad pages	1	1	1	1

Banner placement in information sections	34	13, 55	53	54
Banner exposures in mandatory registration period	1208707	86251	499171	38903
Active exposures in mandatory registration period	19070	1083	8	12
Click rate (mandatory registration period)	1.57%	1.26%	0.0016%	0.031%
Banner exposures in voluntary registration period	3952	34569	80310	27710
Active exposures in voluntary registration period	55	569	1351	2107
Click rate (voluntary registration period)	1.39%	1.65%	1.68%	7.60%
Overall share of passive ad exposures at Web Site	6.87%	1.41%	0.36%	4.9%
Overall share of clicks at Web Site	2.85%	2.71%	3.92%	1.75%

A sample of registered consumers was selected using the following selection criteria: consumers must have (i) registered at the Web site during the mandatory registration period (21,783), (ii) been exposed to at least one banner ad either for advertiser #15 or #34 during the mandatory registration period (8,497), (iii) accessed at least one Web page in the thirteen information “parent” section areas during the mandatory registration period (4,566), (iv) clicked on at least one passive ad for any advertiser at the Web site during the study period (2,085), and (v) registered navigation activity at the site during the calibration and prediction period (1,056). This selection rule yielded 1,056 consumers with 602,930 Web page accesses during the 8-month mandatory registration period. Navigational activity of 1,056 consumers were tracked over 257 days (01.08.95-07.14.95) for the model estimation and over 29 days (07.15.95-08.13.95) to test predictive ability.

### 4.3. Modeling Consumer Navigation

Consumer network navigation during a particular **session** includes **visit(s)** to one or many Web sites until the consumer exits the Web environment. The time between the consumer entry and exit from a Web site is defined as the **visit duration** (or **session duration** if only one site was visited during the session). Popular ad-supported sites carry ads for many sponsors, and one or more passive ads on any Web page. Passive ads for each sponsor may be placed (sometimes dynamically) on one or more editorial Web pages at the site. Every time a consumer clicks on an internal hyperlink a record or “hit” is made *for each of the files* comprising the destination document in the server’s access log. The content organized on Web pages at an ad-supported Web site can be classified into three categories: editorial (includes pages with passive ads for sponsor  $k$ , pages with passive ads for competing sponsors and pages with no ads), focal sponsor  $k$ ’s active ad pages, and active ad pages of other sponsors. Navigating through content at the Web site can be conceptualized as conferring utility to the consumer by providing information or intrinsic enjoyment of the browsing activity. The consumer’s utility function for soliciting information at the Web site during a visit is,

$$V_j(\text{ClickAd}_k, \text{Ed}, \text{PAd}_k, \text{OthClickAd}, \text{OthPAd}, \text{VC}_i, A_j) \quad (1)$$

where  $\text{ClickAd}_k$  is number of active ad pages accessed for sponsor  $k$ , and  $\text{Ed}$  is editorial content browsed by consumer during a particular visit.  $\text{PAd}_k$  is the number of passive ad exposures which may affect the utility a consumer may obtain from clicking on a passive ad.  $\text{OthPAd}$  and  $\text{OthClickAd}$  represent number of passive and active ads of competing sponsors viewed by the consumer during the same visit to the Web site.  $\text{VC}_j$  represents visit characteristics that affect the consumer’s utility by clicking on active ad  $k$ .  $A_j$  stands for consumer-specific demographic and unobserved psychological factors in utility. The opportunity to respond to, or click on a passive ad for the focal sponsor refers to the extent to which constraints of time affect attention to and processing of

brand information on the focal ad. The source of distraction<sup>4</sup> relevant during site navigation is “demand from competing activities.” When a consumer encounters passive ads during network navigation, the editorial content at the site competes with active advertising content for the focal sponsor  $k$  and other sponsors for share of consumer navigation time and clicks. The duration of a visit ( $VDur_j$ ) is the sum of time spent interacting with editorial pages, the time spent interacting with focal sponsor  $k$  active advertising ( $t_{ad} * ClickAd_k$ ) and time spent interacting with other sponsor active ad pages ( $t_{othad} * OthClickAd$ ). Hence consumers can be assumed to maximize their utility subject to the time constraint

$$VDur_j = Z + (t_{ad}) ClickAd_k + (t_{othad}) OthClickAd + (t_{ed}) (Ed + PAd_k + OthPAd) \quad (2)$$

where  $VDur_j$  is total time spent at a site by consumer  $i$  during the present visit,  $t_{ad}$  is average time spent at each active advertising page for sponsor  $k$ , and  $t_{ed}$  is average time spent at an editorial page. Note that editorial pages are comprised of pure editorial pages  $Ed$ , and editorial pages with passive (banner) ads  $PAd_k, OthPAd$ . Non-negativity constraints  $ClickAd_k \geq 0, Ed \geq 0, OthClickAd \geq 0, OthPAd \geq 0, PAd_k \geq 0$ , must be met. The resulting demand equations show that continuous optimum is given by

$$ClickAd_k^* = c(PAd, OthClickAd, OthPAd, VDur, Ed, A) \quad (3)$$

Thus, the observed decision to click on passive ad (i.e., to interact with the active ad) is a function of number of passive ad exposures, other sponsor passive and active ad exposures, editorial content accessed at the site, visit duration, and unobservable consumer-specific factors. Demographic, psychological and individual differences in involvement with ad/brand/sponsor are a part of the unobservable factors affecting consumer’s utility for clicking on the passive ad for the sponsor  $k$ .

---

<sup>4</sup> Sources of distraction in consumer processing of ads and editorial content identified by MacInnis, et al. (1991) include the differences in speed of information flow and consumer processing, etc., which are not relevant in the Web medium.

### Click Response Model

Let  $PAd_{ikt}$  denote the event that consumer  $i$  is exposed to the passive ad for sponsor  $k$  during visit  $t$ . If  $ClAd_{ikt}$  denotes the passive ad click decision by consumer  $i$  conditional on the consumer being exposed to passive ad (banner) then  $ClAd_{ikt} = 0$  if passive ad for sponsor  $k$  was not clicked on visit  $t$  and  $ClAd_{ikt} = 1$  if the passive ad was clicked and active ad viewed. We assume that the consumer click on sponsor  $k$  passive ad indicates that the balance of benefits in satisfying the consumer's navigational goals is positive and that not doing so indicates that the balance of benefits is negative. If the editorial page has passive ads for more than one sponsor, the probability of clicking on passive ad for sponsor  $k$ , conditional on passive ad exposure ( $PAd_{ikt} = 1$ ) and consumer entry ( $R_{it} = 1$ ), can be modeled as a multinomial logit function,

$$\Pr(ClAd_{ikt} = 1 | PAd_{ikt} = 1, R_{it}) = \frac{(\exp V_{ikt})}{(\sum_{j \neq k} \exp V_{ijt})} \quad (4)$$

$R_{it}$  denotes the consumer decision to enter the site. Note that the consumer has a choice of clicking on a passive ad for any one of the  $J$  advertisers on that page or *not clicking on any passive ad* at all, resulting in  $(J+1)$  alternatives. The most advantageous format for the advertiser is a single passive ad on a Web page, thus eliminating proximity clutter at the level of the Web page. Since the ad-supported Web site in this study had only one passive ad (i.e., banner) on each page, we develop the model assuming one passive ad on each Web page. Then the probability of clicking on the passive ad on a Web page can be modeled with the binary logit function

$$\Pr(ClAd_{ikt} = 1 | PAd_{ikt} = 1, R_{it}) = \frac{\exp(V_{ikt})}{(1 + \exp(V_{ikt}))} \quad (5)$$

where the deterministic component of click response utility,  $V_{ikt}$ , is written as

$$V_{ikt} = \mathbf{b}_0 + \mathbf{b}X_{ikt} \quad (6)$$

The parameter  $\beta_0$  is an intercept term, and  $\beta$  is a vector of response coefficients for variables  $X_{ikt}$ .

The terms in the  $X_{ikt}$  vector are discussed in table 2. The analysis (including the measurement of ad

exposures) is at the level of visit.

The operationalization of variables is discussed in the appendix. Consumers are heterogeneous with respect to their involvement with the product/product class/brand/

Table 2 Effect of Explanatory Variables on Click Response

Hypotheses	Explanatory variables in Utility Specification - $X_{ikt}$ vector	Notation	Hypothesized Relationship
H1a	Repeated exposures to passive ads for sponsor $k$ during the visit occasion	$\Sigma PAD_{ikt}$	$b_{1a} < 0$
H1b	Nonlinear response to repetitive passive ad exposures for sponsor $k$ during visit	$(\Sigma PAD_{ikt})^2$	$b_{1b} > 0$ , $ b_{1a}  >  b_{1b} $
H1c	Cumulative effect of exposure to passive ads for sponsor $k$ across visits to the site	$CumPAD_{ik(t-1)}$	$b_{1c} \neq 0$
H2a	Clutter effects of competing passive ads in present visit	$(\Sigma OthPAD_{ikt})$	$b_{2a} < 0$
H2b	Clutter effects of competing active ads in present visit	$(\Sigma OthCIAd_{ikt})$	$b_{2b} < 0$
H3a	Cumulative effect of exposure to active ads for sponsor $k$ across visits to the site	$CumCIAd_{ik(t-1)}$	$b_{3a} \neq 0$
H3b	Consumer has already accessed active ad for sponsor $k$ earlier during the visit	$CIAd_{ikt} = 0, 1$	$b_{3b} > 0$
H4a	Consumer interaction: duration of the visit	$VDur_{it}$	$b_{4a} > 0$
H4b	Consumer interaction: number of pure editorial pages browsed during the visit	$Ed_{it}$	$b_{4b} < 0$

sponsor and exposure to ads. Observing past behavior can capture much of the heterogeneity. We include the terms  $CumCIAd_{ik(t-1)}$  (hypothesis 3a) and  $CumPAD_{ik(t-1)}$  (hypothesis 1c) in the utility equation to capture these cumulative effects of exposure to sponsor active and passive ads in consumers' earlier visits to the Web site. The cumulative effect of active ad exposures in earlier visits captures the enduring involvement of the consumer with the product/product class/brand/sponsor.  $CumCIAd_{ik(t-1)}$  is defined as

$$CumCIAd_{ik(t-1)} = g_a CumCIAd_{ik(t-2)} + (1-g_a) CIAd_{ik(t-2)} \quad (7)$$

where  $CIAd_{ik(t-2)} = 1$  if consumer  $i$  clicked on the passive ad for sponsor  $k$  on exposure at visit occasion (t-2), and  $CIAd_{ik(t-2)} = 0$  otherwise.  $g_a$  is the carryover constant, as in Guadagni and Little's

(1983) study. Operationally cumulative effect of past active exposures is taken as an exponentially weighted average of past click behavior treated as 0-1 variables. The cumulative effect of passive exposures in prior visit occasions  $CumPA_{ik(t-1)}$  is operationalized similarly.

The single-segment approach discussed above assumes that consumers share common response coefficients to focal sponsor passive ads and competing sponsor ads. To relax this assumption, this approach calls for a multiple-segment approach. The segment-level click response probability may be written as

$$\Pr_s (ClAd_{ikt} = 1 | PAd_{ikt} = 1, R_{it}) = \frac{\exp(V_{ikt} | s)}{1 + \exp(V_{ikt} | s)} \quad (8)$$

where  $V_{ikt}/s$  is utility based on response parameters specific to segment  $s$ . The deterministic portion of utility within segment  $s$  is

$$V_{ikt} | s = \mathbf{b}_{0s} + \mathbf{b}_s X_{ikt} \quad (9)$$

The  $\beta_s$  vector differs across segments accommodating heterogeneity in response parameters.

## 5. Results

### 5.1. Click Response Model

Table 3 reports the analyses of consumers' click decisions for sponsor #15 using equations 4 and 5. The order of variables in the full model results for sponsors #15 and #34 depends on their relative contribution to the likelihood ratio index  $U^2$  (Guadagni and Little 1983), i.e., the variable that contributes most to  $U^2$  enters first, followed by the variable with the second largest contribution. The coefficients of cumulative effect of passive ad exposures in prior visits (H1a),

number of clicks on competing sponsor passive ads, visit duration and number of editorial pages browsed were insignificant and did not lead to any improvements in predictive ability of the model. Hence hypotheses 1c, 2b, 4a and 4b respectively were rejected by the data, and the corresponding variables were dropped from the model. In terms of *t*-statistics or partial correlation coefficients, the prior click response during the visit is the best explanatory variable of click decision supporting hypothesis 3a; i.e., if a consumer has clicked on a banner for sponsor once, s/he is more likely to click on it again on exposure during the same visit. On the same bases, the next 3 important variables are the number of passive (banner) ad exposures during the visit (H1a), the cumulative effect of active ad exposures in earlier visits (H3a), and the number of competing banner exposures during the same visit (H2a). Cumulative or enduring click behavior does not explain click decision as much as present visit or situational click behavior does. The intercept would be zero (if all variables have zero mean) and insignificant if all variables that explained the difference between click/no-click response were found.

Note that both the linear and quadratic terms of the effects of passive advertising (or banner) exposure are significant. Further, the coefficient of linear effect of passive advertising is negative and significantly larger than the positive quadratic term. Thus, Hypotheses 1a and 1b are supported. This indicates that passive ad wearout occurs earlier in the Web medium. The small positive, but significant quadratic coefficient indicates that the negative effect of repeated passive ad exposures decreases and changes to an increasing positive effect after a finite number of repeated passive exposures. This is in sharp contrast to results obtained for television ad exposures in earlier empirical studies (Stewart and Furse 1986; Tellis 1988). Marginal effect analysis indicates that the probability of click decreases by 0.0526 for every unit increase in number of passive ad exposures for sponsor #34 beyond 0.999. Hence for sponsor #34 increasing the number of passive ad exposures during a visit beyond 1 does not improve clickthrough probability.

Table 3

## Binary Logit Analysis for Passive Ad (Banner) Click Response

(Note: t - statistics are in parentheses)

Specification $\Rightarrow$	Full Model		Selected Model		M1	M2	M3
Sponsor $\Rightarrow$	#15	#34	#15	#34	#34	#34	#34
$U^2$	0.5117	0.216	0.5117	0.214	0.049	0.187	0.164
<b>H3b:</b> $ClAd_{kt}$ Clicks present visit	6.756 (34.82)**	3.24 (13.9)**	6.769 (35.23)**	3.347 (11.72)**	----	----	3.135 (11.2)**
<b>H1a:</b> $SPAd_{kt}$ Banner exposures present visit	-1.184 (-10.7)**	-0.915 (-6.5)**	-1.179 (-10.7)**	-1.100 (-5.63)**	-1.074 (-7.3)**	-1.109 (-6.2)**	-1.22 (-5.9)**
<b>H3a:</b> $CumClAd_{k(t-1)}$ Cumulative effect of prior visit clicks	14.926 (9.9)**	16.49 (13.3)**	14.925 (9.982)**	14.49 (8.86)**	----	14.642 (8.88)**	----
<b>H1b:</b> $(SPAd_{kt})^2$	0.009 (7.3)**	0.049 (1.94)**	0.009 (7.363)**	0.059 (0.89)	0.068 (3.01)**	0.07 (2.4)**	0.05 (0.7)
<b>H2a:</b> $SOthPAd_t$ Competing banner exposures present visit	0.0312 (1.67)*	0.028 (1.7)*	0.0333 (1.867)**	0.02 (1.67)*	0.034 (3.36)**	0.01 (0.6)	0.02 (1.7)
<b>H2b:</b> $SOthClAd_t$ Clicks on other advertisers present visit	0.0302 (0.477)	-0.206 (-1.3)	----	----	----	----	----
<b>H1c:</b> $CumPAd_{k(t-1)}$ Cumulative effect of prior visit passive exposures	-2.17 (1.42)	-1.96 (-0.6)	----	----	----	----	----
<b>H4a:</b> $VDur_t$ Visit duration	-0.003 (0.003)	-0.003 (-1.1)	----	----	----	----	----
<b>H4b:</b> $Ed_t$ Editorial pages browsed	0.015 (0.009)	0.0001 (0.024)	----	----	----	----	----
Intercept	-2.90 (17.3)**	5.41 (3.46)**	-2.902 (-17.3)**	-6.075 (-12.6)**	-1.9 (-8.5)**	-6.14 (-12.6)**	-2.03 (-8.2)**
Log Likelihood	-1118.1	-541.7	-1118.03	-536.38	-648.21	-486.7	-570.2
Correct prediction %	98.1	93.5	98.10	93.64	90.31	91.74	92.92
Clicks correctly predicted %	40.78	19.94	40.79	19.94	4.48	9.37	17.26

\*\* Significant at 0.001 level \* Significant at 0.05 level

The effect of a unit increase in passive ad exposure decreases the probability of clicking a banner for sponsor #15 by 0.006 at 1.287. The cumulative effect of passive ad exposures is insignificant indicating that long-term exposure effects do not induce consumers to click in this data, hence hypothesis 1c is rejected. Competing sponsor passive ads have a significant but positive effect on click probability thus the directionality of hypothesis 2a is not supported. Hypothesis 2b is rejected, the effect of competing sponsor active ads is insignificant.

In the holdout sample, the model correctly predicts 98.1% of consumer click decisions for sponsor #15. However this prediction accuracy should be interpreted with caution, given that there are only 457 (40.79% predicted correctly) observations for a positive click response, and 15,858 (99.7% predicted correctly) observations for no click response. For sponsor #34 the model correctly predicts 93.64% of consumer click decisions in the holdout sample; there are only 131 (19.94% predicted correctly) observations for a positive click response, and 2,251 (99.2% predicted correctly) observations for negative click response. The average predicted probability in logit/probit models equals the proportion of individuals in the sample who experienced the outcome/event of interest. When the base probability of outcome is very high or very low (as in this model) van Houwelingen and le Cessie (1990) suggest that in the interest of sensitivity and specificity the cutoff point should be raised or lowered to equal the mean of the sample. The model estimates are fairly similar for sponsor #34 except for a poorer fit ( $U^2=0.214$  and larger value for the intercept) and insignificant coefficient for quadratic term of passive ad exposure.

The necessity of including both the present visit effect ( $CIAd_{k,t}$ ) and cumulative effect of active exposures ( $CumCIAd_{k(t-1)}$ ) is demonstrated in Table 3 for sponsor #34. The specification M1 in Table 3 shows the estimation results if both  $CIAd_{k,t}$  and  $CumCIAd_{k(t-1)}$  are dropped from the estimation and other variables are retained. The signs of the other variables remain the same, but the quadratic component of repeated passive ad exposure gains significance. However, the

predictive ability of the model drops drastically, as does its explanatory power ( $U^2=0.049$ ). Including the cumulative effect of prior visit active exposures only (specification M2) improves the predictability and explanatory power to a lesser extent than the specification M3 which includes the term  $ClAd_{k,t}$  capturing situational involvement of the consumer. Thus both the cumulative effect ( $CumClAd_{k,(t-1)}$ ) and present visit effect ( $ClAd_{k,t}$ ) of active exposures must be included in the model to get maximum predictability of click response.

## 5.2. Segmenting Web Site Consumer Base

Information regarding the frequency of consumer visits to the ad-supported Web site is

Table 4 Descriptive Statistics for Visitor Segments of Consumers at Web Site  
(Note: Standard deviations are in parentheses)

	Segment 1	Segment 2	Segment 3	Segment 4	Total Sample
No. of visits (range)	3-8	9-19	20-42	42-601	1-601
Visit duration $VDur$	9.8485 (28.0927)	11.683 (30.9955)	15.452 (34.3385)	25.304 (56.4394)	20.531 (45.39)
Passive ad exposures per visit	2.2697 <sup>a</sup> (3.3342)	2.2034 <sup>a</sup> (2.7986)	2.4850 (2.9847)	2.7846 <sup>b</sup> (3.7565)	2.5502 <sup>b</sup> (3.3903)
No. of clicks on ads per visit	0.2490 <sup>a</sup> (0.8353)	0.2150 <sup>b</sup> (0.6988)	0.1536 <sup>c</sup> (0.6238)	0.0910 <sup>d</sup> (0.4054)	0.1743 <sup>e</sup> (0.5971)
Segment size	271(0.256)	255(0.241)	272(0.257)	258(0.244)	1056

Differently lettered superscripts across segment means denote differences that are significant at 0.05 level; segment means with like superscripts are not significantly different from each other; segment means with no superscript are not significantly different from the means that are immediately higher or lower.

used to segment the consumers at the Web site. The consumers in the research sample were segmented into 4 groups of roughly equal sizes based on their total number of visits to the Web site during the entire mandatory period. Table 4 shows the descriptive statistics for consumers in the four visitor segments. As frequency of visits increase, consumers tend to stay longer at the site, are exposed to more passive ads, but click on fewer passive ads (or browse through active ads) during visits.

Segmenting consumers on the basis of their frequency of visits yields many insights. Tables

5 and 6 show the results of estimating the click response model for the four visitor

Table 5 Visitor Frequency Segment Analysis for Click Decision (Advertiser #15)  
(Note: t - statistics are in parentheses)

Specification $\Rightarrow$	Segment 1	Segment 2	Segment 3	Segment 4	Total Sample
$U^2$	0.38	0.403	0.58	0.553	0.512
<b>H3b:</b> $ClAd_{kt}$ Clicks present visit	5.835 (7.712)**	5.68 (9.731)**	7.71 (14.25)**	6.923 (29.075)**	6.769 (35.23)**
<b>H1a:</b> $SPAd_{kt}$ Banner ad exposures present visit	-0.329 (-0.982)	-0.618 (-1.004)	-0.463 (-1.22)	-1.30 (-8.178)**	-1.179 (-10.69)**
<b>H3a:</b> $CumClAd_{k(t-1)}$ Cumulative effect of prior visit clicks	51.661 (7.495)**	24.57 (7.527)**	18.81 (6.197)**	5.235 (1.878)*	14.925 (9.982)**
<b>H1b:</b> $(SPAd_{kt})^2$	0.011 (0.257)	-0.607 (-1.768)*	-0.285 (-3.449)**	0.0101 (6.981)**	0.009 (7.363)**
<b>H2a:</b> $SOthPAd_t$ Competing banner ad exposures present visit	-0.0037 (-0.052)	0.128 (2.076)**	0.036 (0.597)	0.054 (2.17)**	0.0333 (1.867)**
Intercept	-21.54 (-8.779)**	-12.65 (-10.58)**	-10.60 (-9.89)**	-2.833 (-17.23)**	-2.902 (-17.31)**
Correct prediction %	99.64	98.04	97.94	99.21	98.1
Clicks correctly predicted %	41.61	36.55	44.7	45.72	40.79
Segment Share of Consumers	271 (0.256)	255 (0.241)	272 (0.257)	258 (0.244)	1056

\*\* Significant at 0.001 level

\* Significant at 0.05 level

frequency segments, respectively for sponsor #15 and #34. To save space we report only those variables that show statistical significance for at least one of the sponsors. The coefficients of variables representing situational and enduring involvement ( $ClAd_{kt}$  and  $CumClAd_{k(t-1)}$  resp.) are unambiguously positive and significant. The heterogeneity across consumers in their involvement and prior click behavior in predicting click response on next exposure to passive ad has been supported by the data (hypotheses 3a and 3b). Both of the terms individually contribute in improving predictability of the click response. The multiple-segment analysis

Table 6 Visitor Frequency Segment Analysis for Click Decision (Advertiser #34)  
(Note: t - statistics are in parentheses)

Specification $\Rightarrow$	Segment 1	Segment 2	Segment 3	Segment 4	Total Sample
$U^2$	0.451	0.224	0.099	0.103	0.214
<b>H3b:</b> $ClAd_{kt}$ Clicks present visit	4.861 (6.19)**	2.745 (3.476)**	2.442 (3.608)**	3.129 (4.956)**	3.347 (11.72)**
<b>H1a:</b> $SPAd_{kt}$ Banner ad exposures present visit	-0.956 (-0.973)	-0.489 (-0.467)	-0.319 (-0.677)	-1.136 (-3.86)**	-1.100 (-5.634)**
<b>H3a:</b> $CumClAd_{k(t-1)}$ Cumulative effect of prior visit clicks	16.33 (4.196)**	19.569 (5.632)**	12.34 (3.749)**	9.312 (2.39)**	14.49 (8.858)**
<b>H1b:</b> $(SPAd_{kt})^2$	0.0136 (0.030)	-0.282 (-0.424)	-0.017 (-0.095)	0.0702 (1.815)*	0.0585 (0.890)
<b>H2a:</b> $SOthPAD_t$ Competing banner ad exposures present visit	0.0494 (1.64)	-0.058 (-0.822)	0.0451 (1.379)	-0.0062 (-0.216)	0.0213 (1.668)*
Intercept	-6.653 (-5.35)**	-7.415 (-7.031)**	-5.318 (-5.653)**	-4.962 (-4.603)**	-6.075 (-12.569)**
Correct prediction %	91.01	91.84	87.47	96.55	93.42
Clicks correctly predicted %	67.39	37.14	10.71	3.11	19.94
Segment Share of Navigation Activity	823 (0.097)	1103 (0.13)	1384 (0.163)	5178 (0.61)	8488
Segment Share of Consumers	271 (0.256)	255 (0.241)	272 (0.257)	258 (0.244)	1056

\*\* Significant at 0.001 level

\* Significant at 0.05 level

predicts better than the single-segment analysis, 99.13% of click decisions and 44.81% clicks were predicted correctly across all segments for sponsor #15. The corresponding percentages for sponsor #34 are 93.9% and 24.36%. The effect of competing sponsor passive ad exposures is mostly insignificant, except for sponsor #15 in segments 2 and 4 indicating that passive ads of competitors displayed while the consumer browses through other pages at the Web site are not detrimental to the focal sponsor in the site under study.

A major change is observed in the effect of repeated banner exposures; which is negative and significant (as hypothesized) only in segment 4 for both sponsors. Moreover, the quadratic

term capturing the non-linear effect of repeated passive ad exposures has a significant negative sign in segments 2 and 3 for sponsor #15, indicating that the negative response to passive ad exposures for sponsor #15 is non-linear in segments 2 and 3 and is captured by the quadratic term.

Consumers in segment 1 visit the site least frequently are not affected by repeated passive ad exposures for both sponsors at all. The probability of click response by frequent patrons drop by 0.005 for sponsor #15 for every unit increase in passive ad exposure at mean 1.445 passive ad exposures, twice that of consumers in the other segments, yielding higher negative returns.

Inspection of the data and the correlation matrix does not indicate collinearity problems. Since consumers who visit the Web site frequently drive most of the estimation results in the single segment model because of their huge share of navigation activity, the gradual (or non-linear) negative response cannot be detected if consumers are not segmented. Consumers in segment 3 have the highest mean probability of clicking on passive ad for sponsor #34 followed by segment 2 and segment 1. As the number of passive ad exposures increases beyond 2 the click probability of consumers in segment 1 is higher than that of consumers in segment 2. Table 7 summarizes the results and the hypotheses supported by the data.

Table 7 Summary of Findings

Effects under investigation	Hypotheses	Empirical Findings
Repeated exposures to passive ads	H1a: $\mathbf{b}_{1a} < 0$	Supported by both sponsors
Nonlinear response to repetitive passive ads	H1b: $\mathbf{b}_{1b} > 0$ , $ \mathbf{b}_{1a}  >  \mathbf{b}_{1b} $	Supported by sponsor #15
Cumulative effect of exposure to passive ads	H1c: $\mathbf{b}_{1c} > 0$	Not significant for both sponsors
Exposure to competing passive ads	H2a: $\mathbf{b}_{2a} < 0$	Opposite effect significant for both sponsors
Cumulative effect of exposure to active ads	H3a: $\mathbf{b}_{3a} > 0$	Supported by both sponsors
During the visit clicks	H3b: $\mathbf{b}_{3b} > 0$	Supported by both sponsors
Consumer interaction: visit duration	H4a: $\mathbf{b}_{4a} > 0$	Not significant for both sponsors
Consumer interaction: pure editorial pages read	H4b: $\mathbf{b}_{4b} < 0$	Not significant for both sponsors

## 6. Discussion

Our empirical results indicate that present visit click behavior contributes the most to the prediction of click response. Unlike the repetition effects in broadcast (off-line) media, the results show that response to repeated passive ad exposures is non-linear, with initial response negative and a small positive, but significant non-linear component. This implies that while the effect of repeated exposures to passive ads is negative in the short-term (during the same visit), there is a positive long-term effect which will be recorded only if the consumer is exposed to passive ads for a sponsor more than 14 times during a visit for sponsor #15 (11 for sponsor #34). Further the slope of negative linear response increases with increase in visitor frequency. While past visit click behavior has a significant positive effect on probability of click, passive ad exposures during prior visits do not. Competing passive ads have a small positive but significant effect on click probability, but the effect of competing active ad exposure is insignificant.

This may be due to the fact that a click on a passive ad implies some commitment to process ad messages for a particular sponsor, which should not have any impact on the consumer's involvement with or click probability for any other sponsor. Thus presence of competing ads during navigation does not have a negative impact on click probability of focal sponsor as long as they are not on the same page as the focal sponsor. However, it is possible that our hypothesis regarding response to competitive passive advertising exposure (H2a) might have been supported if we had been able to evaluate the effect of multiple passive ads per page.

*Banner Ad Burnout:* Placing passive ads at an advertiser-supported Web site requires at least two major decisions – how many and where. Advertisers typically desire to place as many passive ads as their online ad budgets allow in order to generate the maximum number of impressions or page views for their passive ads. However, if maximizing clickthrough is the marketing outcome of interest, our results indicate that increasing numbers of passive ad exposures to individual

consumers may initially generate decreasing returns. However, H2a suggests the positive effect of many different banners for the same sponsor throughout the site. This result indicates that ads for sponsor #34 should be placed (dynamically) in such a way that no consumer is exposed to more than two exposures (and four exposures for sponsor #15) of the same passive ad during a visit.

Note that the effect of repeated exposures to passive ads will be moderated by the frequency of change in passive ad executions as proposed by the “encoding variability hypothesis” (Unnava and Burnkrant 1991). It was not possible to investigate these effects directly since data on ad execution features were unavailable. However, hypothesis H2a seems to suggest that advertisers may be able to forestall wearout due to repetition by changing ad copy and execution frequently. This is relatively easy to achieve in the Web environment, given its very low marginal production costs.

*Identifying Attractive Response Segments:* Segmenting consumers improves predictive ability of click response and identifies differences in response coefficients. Consumers who visit more frequently have a higher probability of clicking on passive ads but are more intolerant of repeated passive ad exposures during the same visit. The effect of competing passive ad exposures is increasingly positive with increase in visitor frequency. This suggests that if a higher proportion of consumers at a Web site are frequent visitors, placing few ads for each sponsor and increasing the number of sponsors will likely be a more effective strategy.

Segmenting consumers on the basis of their content choices may yield useful insights; however, the lack of detailed information about content prevented us from doing so. Response-based segmentation can be used to quantify the relative sizes of the response segments and investigate how consumers in each of these segments differ in their propensities to respond to a particular sponsor’s passive ad. This has implications for media efficiency. If a similar exercise is conducted across sites, an advertiser can rank sites on the basis of their sizes of “response prone”

segments. This will provide valuable input to any media planning exercise. Given the nature of data sophisticated segmentation procedures will no doubt become popular in the future.

*The Identification Decision:* The ability of a media vehicle to deliver qualified advertising exposures largely determines its ability to command attract advertising revenues. Most advertiser-supported Web sites implement identification procedures using cookies or registration. Most, however, are reluctant to require mandatory registration. This research highlights the importance of implementing mandatory identification procedures to enhance the value of clickstream data. Past click behavior at the site across visits can be used to capture heterogeneity across consumers in their involvement with the sponsor/ad/product and their click propensity. This was an important component of the model across all segments for ads of both sponsors in this study. In absence of a mandatory identification procedure, analysis of data over time will be limited to cross-sectional effects and will lead to considerably weaker performance in terms of prediction ability.

*Limitations:* We emphasize the preliminary nature of our analysis. We concentrated on click behavior of consumers at one advertiser-supported Web site. Investigation across a network of sites would provide more realistic results, since consumers typically visit more than one site during a session and there may be spillover effects of ad exposure. However difficulties in identifying navigational activity for individual consumers over time across real sites remains as a problem. Consumer-centric clickstream data may provide a solution to this problem.

We operationalize passive ad variables in terms of the number and not duration of exposures. Further, we were not able to investigate the effect of executional features of passive ads (Dreze and Zufryden 1997) because of lack of data. The effect of exposure duration and its interaction with executional features may yield important insights. We developed the model assuming one passive ad placed on each page. While this is the dominant placement policy at most sites, and extending the model accommodate multiple banners on each page may be theoretically

possible, estimating the extended model will be a challenge. Since each passive ad can appear with a different set of passive ads at each click choice occasion, there will be  ${}^aC_p$  click choices if there are  $a$  advertisers and  $p$  ads on each Web page. The estimation becomes very difficult for most Web sites that are ad-supported. We use visitor frequency as the basis for segmenting consumers. Clearly, other bases that capture the effect of editorial content in inducing click behavior and exploit the personalization ability of the medium in delivering advertising stimuli need to be considered as more Web sites start using collaborative filtering techniques. Consumer preferences of editorial content can be analyzed to identify segments that are more or less responsive to an advertiser's communication (Chatterjee, 1998).

The inability of log files to precisely record exposure to floating options and external active ads for competitors implies that the findings due to clutter effects, in this study, need to be interpreted with caution. If Web pages at the site carry more than one banner, the complexity increases. This is a matter of utmost concern to publisher sites dependent on ad revenues. If passive ad exposures for other sponsors encountered during navigation create a significant negative effect due to clutter, it has implications for the maximum number of clients and banners a site can accommodate in order to generate an appreciable number of clickthroughs.

## **7. Conclusions**

In this paper, we have proposed a model for predicting advertising click response at an advertiser-supported Web site using clickstream data gathered at the individual level. Empirical analysis of clickstream data collected over a seven month period at an ad-supported site was used to study the effect of other advertising stimuli encountered by the consumer during navigation and heterogeneity in within-visit and across-visit click behavior at the Web site.. The results suggest that increasing the number of same passive ad insertions will lead to negative returns initially and level off at higher level of exposures. The negative impact will increase with increase in frequency of

visits to the site. The publisher can accommodate many different advertisers at the site, without adverse effects as long as a single passive ad is placed on a page. This research answers the call by researchers for testing advertising theories in realistic contexts (Stewart and Furse 1986) over time.

To our knowledge this is the first paper to use clickstream data of navigation at an actual commercial Web site to study consumer response to advertising over time. The benefits of using clickstream data as discussed above notwithstanding, empirical analysis of clickstream data from any commercial Web site poses many challenges. First, the task of parsing and analyzing the data even at a preliminary level is gargantuan, records of a single month of activity may occupy 75 megabytes at the least. Second, an interdisciplinary approach is required to understand the process of consumer behavior in online environments in order to interpret the data. Knowledge of how programs operate to log activity and network traffic modeling is needed to interpret the data generated by the Web servers before modeling approaches can be developed to investigate advertising and consumer behavior. This paper represents an initial approach to analyzing and interpreting data generated from the clickstream and identifies problems that need to be addressed in collecting clickstream data. This work may have relevance beyond advertising, given the potential of clickstream data to other marketing problems.

Although we treat consumer response to banner ads, our framework can be applied to other passive ads, including Virtual-tag enhanced banners or Java/Shockwave enabled intermercials. The framework may also be extended to direct advertising like infomercials, direct mail and other forms of electronic advertising (e.g. online and videotex services) where clickstream data are captured.

*Future Research:* Numerous extensions of this model await further research. The ability of editorial content to attract consumers who are more (or less) responsive to advertising for a particular sponsor is intuitive. Use of network theory in analyzing consumer navigation patterns can lead to segmentation procedures that yield better predictability of response and help in decisions regarding

dynamic ad personalization and placement. Differences in response to passive exposures in each segment can be used to decide how many and which pages/ information sections should carry ads for a particular sponsor. Segmenting consumers on the basis of their browsing preferences and then examining the differences in response to ads for a particular sponsor will assist Web publishers in identifying editorial content that is most conducive to generating clickthroughs and repeat visits to the site. It will also provide insights into the types of content patronized by consumers who are most responsive to sponsor's ads, which the publisher can use in designing new content areas and developing existing popular content areas.

The continuing rapid evolution of media warrants understanding and analyses of the ways in which individuals interact with and act upon media (Stewart and Ward 1984). Consumer clickstream is a rich source of process tracing data, not only for investigating advertising effects, but also consumer information acquisition, and decision making strategies on choice tasks in real world choice environment, without large financial outlays involved in Mouselab or eye-tracking experiments (Lohse and Johnson 1996). Dynamic modeling of usage and navigation patterns over time will help provide insights on better interface, information structuring and advertising placement in this medium.

## Appendix. Variable Operationalization

As this is the first modeling analysis of clickstream data in the marketing literature, we discuss variable operationalization and descriptive statistics for the click response model. For detailed description of variables and their distributions refer to Chatterjee (1998).

*Active Ad Exposure:* The clickstream data consists of choice records for consumer click decisions, the dependent measure for the Click Response Model. The click response or active ad exposure variable ( $ClAd_{ik}$ ) is defined as  $ClAd_{ik} = 1$ , if the consumer “hits” the active ad Web page for sponsor  $k$  at visit  $t$ , 0 otherwise. Since clicks on ads may be inflated when consumers become impatient and click the page with the banner ad multiple times; we discarded those extra clicks. Only those clicks that were immediately preceded by a banner ad exposure are included in the analysis. The data indicates that more consumers in the sample click on being exposed to an ad for sponsor #34 than an ad for sponsor #15. 91% (71.4% for sponsor #34) of consumers who were exposed to banner ads for sponsor #15 at least once during the study period did not click at all. Please refer to Table 8 for details.

The explanatory variables for the Click Response Model (discussed in table 2) are all recorded at the level of a visit for each consumer. The total number of visits by consumers in the research sample during the study period is 19,628. The average number of visits per consumer is 18.58. *Visit duration* ( $VDur_{i,t}$ ) is measured as the time, in seconds, spent browsing Web pages at the site by a consumer  $i$  during visit  $t$ . *Passive ad exposure* ( $\acute{a}PAd_{ikt}$ ) is operationalized as the total number of passive (banner) ad exposures by consumer  $i$  for the focal sponsor during the visit till the choice situation. The range of passive ad exposures during a visit is 1 to 33. The data indicate that the probability of being exposed to more than 6

Table 8 Description of Advertising Exposure Variables

	Calibration period		Prediction period	
	#15	#34	#15	#34
Advertiser ⇒	#15	#34	#15	#34
Probability of banner ad exposure	0.134	0.0126	0.13	0.019
Probability of click on banner ad exposure	0.0209	0.056	0.024	0.053
Total number of banner ad exposures	64,196	6,047	16,315	2,441
Total number of banner ad clicks	1,348	339	457	131
Sample click rate at Web Site	2.09%	5.6%	2.4%	5.32%

passive ads during a single visit for sponsor #15 or sponsor #34 is extremely low. *Prior click response during the visit* ( $ClAd_{i,k,t}$ ) is operationalized as a dummy variable indicating if the consumer  $i$  has already clicked on a passive ad for sponsor  $k$  earlier in the visit. *Competitive (or other) passive ad exposure* ( $\dot{a}OthPAd_{i,k,t}$ ) is analogous and measured as the total number of exposures to passive ads for other advertisers at the Web site during the visit until the choice situation. In at least 82% of click decision situations, consumers encounter one or more passive ads for competing sponsors during the same visit. *Competitive (other) active ad exposure* ( $\dot{a}OthClAd_{i,k,t}$ ) is measured as the number of times a consumer clicked on a passive ad for competing advertisers prior to the click choice event for the focal sponsor. In 3.4% of click decision situations for sponsor #15 (1.3% for sponsor #34) consumers have already accessed an active ad for at least one competing advertiser in the same visit. *Cumulative effect of prior advertising exposures* has two parts. The *cumulative effect of prior active advertising exposures* ( $CumClAd_{i,k,t-1}$ ) is operationalized as the exponentially weighted average of past click decisions made on passive exposures during prior visits at the site, treated as 0-1 variables. The active advertising carryover constant is  $\gamma_a$ . *Cumulative effect of prior passive ad exposure* ( $CumPAd_{i,k,t-1}$ ) is analogous, and measured as the exponentially weighted average of passive exposures for the focal advertiser during prior visits at the site. The passive advertising carryover constant is  $\gamma_p$ . The

expressions for cumulative effects of advertising exposures are described in equation 7. Navigation activity of 1056 consumers in the initial 80 days was used to initialize cumulative advertising effects. The cumulative effect variables are particularly important since they carry not only much of the cross-sectional heterogeneity but also a good part of the visit-to-visit dynamics. The smoothing constants  $\gamma_a$  and  $\gamma_p$  are obtained by using an iterative scheme to get approximate values for  $\gamma_a$  and  $\gamma_p$ . To assess the global optimality of our results we performed a grid search. We derive estimates 0.847 for  $\gamma_a$  and 0.908 for  $\gamma_p$ .

## References

- Bawa, Kapil and Robert Shoemaker (1987), "The Effect of Direct Mail Coupon on Brand Choice Behavior," *Journal of Marketing Research*, 24 (November), 370-376.
- Briggs, Rex, and Nigel Hollis (1997), "Advertising on the Web: Is There Response before Click-Through?" *Journal of Advertising Research*, (March-April), 33-45.
- Carter, David (1968), "Newspaper Advertising Readership: Thick vs. Thin Issues," *Journal of Advertising Research*, 8(3), 39-41.
- Chatterjee, Patrali (1998), "Modeling Consumer Network Navigation at World Wide Web Sites - Implications for Advertising," doctoral dissertation, Vanderbilt University.
- Clark, Don (1997), "Facing Early Losses, Some Web Publishers Begin to Pull the Plug," *The Wall Street Journal*, January 14, A1, A8.
- Consumer Experience Probe (1996), "Web Advertising Report", Jupiter Communications/The Westport Company. [URL: <http://www.webtrack.com/interad/>]
- DoubleClick (1996), Frequency and Banner Burnout.  
[URL: <http://www.doubleclick.net/nf/general/frequetset.htm>]
- Dreze, Xavier and Fred Zufryden (1997), "Testing Web Site Design and Promotional Content," *Journal of Advertising Research*, March-April, 77-91.
- Ehrenberg, A. S. C. (1966), "The Time and Place for Readership Panels," *Journal of Advertising Research*, 8 (2), 19-22.
- Fitzgerald, Michael (1998), "Slate halfway to subscription goal," *ZDNET*, March 4.  
[URL: [www5.zdnet.com/zdnn/0304/290963.htm](http://www5.zdnet.com/zdnn/0304/290963.htm)]
- Guadagni, Peter M. and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (Summer), 203-238.
- GVU WWW User Surveys (1998), "GVU's 8th WWW User Survey."  
[URL: [http://www.cc.gatech.edu/gvu/user\\_surveys/](http://www.cc.gatech.edu/gvu/user_surveys/)]
- Hoffman, Donna L., Thomas P. Novak and M. A. Peralta (1998), "Information Piracy in the Marketplace: Implications for the Commercial Uses of Anonymity on the Web," conditionally accepted, *The Information Society*.
- Houston, F. S. and Diane Scott (1984), "The Determinants of Advertising Page Exposure," *Journal of Advertising*, 13 (2), 27-33.
- Hoyer, Wayne D. and Steven P. Brown (1990), "Effects of Brand Awareness on Choice for a Common Repeat-Purchase Product," *Journal of Consumer Research*, 17(2), September, 141-148.
- Internet Advertising Bureau (1997), "Q2 1997 Spending Reaches \$214.4 Million," Oct. 7.

[URL:<http://www.iab.net/news/content/new%20/q2.html>]

Internet Advertising Bureau (1998), "Internet Advertising Approaches \$1 Billion," April 6.  
[URL:<http://www.iab.net>]

Lohse, Gerald L. and Eric J. Johnson (1996), "A Comparison of Two Process Tracing Methods for Choice Tasks," *Organizational Behavior and Human Decision Processes*, 68(1), October, 28-43.

MacInnis, Deborah J. , Christine Moorman and Bernard J. Jaworski (1991), "Enhancing and Measuring Consumers' Motivation, Opportunity, and Ability to Process Brand Information from Ads," *Journal of Marketing* 55 (October), 32-53.

Meeker, Mary (1997), *The Internet Advertising Report*, Morgan Stanley.

Novak, Thomas P. and Donna L. Hoffman (1997), "New Metrics for New Media: Toward the Development of Web Measurement Standards," *World Wide Web Journal*, Winter 2(1), 213-246. [URL: <http://www2000.vanderbilt.edu/novak/web.standards/webstand.html>]

Pechmann, Cornelia and David W. Stewart (1989), "Advertising Repetition: A Critical Review of Wearin and Wearout," in *Current Issues and Research in Advertising 1988*, James H. Leigh and Claude R. Martin Jr., eds. Ann Arbor, MI: University of Michigan, 285-329.

Sawyer, Alan (1981), "Repetition, Cognitive Responses and Persuasion," in *Cognitive Responses in Persuasion*, eds. R. E. Petty, T. M. Ostrom and T. C. Brock. Hillsdale, NJ: Lawrence Erlbaum Associates, 237-261.

Schumann, David W., Richard E. Petty, and D. Scott Clemons (1988), "Predicting the Effectiveness of Different Strategies of Advertising Variation: A Test of Repetition-Variation Strategies Hypothesis," working paper from the College of Business Administration of the University of Tennessee at Knoxville.

Stewart, David W. and David H. Furse (1986), *Effective Television Advertising*. Lexington, MA: Lexington Books.

Stewart, David, Connie Pechmann, Srinivasan Ratneshwar, Jon Stroud and Beverly Bryant (1986), "Methodological and Theoretical Foundations of Advertising Copy testing: A Review," *Current Issues and Research in Advertising* 1-74.

Stewart, David W. and Scott Ward (1994), "Media Effects on Advertising," in *Media Effects, Advances in Theory and Research*, Jennings Bryand and Dolf Zillman, eds., Hillsdale, NJ: Lawrence Erlbaum Associates.

Tellis, Gerard J. (1988), "Advertising Exposure, Loyalty, and Brand Purchase: A Two-Stage Model of Choice", *Journal of Marketing Research*, vol. 25 (May), 134-44.

Unnava, H Rao. Burnkrant, Robert E (1991), "Effects of Repeating Varied Ad Executions on Brand Name Memory," *Journal of Marketing Research*., 28(4), Nov., 406-416.

Urban, Christine (1976), "Correlates of Magazine Readership," *Journal of Advertising Research*, 19(3)

June, 7-12.

Van Houwelingen, J.C. and le Cessie, S (1990), "Predictive Value of Statistical Models," *Statistics in Medicine*, 8, 1303-1325.