Article

Journal of Marketing Research
2019, Vol. 56(1) 57-75

DOI: 10.1177/0022243718820587
journals.sagepub.com/home/mrj

Wearout or Weariness? Measuring Potential Negative Consequences of Online Ad Volume and Placement on Website Visits

Inyoung Chae, Hernán A. Bruno, and Fred M. Feinberg

Abstract

The global importance of online advertising calls for a detailed understanding of consumer-specific responses to online ad repetitions. A key concern for advertisers is not only whether some consumers display degrees of "wearout" but also whether they can surpass a point at which additional exposures have a *negative* marginal effect: "weariness." The authors examine a large-scale advertising campaign aimed at driving viewers to a target website, which comprises more than 12,000 users across over 400 websites. These data are analyzed using a flexible discrete mixture specification that accommodates different response shapes over ad stock and timing and parcels ad viewers into response classes based on their internet usage metrics. The resulting classes display varying degrees of wearout, with one subgroup, accounting for about 24% of the sample, evincing weariness. The model also estimates differential publisher effectiveness, with the most effective publisher being nine times more effective than the one 26 places down. The authors demonstrate that the finding of weariness is robust to all the model's main components, with one key exception: heterogeneity in users' ad response. Analysis further suggests that an appropriate "profiling and capping" strategy can improve ad deployment by as much as 15% overall for these data.

Keywords

advertising models, heterogeneity, hierarchical Bayes model, online advertising

Online supplement: https://doi.org/10.1177/0022243718820587

Global advertising spend stands at over $530 billion, growing at a 4% clip over the last five years. Most of this growth is spurred by digital advertising, which currently accounts for more than a third of all ad expenditures and nearly 20% of average growth (Interactive Advertising Bureau and PricewaterhouseCoopers 2017). Online advertising has emerged as a critical marketing tool, and its effectiveness has been affirmed through consumers' click-through rates (Chatterjee, Hoffman, and Novak 2003; Hoban and Bucklin 2014; Schwartz, Bradlow, and Fader 2017), web browsing (Rutz and Bucklin 2012), online repurchase decisions (Manchanda et al. 2006), offline sales (Lewis and Reiley 2014), and long-term brand awareness (Drèze and Hussherr 2003).

Central to understanding advertising effectiveness is the concept of the response function—namely, the relationship between an outcome of interest (e.g., site visits, downstream sales) and advertising effort, assessed through various metrics (e.g., exposures, spend, ad stock). The broad consensus both in academia and managerial practice is that market response to advertising effort is nonlinear. That is, more intensive advertising leads to a stronger market response, but with diminishing marginal effectiveness—a concept sometimes generically referred to as "wearout." It has long been theorized that too many exposures beyond some threshold can even have a *negative* impact on market- or individual-level response (Calder and Sternthal 1980; Pechmann and Stewart 1988; Tellis 1988).

Inyoung Chae is Assistant Professor of Marketing, Goizueta Business School, Emory University (email: inyoung.chae@emory.edu). Hernán A. Bruno is Professor of Marketing and Digital Environment, University of Cologne (email: hernan.bruno@wiso.uni-koeln.de). Fred M. Feinberg is Handleman Professor of Marketing and Professor of Statistics, Ross School of Business, University of Michigan (email: feinf@umich.edu).

Online advertising environments encourage intense competition among advertisers for consumers' attention, which in turn results in repetitive exposures, ostensibly to break through the clutter (Yaveroglu and Donthu 2008). For instance, a typical internet user is exposed to more than 1,700 banner ads per month (Morrissey 2013). Compounding the problem is the nature of the medium: vivid ads are displayed in parallel to the website's main content (Drèze and Hussherr 2003) and seem to actively encourage divided attention (or even inattention) compared with the intrinsically sequential nature of radio or TV ads. Internet users have voiced concerns about unpleasant experiences from too many repeated exposures haunting their online paths. According to a survey by Upstream (2012), two-thirds of adults in the United States and United Kingdom reported experiencing digital advertising overload and negative reactions, such as "stop using the brand" (27% of U.S. adults) and "avoid the future message" (28% of U.S. adults).

In this article, we characterize the shape and dynamics of the online display advertising response function at the individual level. Specifically, we investigate the relationship between exposures to banner advertising from a financial service provider (hereinafter referred to as the "advertiser") and the visits to their website (i.e. the view-through). Compared with metrics such as site actions (e.g., requesting information) or downstream sales, view-through provides a more direct measure of campaign effectiveness, one less prone to contamination by other contextual variables. For instance, a consumer's decision to purchase a specific product over its rivals is typically affected by factors such as local availability, the point-of-purchase marketing mix, alternative options, purchase timing, and so on that are not necessarily related to the advertisement. How likely users are to make a site visit provides a rough approximation of their conversion value to the firm (Dalessandro et al. 2012) and has consequently served as a dependent measure (Braun and Moe 2013; Manchanda et al. 2006); moreover, it is less volatile than incremental sales or conversions, which often present a lack of statistical power (Lewis and Rao 2015). Park and Park (2016) showed, for example, that consumers' online store visits tend to vary in intensity, with a larger number of visits predicting higher conversion probability.[1]

We develop a model that relates a cumulative measure of ad exposure and the spacing between exposures to subsequent website visits to the advertiser, thereby allowing ads to be deployed more effectively. The model is flexible enough to accommodate a wide range of qualitatively distinct contours, particularly with respect to advertising repetition. Specifically, the model must be capable of capturing not only gains and regions of indifference, but also negative consequences as exposures accumulate, if such effects indeed exist. For this reason, the model will allow for a variety of individual-level response shapes but will not mandate them; it does not impose linearity, or even monotonicity, much less an eventually downward-sloping ad response function.

Moreover, it takes detailed account of the empirical nature of observable daily response, which is frequently zero visits, and—conditional on visiting on a given day—the modal number of visits is one. We use a mixture model that accounts for the excessive number of non- and single daily visits, as well as accommodating users who visit frequently. Importantly, the model also allows natural classes of response shapes to emerge from a general account of parametric heterogeneity. If distinct (parametric) user segments exist, the model can detect them alongside other qualitative differences in response. Because we do not impose any a priori functional differences across these (latent) segments, any manifestation of different response shapes across the classes is not guaranteed unless the data themselves support them. Individuals' membership in these segments is further linked to their observable characteristics, which, in our implementation, includes several distinct aspects of their browsing patterns.

The proposed model accounts for another key feature of display advertising: the differential effectiveness of publishers (i.e., the sites posting ads for the focal campaign) to drive visits to the advertiser's website. With the large and growing number of online publishers (e.g., 13 million alone in Google AdSense [2018]) and the variety of contracts available (e.g., real-time ad exchange, reservation-based ad contract), the need for advertisers to distinguish the relative efficacy of alternative publishers is ever more important. Although several studies have attempted to address this issue analytically (Balseiro et al. 2014; Berman 2017), evidence remains equivocal. Our empirical model measures the differential effectiveness of various publishers among the users who elect to visit their sites. Note that this differential effectiveness is that of *publishers*, not different media such as email and paid search advertising, as has been addressed in other marketing studies (e.g., Danaher and Dagger 2013; Kireyev, Pauwels, and Gupta 2013; Li and Kannan 2014; Shao and Li 2011).

We apply our model to a data panel containing individuals' advertising exposures (timing and publisher information) and their subsequent visits to the target site. Our results show clear evidence for the existence of a subset of users who are past the point at which an additional ad serves to increase their number of visits to the advertiser's website and, in fact, reduces it. We use the term "weariness" to refer to situations in which a user's advertising response curve slope becomes negative beyond some specific degree of cumulative advertising impressions. Investigating weariness is particularly critical in online advertising environments, because their low cost, ease of creation, and seemingly limitless channels can lead to excessive exposures to some users. This central finding of weariness critically relies on allowing for response heterogeneity. With a

[1] Park and Park (2016) also found a temporal variation period of $\sim$.81 days, suggesting that modeling *daily* number of visits (to capture conversion value) is particularly appropriate. When consumers are actively deliberating on (financial service) information, multiple site visits are common, so daily visit volume is a proper metric to capture ad performance. We note as well that view-through is distinct from click-through, which records actions to specific ads, and is well-suited to capture longer-term ad exposure effects.

homogeneous response function (as a benchmark), we observe a globally increasing and concave response function, consistent with other empirical findings in the literature.

We also observe each user's browsing activity (e.g., online usage frequency, breadth of searching, browsing preferences of specific topics), which is used to profile the revealed ad response classes. Such individualized data allow for not only a finer-grained post hoc portrait of reaction to advertising within the range it is actually practiced but also an assessment of individual user value before any ads are served in a focal campaign. Coupled with publisher effectiveness measures from the model, such information helps advertisers assess when specific users may be approaching the point at which additional advertising is overtly detrimental and thereby set appropriately individualized advertising strategies.

The remainder of the article is organized as follows. In the next section, we briefly discuss findings for advertising response functions in traditional and online media. We then elaborate the proposed model, estimation procedures, and a description of the empirical setting. In the remaining sections, we discuss inference and present estimation results, followed by substantive conclusions and discussion.

Advertising Response Functions

The history of advertising research is rich, from lab studies to empirical modeling, to identify functional relationships—that is, response functions—to advertising over intensity (i.e., repetition) or time (Calder and Sternthal 1980; Craig, Sternthal, and Clark 1976; Little 1979). Empirically estimated response function shapes have suggested that additional ad exposures lead to stronger response but that the marginal effect of each exposure can differ over the observed range of ad intensity.

To address this nonlinearity, advertising researchers introduced, and empirically validated, both the wear-in and wearout of advertising (Corkindale and Newall 1978; Naik, Mantrala, and Sawyer 1998; Pechmann and Stewart 1988). "Wear-in" refers to increasing the marginal effectiveness of advertising (on the outcome of interest, usually sales) with successive exposures, ordinarily detected at the outset of a given advertising campaign (Rao and Miller 1975; Terui and Ban 2008; Vakratsas et al. 2004). By contrast, "wearout" is the leveling out over ad repetitions in the so-called long-run (i.e., worn-in) level of sales. However, this definition includes several possible patterns, among them two that are of primary interest in the present study: (1) diminishing, but always positive, marginal response and (2) eventual negative marginal effects.

Complicating matters is that the empirical literature on advertising effectiveness has adopted a variety of functional definitions for wearout. For example, studies using a state-space model for advertising stock ("AdStock") define wearout as diminishing returns on otherwise positive effects (Bruce, Foutz, and Kolsarici 2012; Naik, Mantrala, and Sawyer 1998; Nerlove and Arrow 1962), whereas other definitions allow advertising effectiveness to hit a peak and thereafter display negative marginal effects (Chatterjee, Hoffman, and Novak 2003; Tellis 1988). A key question, therefore, is whether advertising can ever negatively affect performance (in sales, primarily, or in some other measure of interest).

Distinct from the general notion of wearout, we refer to this decreasing response—wherein the response function has negative slope beyond a threshold degree of ad exposure—as "weariness." Such concepts do have antecedents in prior marketing literature. For example, high levels of advertising leading to diminishing marginal response (i.e., wearout) have been referred to as "saturation" and, by analogy, high-enough levels to actually worsen response (i.e., weariness) as "supersaturation" (Doyle and Saunders 1990; Van Diepen, Donkers, and Franses 2009).[2]

Laboratory studies suggest that repeated ad exposures can have negative effects, such as decreased attention, comprehension, and, ultimately, adverse reaction, by inducing tedium (Calder and Sternthal 1980) and negative cognitive responses (Belch 1982). The ensuing behavioral literature has clarified mechanisms leading to diminishing or decreasing effects (e.g., Anand and Sternthal 1990; Batra and Ray 1986), shed light on different advertising or attitudinal metrics affected by repetition (e.g., Haugtvedt et al. 1994), and investigated moderating factors between repetitions and attitudinal response (e.g., Campbell and Keller 2003).

In contrast to these behavioral findings, empirical evidence for weariness in field data is remarkably scarce. For instance, Naik, Mantrala, and Sawyer (1998) tested for repetition wearout effects in two sets of market data and found an insignificant result in one setting and unimproved model performance in the other (a model with a wearout component did not perform better than a benchmark lacking it). Similarly, Bass et al. (2007) did not discover repetition wearout when an advertiser employed different themes in the campaign. More recently, in an e-commerce setting in which competitors coexist, Sahni (2016) demonstrated increasing ad effectiveness with mild nonlinearity across discretized exposure brackets (1–3, 4–7, 8–10, and >10), with equivocal results with regard to the existence of weariness (albeit not a focus of his study). Finally, Lewis (2017) investigated degree of wearout heterogeneity from 30 field experimental campaigns, during which online users were repeatedly exposed to advertisements, with their click-throughs and conversions traced. He categorized wearout effects for the 30 campaigns into four types (i.e., constant return to scale vs. {mild, moderate, extreme} wearout), but none of the 30 campaigns suggested evidence of negative consequences of high levels of ad repetition (i.e., weariness).

The lack of consistent evidence for the existence of weariness across laboratory and field settings might be attributed to

2. Despite some degree of conceptual overlap and mixed usage, saturation and wearout can be understood as the result of the former phenomenon affecting the latter; for example, media saturation over a short space of time can lead to ad wearout (Brown 2012). Because the present study does not address underlying mechanisms (e.g., supersaturation), we refer exclusively to the resultant concept, weariness.

three potential sources. First, it may reflect differing levels of analysis and/or circumstances (Craig, Sternthal, and Clark 1976; Pechmann and Stewart 1988): lab experiments study individual responses to displayed ads, whereas empirical (field) studies nearly always analyze market-level exposure data (for recent exceptions, see Sahni [2016] and Lewis [2017]). Second, there is weak evidence for weariness because the range in which advertising actually occurs is generally appropriate, neither too high nor too low. That is, observant advertisers might learn through trial and error to avoid costly ad repetitions at the point where overtly negative effects are apparent. Third, several previous models have, for reasons of parsimony and estimation efficiency, imposed nonnegativity constraints on additional ad exposure effects (e.g., those in the Nerlove and Arrow [1962] framework, where rate of change in advertising goodwill relates positively, with diminishing returns, to ad expense and/or number of repetitions; Bass et al. 2007; Braun and Moe 2013; Naik, Mantrala, and Sawyer 1998). Although such constraints can enhance model fit and empirical identification, they also smooth out (or rule out) the potential to detect weariness.

Our article contributes to this literature by providing evidence that there can be threshold levels of ad repetition past which marginal response is indistinguishable from zero and even demonstrably negative. Previous research has found an average ad frequency far greater, sometimes nearly 50 times so, for online advertising compared with traditional channels.[3] The uptick in ad repetition provided by the online platform provides researchers, for the first time, a firm basis on which to empirically examine the existence and extent of weariness.

The proposed model affords flexibility along several dimensions, sidestepping functional restrictions found in prior literature that may mask the potential manifestation of negative consequences of advertising repetition. Allowing for heterogeneity in response function parameters is critical to detect segments that evidence weary response functions. The latent class formulation effectively clusters those users with similar response patterns, discriminating those whose response is fairly linear from those exhibiting pronounced nonlinearity, including users with significantly nonmonotonic (i.e., weary) relationships to advertising exposure.

As alluded to previously, we show that observable, individual-level behavioral characteristics (i.e., browsing behaviors such as online frequency and breadth of browsing) help profile consumers into the identified segments. Such browsing behavior variables are readily collectable in online marketing systems through server log data stored in the ad network (as in this study). Scraping users' online cookies is an alternative method of collection, though this presents challenges to current data collection systems owing to the natural churn or removal of cookies from user devices after a specified period. Identifying individuals' membership within segments that vary in response (and curvature) across advertising control variables (e.g., repetition, timing) is critical for advertisers to plan user- or segment-specific advertising schedules.

Model Development

We model the number of visits y_{it} made by individual i on day t to an advertiser's website[4] by consumers who have been exposed to banner ads placed on various publishers' sites ($j = 1, \ldots, J$). Each user provides insufficient data to calibrate an individual-level ad response function, but these can be grouped (into latent classes) on the basis of similarity of response; critically, users' browsing behaviors can differentially "profile" them into the derived latent class groups.

The model is motivated both by the nature of the outcome variable and by several empirical patterns apparent in our data (as well as in prior studies of online advertising; Manchanda et al. 2006; Rutz and Bucklin 2012). Prominent among these patterns is one in Figure 1, Panel A: a high proportion—96% of the total—of "no visits" ($y_{it} = 0$) to the target website, as well as a large mass of *single* daily visits; although observations with $y_{it} = 1$ comprise only 1.6% of all observations, they account for more than 44% of all nonzero counts. A second important characteristic of the distribution of daily visits is its heavy right tail: conditional on visiting more than once, there is a small but important mass of observations with many visits.

This empirical distribution reveals three distinct categories of daily visit counts (i.e., 0, 1, or "more than 1"). We can conceptualize these categories as having different roles in the process of site visitation. That is, there are both quantitative and substantive differences between zero, one, and more visits, related to successive decisions about "Should I visit at all?," "Should I visit again?," and "How many times?" Models that *only* account for excessive zeros would lead to a poor fit, given the empirical distribution of our response variable y_{it}: either the mixture distribution would focus on the single-visit occurrences, ignoring the heavy tail of larger numbers of visits, or it would accommodate the tail at the expense of underestimating the mass of observations at $y_{it} = 1$.

Response Model for Site Visit Counts

We therefore tailor our model to account for excessive zeros and ones by defining a discrete mixture,

$$\Pr(Y = y_{it}) = \begin{cases} \phi_{0it} \text{ if } y_{it} = 0 \\ \phi_{1it} \text{ if } y_{it} = 1 \\ \phi_{2it} \cdot p_2(y_{it}) \text{ if } y_{it} \geq 2 \end{cases}, \tag{1}$$

3. Online users have been documented as receiving 20 repeated brand advertisements over 32 weeks (Chatterjee, Hoffman, and Novak 2003), 5 over 10 weeks (Braun and Moe 2013), and 6 over 10 weeks (in the present study). By contrast, pre-internet advertising entailed radically smaller rates, with 1.3 repeated brand ads observed over 84 weeks (Deighton, Henderson, and Neslin 1994) and 6 over 96 weeks (Siddarth and Chattopadhyay 1998).

4. A "visit" consists of a series of page views, each separated by less than ten minutes, from the same source.

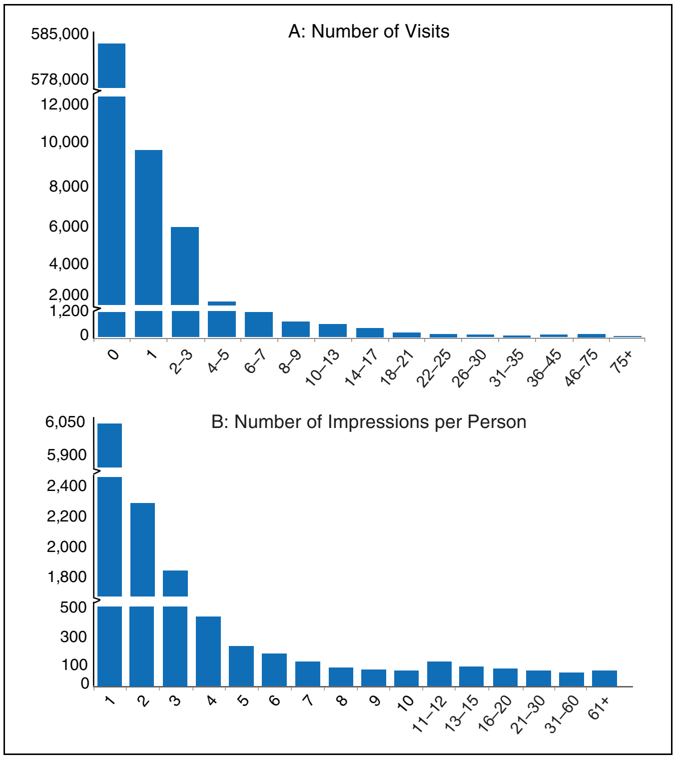

Figure 1. Histograms of number of visits and number of impressions per person.

where $\phi_{0\,it}$ and $\phi_{1\,it}$ capture the probability mass at zero and one, respectively. For visit counts larger than one, we allow for a discrete distribution, $\phi_{2it} \times p_2(y_{it})$, with support over $y_{it} \geq 2$; probability masses must sum to one: $\sum_{q=0}^{2} \phi_{qit} = 1$ and $\sum_{y=2}^{\infty} p_2(y) = 1$.

The second important characteristic of the distribution of daily visits is its heavy right tail. Simple maximum likelihood tests on our data set[5] reveal that the log-normal performs far better than the normal, the Poisson, and the negative binomial distribution densities to capture the "greater than one" visit distribution. This is unsurprising regarding the Poisson, which as a single-parameter distribution can be expected to perform poorly with overdispersed data. The superior performance of the log-normal over the normal distribution in particular can be explained by a better fit for skewed distributions (Limpert, Stahel, and Abbt 2001). To accommodate heavy tails, the discrete probability mass function $p_2(y_{it})$ is based on a discretization of the log-normal distribution, defined for integers larger than one (i.e., shifted by two) as follows:

$$p_2(Y = y_{it}|\mu_{it},\ \tau_i) = \begin{cases} \Phi\left(\frac{\ln(y_{it}-1.5)-\mu_{it}}{\tau_i}\right) & y_{it}=2 \\ \Phi\left(\frac{\ln(y_{it}-1.5)-\mu_{it}}{\tau_i}\right) - \Phi\left(\frac{\ln(y_{it}-2.5)-\mu_{it}}{\tau_i}\right) & y_{it}>2 \end{cases}, \tag{2}$$

where $\Phi(.)$ denotes the standard normal cumulative density function. Note that this is not an approximation, but a legitimate probability mass function with prior applications to count data (Chakraborty 2015).

To measure the effects of advertising on a consumer's view-through behavior, we model the probability of a nonvisit ($\phi_{0\,it}$), a single visit ($\phi_{1\,it}$), multiple visits ($\phi_{2\,it}$), and the mean parameter (μ_{it}) of the count values of further visits as functions of individual-specific advertising exposure covariates (A_{it}, z_{it}, and ν_{it}), individual-specific baselines (i.e., random intercepts u_i), and idiosyncratic errors (i.e., extreme value and normal distributions, respectively), unrelated to advertising effects. We model the zero-one-multiple visit masses as a conditional multinomial-logit,

$$\phi_{qit} = \frac{\exp(V_{qit} + u_{iq})}{\sum_{r=0}^{2} \exp(V_{rit} + u_{ir})} \quad \text{for } q \in \{0, 1, 2\}, \tag{3}$$

and the mean parameter μ_{it} of the discretized log-normal $p_2(y;\ \mu_{it},\ \tau_i)$ as

$$\log(\mu_{it}) = V_{3it} + u_{i3}, \tag{4}$$

where, in all cases, V denotes a deterministic function of covariates and u is a consumer-specific random intercept term. For identification purposes, the value of the "no visit" option ($q = 0$) utility is taken to be zero: $V_{0\,it} + u_{i0} = 0$. We use the same functional form for all three V_{qit}, $q \in \{1,\ 2,\ 3\}$,

$$V_{qit} = \alpha_{q1} + \alpha_{q2}A_{it} + \alpha_{q3}A_{it}^2 + \alpha_{q4}z_{it} + \alpha_{q5}z_{it}^2 + \alpha_{q6}\,\nu_{it} + \alpha_{q7}\,\nu_{it}A_{it} + \alpha_{q8}\,\nu_{it}z_{it}. \tag{5}$$

Here, A_{it} is i's advertising stock ("AdStock," as defined subsequently) at time t, z_{it} is elapsed time since a user's last ad exposure (Calder and Sternthal 1980), and ν_{it} is an indicator for whether the user has visited the site within the seven-day period before time t. This previous visit variable is included to control for unobserved, individual-level factors that potentially influence both having already visited and visiting in the current period. Such inclusion of lagged dependent variables can help mitigate potential autocorrelation (Germann, Ebbes, and Grewal 2015) and provide more robust estimates (Keele and Kelly 2005). To allow for potential nonlinearity in the marginal response, we include quadratic terms for both AdStock and time from the last exposure, as well as interaction terms with the within-a-week visit variable, ν_{it}.[6] To enable the model to be built up in stages, we

[5] Specifically, the maximized "zero-one-more" mixture model log-likelihoods are -168,450.7 for the Poisson (3 d.f.), -142,443.5 for the normal (4 d.f.), and -139,462.1 for the log-normal (4 d.f.). These analyses do not include covariates and heterogeneity.

[6] We found the one-week "Did they visit at all?" window to be empirically superior to other common measures that summarize prior visit history, such as

mean-center all variables, so that adding or removing one will have a far smaller effect. This also helps interpretations of higher-order terms and interaction effects (Irwin and McClelland 2001), and, in our application, with estimation efficiency.

The individual-specific random intercept terms u_i are included to capture visit tendencies driven by individual factors other than advertising exposure (e.g., word of mouth). We therefore include random intercept terms for both the multinomial ($\{u_{i1}, u_{i2}\}$) and the discretized log-normal (u_{i3}) components and allow them to be correlated across the specifications: $u_i \sim N_3(0, \Sigma)$ for $u_i = [u_{i1}\ u_{i2}\ u_{i3}]$. Note that these are individual-specific deviations from the respective overall intercept terms in each specification, such that the joint mean is identically zero.

AdStock Across Publishers with Differential Impact

A key component of our formulation is an individual- and time-specific measure of cumulated advertising exposure, or AdStock, that considers both the carryover property of classic AdStock dynamics (Naik, Mantrala, and Sawyer 1998) and the differential effects of each publisher. We thereby specify AdStock to allow for three key properties: (1) geometric weighting over time periods, (2) weighted aggregation across publishers, and (3) diminishing marginal returns to advertising exposures. First, we adopt the Koyck specification (Bass and Clarke 1972) for creating a geometrically smoothed index of each publisher's exposures:

$$x_{ijt} = (1 - \delta)\iota_{ijt} + \delta \times x_{ijt-1},$$

where ι_{ijt} is subject i's number of exposures of from publisher j, in day t, and $\delta \in [0, 1]$ is a carryover parameter. Note that, if advertising is held constant at $\iota_{ijt} = \bar{\iota}_{ij}$, in the long run $x_{ijt} \rightarrow \bar{\iota}_{ij}$.

One convenient, yet seldom empirically justifiable, assumption is that a given ad's impact is largely independent of the website on which it was shown. However, some sites have much higher ad density than others, place ads in less prominent positions, or otherwise differentially compromise the potential for the ad to spur the target behavior. Therefore, we construct the "smoothed advertising index" (SAI), a summary measure of Koyck indices weighted by each publisher's estimated relative effectiveness parameter, λ_j:

$$\mathrm{SAI}_{it} = \sum_{j=1}^{J} \lambda_j x_{ijt}. \tag{6}$$

As we detail subsequently, we operationalize $\{\lambda_j\}$ to nest the usual homogeneous publisher effects model ($\lambda_j = 1$) via a gamma density with mean 1. This avoids inflating the *overall* impact of advertising and offers the benefit of measuring its dispersion via the gamma variance. That is, if publishers' messages are roughly equally effective, $\{\lambda_j\}$ will have low variance.

Advertising has been applied in explicitly logarithmic form in a wide variety of empirical settings (Bass 1969; Rao 1972; Steenkamp et al. 2005; Van Heerde, Gijsbrechts, and Pauwels 2015, p. 683), capturing diminishing marginal returns, and we adopt this as follows:

$$\log[\mathrm{SAI}_{it} + 1] = \log\left[\left(\sum_{j=1}^{J} \lambda_j x_{ijt}\right) + 1\right].$$

Note that this measure relies primarily on three former treatments, namely Doyle and Saunders (1990), Manchanda et al. (2006), and Danaher and Dagger (2013), each of which helps incorporate advertising efforts spread across venues, consumers, and time in a manner exhibiting diminishing marginal returns to additional impressions. Manchanda et al. (2006) constructed an individual-level summary measure of the form log(1 + A), with contemporaneous, not cumulative, effects of the number of ad impressions; their study does not, however, differentiate ad effects across different media or publishers. By contrast, both Doyle and Saunders (1990) and Danaher and Dagger (2013) aggregate the effects of ad expenses across different media types—the former using aggregate data to smooth over time periods and campaigns, and the latter using "before–after" data to cumulate across a variety of media at the individual level. Although both articles use a weighted sum formulation, their data consists of different messages/channels, allowing them to study diminishing advertising effectiveness in each message/channel separately. In our study, the same message is delivered across publishers, whose effects are aggregately cumulated into the AdStock. This distinction is manifested in the modeling assumptions: while the models from Doyle and Saunders and Danaher and Dagger enacted the logarithmic transformation for each message type separately, after which they were linearly weighted, we impose the log-transformation after aggregating each publisher's effect and its corresponding AdStock x_{ijt}.

Finally, we subtract off the same "mean" (the one corresponding to $\lambda_j = 1$) to obtain what we refer to subsequently as "AdStock," a measure that nests the classical, mean-centered, variant:

$$\mathrm{AdStock}_{it} = \log\left[\left(\sum_{j=1}^{J} \lambda_j x_{ijt}\right) + 1\right] - \bar{x}, \tag{7}$$

where $\bar{x} = \dfrac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \log\left[\left(\sum_{j=1}^{J} x_{ijt}\right) + 1\right]}{\sum_{i=1}^{N} T_i}$ is the mean of the number of exposures for the homogeneous model (i.e., equal publisher weights, $\{\lambda_j = 1\}$).

Heterogeneous Users

In addition to heterogeneous individual baseline intercepts u in the three V portions of the model in Equations 3 and 4, it is critical

the number of times the user had visited in a week, whether (s)he had ever visited before, or several variants thereof.

to include heterogeneous sensitivity to advertising itself across users. To capture a variety of advertising response profiles, we introduce a latent categorical variable C_i that takes the value k ($k = 1, \ldots, K$) if user i belongs to class k. Thus, the full model can be specified via the following conditional likelihood for observing a specific number of site visits, y_{it}, by user i on day t:

$$f(y_{it} | C_i = k, \phi_{1itk}, \phi_{2itk}, \mu_{itk}, \tau_k) = (1 - \phi_{1itk} - \phi_{2itk})^{I(y_{it}=0)} \; \phi_{1itk}^{I(y_{it}=1)} \; [\phi_{2itk} p_2(y_{it} | \mu_{itk}, \tau_k^2)]^{I(y_{it}>1)}, \quad (8)$$

where the probabilities are defined as in Equations 2–5, with ϕ_{qitk} and μ_{itk} reparameterized as functions of latent class parameters α_{qk}. The variance parameters τ_k^2 and Σ_k are also subscripted to reflect their being class-specific.

The model is completed with a specification of the categorical class indicator, C_i, which takes value k with probability m_{ik}, and is in turn related to an individual-level covariate vector w_i through a multinomial logit model:

$$C_i \sim \mathrm{Cat}(m_{i1}, \ldots, m_{iK}) \quad (9)$$

$$m_{ik} = \frac{\exp(w_i \gamma_k)}{\sum_{j=1}^{K} \exp(w_i \gamma_j)}.$$

To enable model identification, the coefficients γ_1 for class 1 are set to zero. In simple terms, this portion of the overall model takes the form of a concomitant latent class model (Dayton and Macready 1988; Kamakura, Wedel, and Agrawal 1994).

The individual-level covariates w_i are used to probabilistically parcel users into K classes, for which user-specific, time-invariant variables (e.g., demographics) are commonly used. However, in the online arena, acquiring user-specific demographic information is uncommon outside limited situations, such as where user registration is required (e.g., online shopping applications); thus, the class probability or mixture proportion m_{ik} is often assumed to be constant across users, as opposed to a hierarchical structure in which different users have distinct assignment probability vectors (Rutz and Bucklin 2012). By way of contrast, we make use of individual users' behavioral metrics (i.e. online browsing patterns) to inform their class membership. Such metrics—including internet usage frequency, breadth of distinct websites visited, and visiting trends that capture degrees of interest in specified topics—can be assessed by advertisers through user cookies. The online behavioral variables (w_i) were collected over a longer time span and averaged across the observation period; thus, they can be regarded as largely static individual-level traits, as opposed to short-term behaviors that can be confounded with ad exposure. Note that given the hierarchical nature of our model, these variables influence only class membership, not the visit behavior. To further assess the viability of treating these variables accordingly, we calculated the variational overlap between an individual's behavioral traits and visits, finding these to be very low, ranging from 1.7% to 4.4% variance explained.

This latent class structure and the user-specific random intercept term discussed previously provide a flexible heterogeneity specification that can account for a wide range of responses. We stress again that the latent classification is not based on response curvature with respect to AdStock (i.e., differences in weariness), but on a general account of parametric heterogeneity for a variety of covariates (i.e., intercept, AdStock, Timing, Visit, and interaction terms), as well as hierarchical class inclusion probabilities based on browsing behavior metrics.

The model is estimated using a fully Bayesian approach.[7] To ensure a well-identified model with posteriors determined almost entirely by the data, we assign weakly informative diffuse distributions to all class-specific parameters. Prior hyperparameters are set identically across classes. For the coefficient parameters ($\alpha_{1k}, \alpha_{2k}, \gamma$), we adopt normal priors, while for variance parameters, we adopt log-normal and inverse-Wishart priors. Finally, the prior distribution of parameters for the differential publisher effects is taken to be gamma with, as described previously, a mean of 1 and a variance of θ, that is, $\mathrm{Gamma}(\text{shape} = \theta^{-1}, \ \text{scale} = \theta^{-1})$, so that larger values of θ indicate greater variation in publisher effectiveness. For θ, we use a noninformative hyperparameter prior, $\theta \sim \mathrm{Unif}(0, \ \infty)$. For completeness, the full model, including latent class indicators, and the prior specification appear in the Web Appendix W-A.

Under prior independence, the corresponding joint posterior is given by

$$\pi(\{\alpha_k\}, \{\Sigma_k\}, \{u_i\}, \{\lambda_j\}, C, \{\gamma_k\} | y) \propto \prod_{i=1}^{N} \left\{ \sum_{k=1}^{K} f(m_{ik} | \gamma_k, w_i) \left[\prod_{t=1}^{T_i} f(y_{it} | \alpha_k, \lambda_j, u_i, \tau_k^2) f(u_i | \Sigma_k) \right] \times \pi(\alpha_k) \pi(\tau_k^2) \pi(\Sigma_k) \right\} \times \prod_{k=2}^{K} \pi(\gamma_k) \times \prod_{j=1}^{J} \pi(\lambda_k | \theta) \times \pi(\theta). \quad (10)$$

For posterior estimation, we use Metropolis–Hastings steps with a treatment for missing online behavioral metrics for nonvisitors to the advertiser's website (Web Appendix W-B). Web Appendix W-C specifies the conditional densities and algorithms in detail. Burn-in was set at 15,000 draws, with an additional 15,000 used for parameter estimation. The convergence of Markov chain Monte Carlo chains is assessed by

7. Various restricted versions of the model—for example, without individual-specific intercepts or heterogeneous publisher effects—may be much more quickly estimated through an expectation–maximization approach (Dempster, Laird, and Rubin 1977). However, our analyses depend on calculations across the entire posterior, not only its mode, necessitating Bayesian methods.

stable trace and density plots, as well as effective sample sizes (all reported parameters have >100 effective samples) and Gelman–Rubin (1992) diagnostics. All parameters of substantive importance (i.e., excepting the thousands of user-level intercepts) show strong evidence of convergence. To assure model validity and estimation accuracy, we simulate a data set using the same covariates used in the actual estimation. Results show that the model reliably recovers "true" parameter values; details appear in Web Appendix Table W-1.

Table 1. Data Set Summary Statistics.

Total number of observations	601,481
Total number of users	12,000
Average number of days observed (SD)	50.1 (13.4)
Average number of visits per user (SD)	7.7 (17.6)
Average number of visits per observation (SD)	.15 (1.7)
Total number of publishers	473
Impressions accounted for by top 27 publishers	83%
Average number of days since the last exposure (SD)	20.2 (15.8)

Data

Our main data provider for this research was numberly 1000mercis Group ("1000mercis" hereinafter), a large advertising agency that purchases online ad space on behalf of its clients. The data set pertains to advertising campaigns, focused on desktop and laptop users, for a French financial firm that offers credit services and loans. The firm runs two distinct types of ad campaigns: acquisition, wherein the stated goal is to entice viewers to visit its website (which is the firm's primary focus), and retargeting (e.g., in the terminology of Lambrecht and Tucker [2013], "generic retargeting"), which encourages viewers to complete desired actions (e.g., contacting sales staff). Consumers subject to retargeting campaigns were systematically selected before the campaign launched, separately from the acquisition campaigns. As an independent check, we confirmed this claim, finding no overlap across campaigns (acquisition and retargeting). Consequently, our data consist solely of an acquisition campaign, one intended to reach a broad population. Strict privacy laws in France prevented user-specific information (i.e., demographics) from being used; the ad agency has further confirmed that its technology did not accommodate user-specific trait targeting for mass online acquisition campaigns. Specifically, the ad allocation mechanism for the target campaign is based on publishers, not users, ruling out a potential source of endogeneity (i.e., in which more responsive users are targeted for advertising). Moreover, ad exposures are randomly served across both users and time within a publisher's website. Such random frequency greatly mitigates confounding selection bias between users' browsing behaviors and ad responses, as in Lewis (2017). In our "Empirical Results" section, we report direct checks that strongly support 1000mercis's claims regarding such nonendogenous targeting.

The banner ad campaign ran for a total of 72 days, from May 20, 2013, to July 31, 2013, urging consumers to visit the advertiser's website, where they could assess their own credit and loan potential. The banners featured generic information (e.g., basic credit rates and repayment information, along with imagery depicting occasions on which people typically seek short-term financial assistance) and did not contain overtly affective elements. The advertised credit service is often purchased repeatedly, as needed, similar to cash-out services offered by credit card firms.

Our data set thereby comprises three distinct individual-level data sources:

- Daily advertising exposures captured from 473 publishers,
- Individual visitor logs collected from the advertiser's website, and
- Users' browsing behaviors captured from an additional 200 websites that belong to 1000mercis's ad network.

We selected users from among those who were exposed to the ad agency's top publishers and retrieved their ad exposure information from other publishers, visitor behaviors to the advertiser website, and browsing behaviors from separate data sources. Because this specific campaign did not use a capping strategy (i.e., limiting the number of ads delivered to each user through a certain publisher or across publishers), we excluded users who had an anomalously high number of ad exposures (>99.99th percentile, or 920 in total) or site visits (>99.99th percentile, or 119 overall), noting that doing so can only bias *against* discovering weariness and never create it as an artifact. (Subsequently, we perform robustness checks using alternative exclusion rules.) Then, we randomly sampled 12,000 online users, who closely matched the population in terms of distribution of ad exposures and visits. Table 1 shows summary information for the main variables.

In our data set, the 12,000 users are exposed to advertisements at least once, for a total of 28,345 impressions and total 92,586 visits to the advertiser's website during the campaign. These exposures are placed across 473 distinct websites, ranging from news portals and social networking venues to dedicated corporate sites, and from local French websites to well-known global sites.[8] The distribution of number of exposures across publishers is right-skewed, with over 83% of impressions being placed by the top 27 publishers. To efficiently estimate the site-specific differential effectiveness indices, we treat these 27 major publishers as separate entities and the rest as a distinct "all other publishers" category. We calculate the variable "time elapsed since last exposure" without regard to publisher placement. This variable serves to help capture "copy wearout," a potential decline in ad effectiveness over time (Naik, Mantrala, and Sawyer 1998).

Users' characteristics, as described previously, consist of various browsing behaviors. To capture these characteristics,

8. Specific sites referred to in our analysis are proprietary and cannot be publicly disclosed.

Table 2. Summary of Browsing Behavior.

numDays: Number of days browsing the websites (SD)	2.7 (3.0)
numSites: Number of websites visited (SD)	1.8 (1.2)
e-Commerce: Percentage e-commerce	8.0% (22.5%)
Media: Percentage media	6.1% (20.0%)
Finance: Percentage financial services	8.5% (7.4%)

we consulted an additional independent data set provided by 1000mercis that contained the same users' browsing histories (across 200 websites in 1000mercis' networks). We defined users' online usage frequencies as the number of days ("numDays") they browsed in the entire set of networks and defined browsing breadth as the number of distinct websites ("numSites") they visited per day. In addition, using category information for the website networks, we constructed covariates that are known to be either topically or behaviorally related to the campaign and associated target website. Specifically, the website networks cover a wide range of categories (about 20 in all), such as e-commerce (38%), tourism (15%), and finance (12%), among others. Among these categories, we considered whether users' browsing interests coincided with the industry to which the advertiser belonged (e.g., finance), purchase intention (e.g., e-commerce), or general information gathering (e.g., media). These covariates used for profiling are not time varying, rather capturing their general online usage traits. Table 2 lists summary statistics of the target users' browsing behaviors.

Empirical Results

Endogeneity Checks

The central inferential results from our model assume that exposures are pseudo-randomized in a specific way: a user's prior behavior does not influence the probability of receiving an exposure (i.e., there is no targeting based on past behavior). As mentioned previously, the focal ad campaign does not avail of any user-specific data to deliver ads; the ad-serving algorithm is limited to publisher-level information, such as number of users and cost per 1,000 impressions. Before presenting model estimation results, we provide a series of data-based checks against common endogeneity confounds.

The key question can be put simply: do users receive more (or fewer) ad exposures after visiting the focal website? Although this may seem simple, note that it can come about in many ways that do *not* connote users being differentially targeted on the basis of their actions, such as exposures increasing overall over time or varying day-of-week usage patterns. To check that the tendency to receive an ad exposure does not hinge on the user's focal behavior (visiting the advertiser's website), we first conducted simple paired t-tests for the number of impressions before and after a user's visit (Table W-2). We perform this analysis for the entire window of observation (total number of impressions before and after the first visit, Column 1); one week before and after the visit, to control different time length effects (Column 2); and six, seven, or eight days before and after the visit, to control both recency and day-of-week effects (Column 3). The results connote the opposite of what a targeting mechanism would (e.g., that users should receive more impressions *after* a visit) and consistently indicate nothing more than a *general* downward exposure trend across the data window.

It is therefore necessary to consider when the first visit occurs as well. We stratified the sample to control for users' tendency to receive certain numbers of impressions at a certain time (i.e., the time when the first impression occurs and the number of impressions received that week). As an illustration, we first isolate users who received one impression during week 2 (labeled "week 2: 1 impression" in Table W-3); we then compare those who did versus those who did not visit the website during the following week and then compute the number of impressions they received the *following* week. If there were some specific targeting, we would expect to see that those who visit in the following week (week 3 in our example) would receive more impressions on average. We conducted this analysis for all combinations of (noninitial and nonending) weeks {2,3,4,5,6} and number of impressions suitable for a statistically *disconfirming* finding, {1,2,3,4,5} (i.e., there were too few observations for week 6 and up). In these $5 \times 5 = 25$ tests conducted simultaneously, two were significant at the .05 level and none at .01. To pool statistical power, we further conduct five tests combining data across blocks, finding none significant even at .05. In summary, we find no evidence that visiting the site yields differences in post visit exposure activity.

Model Selection

Selected values of several fit metrics are examined for carryover parameter δ from 10% to 90% in 10% increments; relevant latent class specifications (homogeneous and the best-fitting models for each carryover parameter), both with and without differential publisher effects; the average diffuseness parameter ($E(\theta)$; higher values indicate greater heterogeneity in publisher impact); and proportions of users falling into each latent class (Table W-4). We include four principal types of fit metric: log-marginal densities (LMD), deviance information criterion (DIC), $\text{elpd}_{\text{WAIC}}$, and $\text{elpd}_{\text{IS-LOO}}$. The first, LMD, is a standard Bayesian model fit metric, but it can be sensitive to prior specifications and posterior sample size. Like LMD, the DIC (Spiegelhalter et al. 2002) provides a likelihood-based measure, but unlike LMD, it penalizes model complexity (i.e., overparameterization). The last two types, expected log predictive density (elpd) measure out-of-sample prediction accuracy from a fitted Bayesian model by approximating the expected log pointwise predictive density: $\text{elpd}_{\text{WAIC}}$ is based on the Watanabe–Akaike information criterion, or "widely applicable information criterion" (Watanabe 2010), while $\text{elpd}_{\text{IS-LOO}}$ makes use of importance sampling weighted leave-one-out cross-validation. Gelman, Hwang, and Vehtari (2014) and Vehtari and Gelman (2014) provide detailed definitions, calculation methods, and differential benefits of each criterion for Bayesian model selection.

Table 3. Class-Specific Parameters.

	K = 1			K = 2			K = 3			K = 4			K = 5		
Parameters	**Mdn**	**2.5%**	**97.5%**	**Mdn**	**2.5%**	**97.5%**	**Mdn**	**2.5%**	**97.5%**	**Mdn**	**2.5%**	**97.5%**	**Mdn**	**2.5%**	**97.5%**
α_{1k} _(Intercept)	**−6.73**	**−6.96**	**−6.51**	**−7.30**	**−9.41**	**−5.99**	**−4.09**	**−4.18**	**−4.00**	**−4.47**	**−4.57**	**−4.37**	**−3.32**	**−3.42**	**−3.21**
_AdStock	**4.83**	**2.53**	**7.22**	**6.03**	**2.86**	**9.69**	**5.68**	**4.43**	**7.01**	**4.58**	**3.70**	**5.69**	**2.47**	**1.22**	**3.76**
_AdStock^2	**−2.73**	**−5.84**	**−.25**	**−4.84**	**−7.01**	**−2.75**	**−3.01**	**−4.37**	**−1.91**	**−2.81**	**−4.22**	**−1.82**	**−1.05**	**−2.14**	**−.11**
_Time	**−.19**	**−.37**	**−.01**	−.21	−1.59	1.46	**−.21**	**−.33**	**−.10**	−.07	−.15	.01	**.25**	**.14**	**.36**
_Time^2	**−.31**	**−.42**	**−.20**	**−.38**	**−.45**	**−.31**	**−.33**	**−.39**	**−.25**	**−.20**	**−.24**	**−.15**	**−.09**	**−.16**	**−.02**
_Visit	**1.36**	**1.06**	**1.60**	**−13.84**	**−26.92**	**−5.75**	**−.41**	**−.60**	**−.24**	**.89**	**.80**	**.99**	**1.29**	**1.18**	**1.40**
_AdStock × Visit	.77	−1.11	2.62	−2.85	−20.06	17.69	**1.66**	**.32**	**3.07**	.45	−.19	1.19	−.52	−1.46	.46
_Time × Visit	**.47**	**.14**	**.79**	1.11	−7.39	11.31	**.61**	**.25**	**.92**	**.13**	**.01**	**.25**	−.16	−.32	.00
α_{2k} _(Intercept)	**−5.53**	**−5.65**	**−5.43**	**−7.15**	**−9.12**	**−6.05**	**−22.79**	**−32.91**	**−14.64**	**−4.21**	**−4.32**	**−4.11**	**−2.84**	**−2.92**	**−2.76**
_AdStock	**6.73**	**5.14**	**8.36**	**7.87**	**4.70**	**11.09**	−3.18	−20.91	15.78	**4.53**	**3.42**	**5.81**	**1.88**	**1.10**	**2.76**
_AdStock^2	**−4.99**	**−7.16**	**−3.14**	**−6.36**	**−8.53**	**−4.45**	−1.82	−20.87	12.40	**−2.23**	**−3.75**	**−1.20**	−.41	−1.18	.21
_Time	**−.27**	**−.38**	**−.15**	−.03	−1.10	1.51	2.07	−7.16	10.86	**−.21**	**−.27**	**−.15**	.04	−.05	.12
_Time^2	**−.34**	**−.41**	**−.28**	**−.45**	**−.52**	**−.39**	−3.37	−11.20	3.25	**−.25**	**−.30**	**−.19**	**−.14**	**−.19**	**−.09**
_Visit	**1.06**	**.89**	**1.23**	**−12.35**	**−24.63**	**−5.53**	−3.59	−19.66	8.66	**1.00**	**.91**	**1.08**	**1.54**	**1.44**	**1.66**
_AdStock × Visit	−.50	−2.15	1.12	−2.99	−19.93	14.29	−2.27	−18.79	16.13	.06	−.61	.66	−.78	−1.51	.02
_Time × Visit	**.47**	**.23**	**.70**	3.05	−3.40	12.46	−.99	−17.03	12.73	.10	−.04	.21	**−.29**	**−.42**	**−.17**
α_{3k} _(Intercept)	.01	−.15	.17	**.21**	**.11**	**.34**	.00	−.19	.17	**.22**	**.15**	**.30**	**.19**	**.11**	**.26**
_AdStock	.00	−.14	.16	.06	−.11	.24	.01	−.16	.20	.05	−.10	.18	.05	−.15	.24
_AdStock^2	−.01	−.17	.12	−.01	−.20	.19	.00	−.15	.19	.03	−.14	.22	.00	−.18	.23
_Time	.00	−.17	.18	−.02	−.11	.05	−.02	−.23	.16	−.03	−.08	.04	.00	−.09	.07
_Time^2	.01	−.18	.19	.02	−.02	.07	.00	−.18	.17	.03	−.01	.08	.02	−.03	.07
_Visit	−.01	−.15	.14	−.02	−.18	.20	.01	−.20	.19	.01	−.09	.11	−.06	−.17	.02
_AdStock × Visit	.02	−.15	.20	.00	−.18	.20	.00	−.21	.19	.02	−.16	.22	.02	−.15	.18
_Time × Visit	.00	−.16	.17	.01	−.16	.22	.00	−.22	.17	.05	−.04	.13	**−.10**	**−.19**	**.00**

Notes: Values in boldface indicate significant coefficient estimates based on a 95% confidence interval.

Estimation of different latent class specifications (K $=$ $1, \ldots, 6$) shows that five latent classes consistently perform best for the range of carryover parameters (δ) above 50%, although the number of classes for the best fitting models fluctuates when the parameter values are low (e.g., three for 10%, two for 20%, and four for 30%–40%). The DIC and the two elpd-based criteria consistently indicate that the model with $\delta = 60\%$ carryover, five latent classes, and differential (i.e., heterogeneous) publisher effects fits best. It is immediately apparent that both sorts of heterogeneity—parametric (latent classes) and publishers—are valuable. Furthermore, class allocation and class-specific parameters are relatively invariant when publisher heterogeneity is included (vs. not); analogously, the diffuseness parameter θ is largely stable across different numbers of latent classes. Unless otherwise stated, the remaining discussion of the results focuses on the best-fitting model, namely K $= 5$ and $\delta = 60\%$. Table W-5 shows that the degree of publisher heterogeneity is stable across K $=$ $1, \ldots, 6$ latent classes (i.e., $E(\theta) =$ .23, .29, .27, .24, .25, .24) as well as listing class sizes.

AdStock, Timing, and Recent Website Visits

Model parameter estimates are listed in Table 3. For most of the parameters comprising α_{1k} and a_{2k}, posterior mass is bounded away from zero.[9] By contrast, for the parameters in α_{3k}, posterior marginal distributions provide little evidence that any specific parameter differs from zero. However, this ignores two points: (1) for the discretized log-normal portion in particular (for the "two or more" portion of the model), intercepts are significant in three of the classes (2, 4, and 5), and these capture the baseline tendency to revisit the advertiser's website, and (2) in a Bayesian model, "significance" needs to be assessed across the joint posterior distribution of all parameters compared with a base model. To this latter observation, we reestimate the "best" model setting $\alpha_{3k} = 0$, except for intercepts, finding it to fit far worse overall, across all fit metrics (decreases: LMD 9%, DIC 10%, WAIC: 9%, IS-LOO 7%). Taken as a whole, parameter estimates indicate that, for this data set, advertising not only determines the relative proportions of zero versus one versus multiple visits but also selectively influences the actual number of multiple visits for a particular group.

The proposed model, though parsimonious in terms of capturing several forms of heterogeneity, has several parameters large enough that viewing their summary measures together (Table 3) does not provide an immediate substantive interpretation. Instead, we project the parameters onto a managerially relevant metric: expected number of visits to the advertiser's

9. It may appear that the correlations between the vectors for α_{1k} and a_{2k} are high within each latent class. This is an artifact of the choice of Visits $= 0$ as the base category; these correlations fail to evince any clear pattern when the model is reestimated, for example, using Visits $= 1$ as a base category.

Table 4. Lifts in Expected Number of Visits (No Change = 1).

	K = 1	K = 2	K = 3	K = 4	K = 5
AdStock	1.57	1.93	1.62	1.52	1.11
Time	.94	.91	.90	.84	.82

Notes: Lifts are calculated for a one-standard-deviation change in AdStock and Time.

site, over a range of focal covariates (e.g., AdStock, time elapsed since last exposure), computed through:

$$E(\hat{y}_k) = \frac{1}{N_k} \sum_{i \in \{i: C_i = k\}} \frac{1}{D} \left[\sum_{d=1}^{D} \hat{\phi}_{1itkd} + \hat{\phi}_{2itkd} \sum_{y_{it}=2}^{\infty} y_{it} p_2 \left(y_{it} | \mu_{itkd}, \tau_{kd}^2 \right) \right], \tag{11}$$

where the simulated values for $\hat{y}_k, \hat{\phi}_{itkd}, \ p_2(y_{it}|\mu_{itkd}, \tau_{kd}^2)$ are as given in Equation 8, using estimated parameters $\{\hat{\alpha}_{kd}, \hat{\mu}_{kd}, \hat{\tau}_{kd}, \hat{u}_{id}\}$, and $d = 1, \ldots, D$ indexes Markov chain Monte Carlo draws.

Managerially relevant summaries of focal input variables—AdStock and Time—can be obtained by calculating the “lift” of the $E(\hat{y}_k)$ summary measure (i.e., the number of expected visits) under one-standard-deviation increases in each input relative to the baseline values based on the actual observation. The calculated lifts (Table 4) are analogous to model coefficient estimates but instead show by what factor the number of expected visits change, providing a scale-free effects measure (against a baseline of 1). These lift factors vary substantially across latent classes, illustrating the degree of heterogeneity in response. Notably, expected visits increase by 11%–93% across classes as AdStock increases by one standard deviation, with the largest bump for the weary class (Class 2). By contrast, all classes show approximately 10% drops as Time (since last exposure) increases by one standard deviation. Lift results are consistent with coefficient estimates (Table 3) but integrate them into single metrics, thereby illustrating strong between-class covariate differences.

Although the lift measures ably summarize marginal covariate impacts, they are only static summary measures. To illustrate both AdStock and dynamic effects (time, prior visits), Figure 2 breaks out expected number of visits as a function of both SAI (SAI_{it} in Equation 6) and Time (since last exposure, in days), for each of the five classes, over its relevant range (95% of the x-axis covariate). Although this produces different abscissa and ordinate scalings for each class (i.e., each subgraph), it mitigates outlier effects and ensures that each is focused on covariate effects within the data range practiced for each of the five classes.

Recall that the firm did not use an individual-level capping strategy. Figure 2 was constructed using all users in the data set and individual posterior draws with the key input variables varying over the 95% observed range. In each subgraph, the dark gray ribbon shows the 50% pointwise highest density region (25%–75%) and the light gray ribbon shows the 95% highest density region (2.5%–97.5%) of the estimated expected visits. This magnitude varies greatly by class, with Class 1 having an expected number of visits about two orders of magnitude smaller than that of Classes 2, 4, and 5. Summary statistics regarding by-class response shapes appear in Table 5.

The patterns of advertising response over SAI show marked differences across the latent classes. Class 2, which contains 24% of users in the sample, shows a pronounced wearout pattern, with clear concavity in the response function, as well as evidence of weariness. To assess the strength of this evidence for weariness, we compute advertising response contours using each user’s posterior draws and check whether each individual contour has an interior maximum. In Class 2, 100% of contours have internal maxima within the (95th-percentile) observed range of SAI. In other words, we can conclude that the (median) response function estimate has a negative slope beyond a particular point well within the range of SAI observed in the data. For Class 2, specifically, weariness sets in at a cumulated ad exposure level corresponding to being exposed to three impressions within two days. (We note that these findings are robust to several alternate measures of cumulated ad spending [e.g., a fixed window and/or alternative values of the smoothing constant]; detailed results are available from the authors.)

Although it cannot be called conclusive, Class 3 also demonstrates substantial degrees of weariness, with 82% of contours having internal maxima (most of the users within the class show weariness). This class, however, is smaller (7.8% of the sample), with fairly low baseline visit propensity (expected number of visits = .13). Despite 72% of contours showing weariness, Class 1 is both small (2.4%) and nearly nonresponsive, with .004 expected visits. The remaining classes (Classes 4 and 5) display strongly monotonic (quasilinear), increasing response functions (solid lines in Figure 2). Thus, in the end, only Class 2 can be claimed to be both responsive to advertising and weary.

The response shapes over “time since last visit” (right side of Figure 2) consistently show concavity with interior maxima. The timings for these maxima vary across classes: earliest in Class 5 (four days) and latest in Class 1 (ten days). Although this result may speak to the potential for superior ad spacing, we note that doing so requires analyzing both AdStock and Timing in a full-scale dynamic optimization framework under budget constraints, as in the line of research on pulsing (Mahajan and Muller 1986).

In summary, the model clearly distinguishes users’ (latent) heterogeneous responses with respect to AdStock and Time (since last exposure). We stress again that the model was not purposely built to distinguish response *shapes*—the latent classes could differ strongly in effects parameters, but none could have exhibited weariness—yet weariness is strongly apparent in one sizeable class (Class 2). As such, despite visual similarities in the contours across classes (e.g., Classes 4 and 5 in Figure 2), the scales of effects (and, as we explore subsequently, individual class probability effects) differ markedly across them.

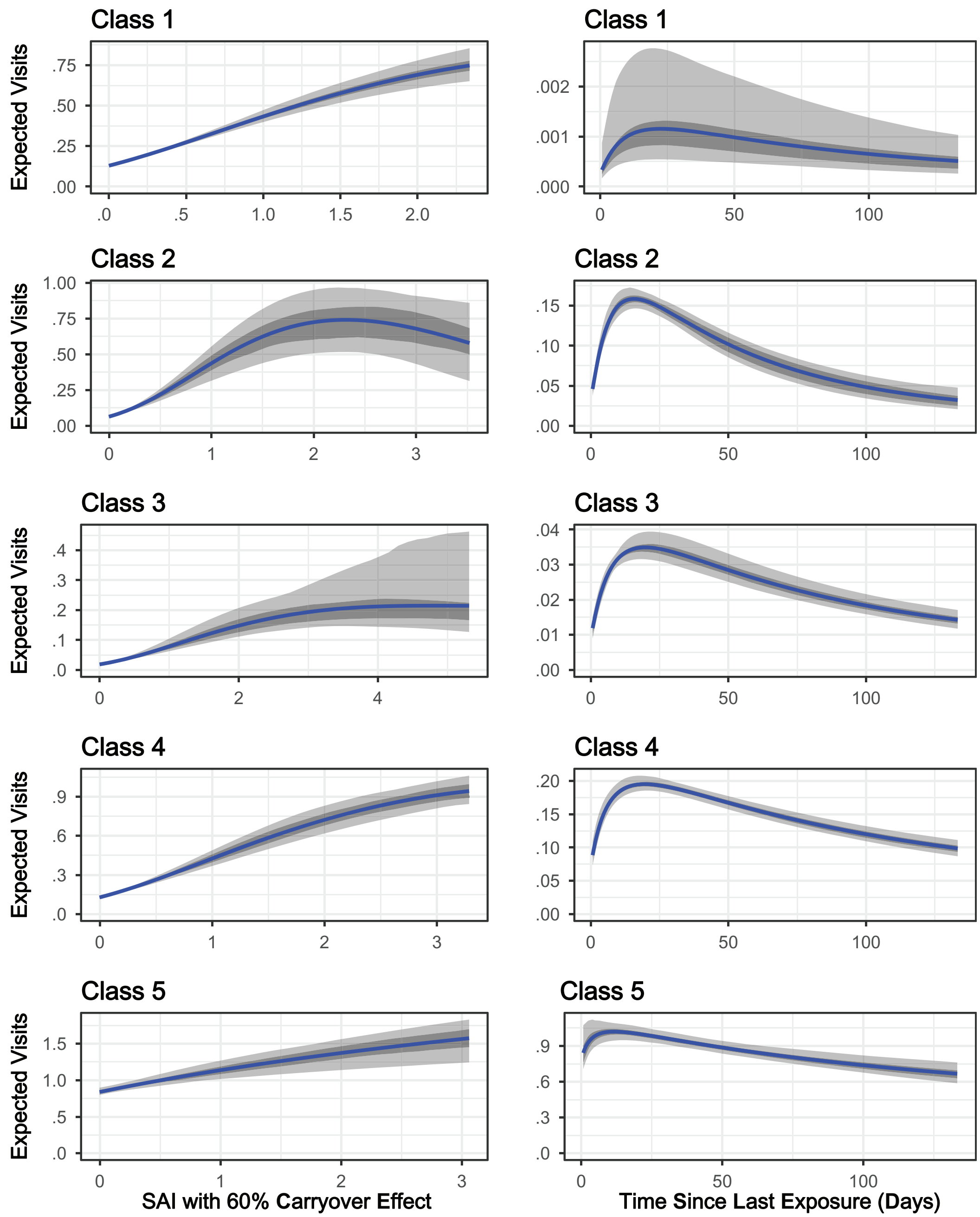

Figure 2. Ad response contours.

Notes: The horizontal and vertical axes are selected for each subgraph to best showcase its contours within the relevant range of ad exposures for the depicted class.

Publishers' Effectiveness

The estimates of $\{\lambda_j\}$, which represent the differential weights of publishers within the AdStock specification, show a substantial degree of variation. We define a metric $H_k(j)$ that captures how much a given publisher j delivers exposures to a particular class k, relative to its entire population baseline exposure. Formally, if $h_k(j)$ is total number of ad exposures to class k by publisher j, we define

$$H_k(j) = \frac{h_k(j)/\sum_j h_k(j)}{\sum_k h_k(j)/\sum_k \sum_j h_k(j)}.$$

In simple terms, $H_k(j) > 1$ means that the publisher is showing more ads to Class k than the average across the user sample.

Table 6 summarizes results for publisher effectiveness $\{\lambda_j\}$, including median and 95th quantile estimates, effectiveness rank, size (proportion of impressions allocated to that

Table 5. Summaries of Response Shapes.

	Class 1	Class 2	Class 3	Class 4	Class 5
Class size (proportion)	2.35%	24.37%	7.80%	57.78%	7.71%
Average exposures per user	2.45	2.60	3.22	2.55	3.78
Average # of expected visits	.004	.48	.13	.53	1.20
Peak # of expected visits	.005	.77	.21	.93	1.61
SAI_{it} at peak	3.29	2.15	4.43	3.29	3.06
Coefficient of variation	1.96	2.05	1.89	2.00	5.43

publisher), and $H_k(j)$. Recall that $\{\lambda_j\}$ are drawn from a gamma density with mean 1 and hierarchical variance parameter $\hat{\theta}$, with an estimated value of .25; this reflects the fact that some publishers' ads seem to have substantially greater impact than others, e.g., $\hat{\lambda}_{19} = 2.16$ (the most effective) is more than nine times that for the 26th, $\hat{\lambda}_{26} = .23$. Note that the publishers are indexed by size (i.e., $j = 1$ has the largest volume of traffic). As one might therefore expect, the first three publishers account for the bulk (in fact, more than 50%) of impressions allocated in the sample. However, estimated $\{\lambda_j\}$ corresponds only roughly to publisher size: the three publishers with highest estimated ad impact ($j = 13, 17, 19$) deliver only 6% of impressions collectively. Combining these patterns in $\{\lambda_j\}$ and $H_k(j)$ with the class-wise response shapes (Figure 2) is broadly suggestive regarding publisher allocation.[10] For example, advertisers may find it worthwhile to increase exposure levels for Publisher 5 within Class 5 because it appears low (93% vs. baseline), given the publisher's high effectiveness ($\hat{\lambda}_5 = 1.47$) and the class's lack of weariness; the same could be said of Class 3 for Publisher 19 (94%; $\hat{\lambda}_{19} = 2.16$). By contrast, advertisers might want to hold off on further exposures for Class 1 (the least effective class) using Publishers 22 and 26, because these publishers already deliver high exposure levels (164% and 171%, respectively) and are not particularly effective ($\hat{\lambda}_{22} = .74$ and $\hat{\lambda}_{26} = .23$).

Class Membership Influence

Thus far, we have found that there are five groups with distinctive ad reaction patterns; some evidence wearout (Classes 1 and 3), one (Class 2) shows clear signs of responding negatively after a certain number of repetitions, whereas the others have positive marginal response to ads across their observed domain. These findings are informative to advertisers overall but are especially valuable when they can apportion users into these response groupings. Advertisers need to capitalize on user-specific information to enact distinct ad strategies without compromising user privacy. As we mentioned in the "Data" section, we used a separate data set precollected by the agency before the ad campaign consisting of browsing behaviors. These variables can be used to compute users' membership indicators (C_i) and thereby enable us to investigate the relationship between membership and browsing behavior, as summarized in Table 7.

An intriguing finding is the relationship between browsing frequency ($\gamma_k^{\text{numDays}}$) and class membership probability (p_k). Notably, users with lower rates of browsing frequency are more likely to show weariness (Classes 1 and 2; e.g., $\gamma_2^{\text{numDays}} = -8.04$), whereas those with greater frequency tend toward quasilinear response shapes—that is, to be nonweary (Classes 4 and 5; e.g., $\gamma_5^{\text{numDays}} = 6.20$). This suggests that if light internet users (i.e., people with low browsing frequency) are repeatedly exposed to ads, they are more likely to tire of and, in turn, respond negatively to those ads. We speculate (i.e., in the absence of attitudinal data) that this phenomenon may arise because light users, who are less familiar with the online environment, are put off when they encounter a large number of ads; by contrast, heavy users are awash in ads and do not find them off-putting when they encounter many extra ones, though confirmation of such a process-based explanation awaits additional dedicated data.

Users in Class 1 visit a large number of websites (numSites), while those in weary Class 2, in addition to having low browsing frequency, visit a small number of sites and show relatively little interest in topics such as e-commerce, media, and finance (strongly negative coefficients for each). Class 3 users fall in between in terms of browsing behavior (numDays) but visit fewer sites, such as those with specific topical and/or behavioral relation to the focal site (finance, e-commerce, or media). Users in Class 4 are frequent and broad internet browsers, but like those in Class 2, they show little interest in topics relevant to advertisers. Finally, the heaviest internet users whose interests align with the advertiser are likely to belong to Class 5. Such links between class membership and behavioral (or other) covariates can serve a function similar to discriminant analysis: overlaying additional data valuable for targeting, thereby parceling incoming (or other) users into previously identified classes and thus helping project their likely ad response pattern.

Discussion and Simulation of Alternative Allocation Policies

Does "Weariness" Exist?

The model presented accounts for three sources of heterogeneity: normally distributed individual-level baselines, gamma-distributed publisher-specific effects, and a discrete (latent class) account of advertising response. Consider a model where the first two sources of heterogeneity were taken into account, but advertising response parameters were presumed identical (homogeneous) for all customers. Such a model would *not*

[10] We note here that a rigorous assessment of whether a particular publisher was truly "overadvertising" would require, at the very least, impression costs for each site, as well as some account of customer lifetime value. Although publishers and advertisers do have such data, their data were proprietary.

Table 6. Differential Publisher Effectiveness.

j	Mdn	2.5%	97.5%	Rank	Size	$H_k(j)$, % Change Above Baseline Exposure				
						Class 1	Class 2	Class 3	Class 4	Class 5
1	.44	.36	.58	26	14%	109%	98%	99%	101%	97%
2	.86	.71	1.00	15	16%	93%	103%	104%	99%	99%
3	1.48	1.21	1.80	4	5%	110%	99%	106%	99%	101%
4	.75	.65	.91	20	8%	95%	100%	96%	101%	97%
5	1.47	1.22	1.70	5	3%	95%	105%	100%	99%	93%
6	1.06	.92	1.27	10	4%	83%	103%	92%	99%	110%
7	.79	.60	.95	19	3%	74%	103%	103%	100%	91%
8	1.26	1.09	1.45	7	3%	89%	100%	101%	101%	97%
9	1.22	.97	1.55	8	2%	113%	103%	89%	101%	91%
10	.99	.87	1.25	11	3%	140%	106%	97%	98%	89%
11	.73	.61	1.03	22	2%	94%	93%	100%	103%	104%
12	.93	.75	1.23	12	1%	122%	104%	101%	97%	101%
13	1.73	1.58	2.28	2	2%	107%	96%	104%	101%	101%
14	.47	.33	.68	25	2%	76%	92%	107%	104%	93%
15	.92	.74	1.36	14	1%	90%	93%	117%	98%	122%
16	1.12	.82	1.38	9	2%	92%	103%	107%	97%	106%
17	1.72	1.26	2.11	3	2%	82%	98%	104%	102%	92%
18	1.37	.81	1.65	6	2%	118%	103%	102%	95%	117%
19	2.16	1.82	2.57	1	1%	68%	101%	94%	96%	141%
20	.27	.19	.36	27	1%	105%	80%	92%	106%	122%
21	.55	.37	.71	24	1%	106%	82%	104%	98%	167%
22	.74	.43	.94	21	1%	164%	98%	103%	97%	108%
23	.81	.61	.99	18	1%	96%	93%	117%	100%	108%
24	.93	.66	1.25	12	1%	140%	102%	85%	102%	84%
25	.85	.56	1.17	17	1%	131%	83%	63%	113%	80%
26	.23	.19	.46	28	1%	171%	95%	77%	107%	66%
27	.59	.43	.73	23	1%	88%	93%	70%	107%	101%
28	.86	.70	.99	15	17%					

Table 7. Class Membership Parameters.

	Class 1	Class 2	Class 3	Class 4	Class 5
(Intercept)	−22.30** (−570.24)	−14.74** (−252.61)	−20.61** (−609.23)	−14.21** (−153.87)	−19.43** (−413.09)
numDays	−.41** (−9.66)	−8.04** (−125.97)	.45** (12.04)	4.30** (42.53)	6.20** (120.40)
numSites	.18** (4.59)	−.13* (−2.30)	−.25** (−7.45)	.16* (1.74)	−.06 (−1.24)
e−Commerce	.51** (12.45)	−4.33** (−71.16)	4.49** (127.11)	−1.11** (−11.54)	.08 (1.72)
Media	.60** (14.14)	−4.01** (−63.46)	3.99** (108.83)	−1.07** (−10.65)	.16** (3.14)
Finance	.24** (6.80)	−1.35** (−25.18)	1.23** (39.53)	−.31** (−3.62)	.04 (.91)

Notes: t-statistics are in parentheses.

show that all viewers were identical, because their baselines to visit the advertiser (intercepts) would still differ, and of course publishers would still have differential effectiveness. The patterns of advertising response over SAI (Equation 6) of such a model appear in Figure 3, and they are in agreement with many previous analyses: response shapes exhibit diminishing marginal returns but still increase notably without an eventual downturn; indeed, *none* of the individual response curves exhibits weariness. In other words, advertising shows the classic contours of wearout, but weariness is nowhere to be found. However, when advertising response heterogeneity is taken into account, distinct reactions over ad repetitions are evident. The majority of users still exhibit diminishing returns, but the *degree* of wearout varies greatly, both in terms of curvature and in the level of exposure when it starts to set in: although Classes 4 and 5 appear to have quasilinearly increasing responses, Classes 1 and 3 exhibit wearout with some degree of potential weariness (but low responsiveness overall), whereas Class 2 shows decisive tendencies toward a weary response pattern. In other words, and in what we believe to be an important novel finding of our analysis: assuming advertising response homogeneity mistakenly implies a lack of weariness, essentially averaging it out of view.

To bolster these findings regarding weariness, we conducted checks to determine whether the existence of weary groups hinges on particular model specifications or sample selection

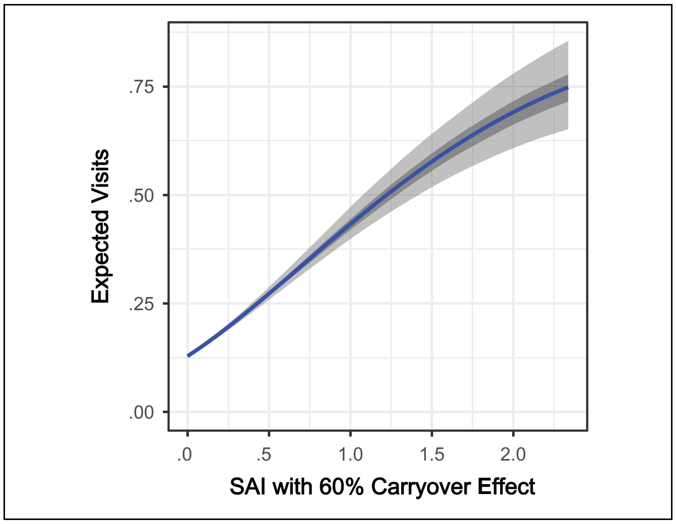

Figure 3. Response shape with single class model.

criteria. First, we substantiate the log-normal portion of the count model. As a benchmark model, we also examine the binary model that distinguishes visits as zero versus one or more, consistently uncovering weary classes; however, neither rises to the 95% contour concavity standard apparent for the full model. We note in passing that it is precisely this information—the number of visits—that is differentially important in determining eventual conversion, as Moe and Fader (2004) point out. Second, we find a distinct weary class even when imposing equal publisher effectiveness ($\lambda_j = 1$). Third, we reestimate the model using the same AdStock specification (Equation 7) but with SAI in place of log(1 + SAI); this results in five classes, one of which displays strong weariness, but a dramatically poorer overall fit (roughly 15% in each of the four fit metrics). Fourth, we excluded outliers using different percentiles (i.e., .1%, 1%, and 5%, vs. the .01% used previously); in each case, there was at least one class with size greater than 30% and over 95% weary contours. The last set of checks concerns the potential for so-called "activity bias" (Lewis and Reiley 2014), whereby very active internet users are more likely to both receive a high number exposures and respond differently to advertising (compared with lighter users). This concern is addressed in two distinct ways: (1) we constructed a daily browsing activity variable (i.e., the number of daily webpages visited), using it as a control covariate, and (2) we constructed two additional sample data sets of 12,000 users based on overall browsing activities, specifically, the second and fourth quintiles of users' total number of webpages visited). As before, the models run in all these cases perfectly replicate our substantive results: a single class with significant, substantial weariness. In short, weariness was robust to all critical sample selection criteria and model constructs, save one: response homogeneity, which we turn to next. (All models pertaining to the robustness checks and corresponding findings are detailed in Web Appendix W-D.)

Failing to Consider Response Shape May Contribute to Wasted Ad Exposures

Considering individual users' responses when weariness sets in may lead to a more nuanced understanding of ad placement performance. This is quantified in the upper part of Table 8, which lists the expected number of visits for each class based on actual ad placement, an improved ad placement across publishers (as described next), and two common benchmark ad-capping strategies: capping by publisher and capping across publishers. The first two rows show the predicted number of visits in-class, whose values very closely match the class means of actual visits (82%–95%), with ads placed as in the data. The next two rows capture ad performance when varying advertisements in terms of the number of repetitions and publishers' effectiveness. Assuming, for the sake of analysis, that users continue visiting publishers in roughly the same proportion, we can reapportion ads over different publishers to improve effectiveness.[11] In this case, more than 1.3% (last column of Table 8) of ads served to the consumer sample appear to be wasted, and even *hurt* the advertiser's performance, driven by the presence of weary users (who are primarily in Class 2, where 3.2% of ads were apparently wasted). By contrast, none of the ads served to Class 5—where the response curve is nearly linearly increasing—were wasted. The improvement in expected number of visits (accomplished by not wasting exposures and improving publisher allocation) is remarkable: through reallocation, the advertiser can achieve a 15% performance improvement in total (and more than 34% improvement in the weary Class 2).

For comparison, we also consider two different types of benchmark strategies common in the advertising industry. The first and simplest of these is to cap the maximum number of ads that each publisher can serve to a user; such a strategy is commonly adopted when an advertiser contracts with a specific publisher to limit the number of ads to a unique user (indexed by user ID). In this case, between 2% and 7% of ads can be saved across five classes, amounting to 4% of total ads served to them (second-to-last row of Table 8). However, if the user visits multiple publishers that serve the ad, the number of ads (s)he is exposed to would exceed this optimum. This issue is readily addressed by the second strategy, enabled by the recent advent of online ad exchanges (in which all advertisers and publishers are connected through networks): capping the number of ads served across all publishers. Such a program shows that 3% to 9% of ads can be saved, or approximately 6% overall (last row of Table 8). As might be expected, our weary Class 2 evinces the highest potential to save ads, at 9.2%. These strategies account for consumer heterogeneity, leading to differential "weariness thresholds" across consumer groups and consequent cost-saving advantages; however, actual

[11] All analyses are based on an exhaustive sequential grid search for the best publisher combinations, carried out in R. Full details are available from the authors.

Table 8. Model Implications for Improved Ad Placement Regimen.

		Class 1	Class 2	Class 3	Class 4	Class 5	Total
Original	Number of ads used	692	7,608	3,017	17,660	3,492	32,469
	Expected number of visits	.001	.09	.03	.15	.88	.20
Enhanced	Number of ads used	692	7,367	2,978	17,530	3,492	32,059
	Expected number of visits	.001	.12	.03	.19	.93	.23
	Performance improvement	12.2%	34.1%	15.5%	21.6%	5.1%	15.1%
	Proportion of ads saved	.0%	3.2%	1.3%	.7%	.0%	1.3%
Benchmark 1	Cap number of ads for each publisher (% ads saved)	3.2%	6.8%	2.1%	3.9%	3.6%	4.4%
Benchmark 2	Cap number of ads across publishers (% ads saved)	4.4%	9.2%	2.7%	5.3%	5.0%	5.9%

performance in terms of increasing expected number of visits is hampered by their ignoring differential effectiveness across *publishers*. This admittedly simple analysis could be enhanced with detailed ad placement cost and customer (exposure and visit) data to help pinpoint exactly which consumers, or groups thereof, are prime candidates for ad wastage for a focal target campaign.

Conclusion

Marketing scholars have theorized for decades about whether a pronounced downturn in cumulative ad effectiveness (weariness) exists but have thus far lacked the detailed, individualized exposure data enabling its detection. In this article, we have examined the results of a comprehensive ad campaign with nearly one million exposures to over ten thousand internet users, finding that weariness manifests for roughly a quarter of the sample within the range typical exposure levels. It is important to note that weariness is not, either conceptually or empirically, a specifically temporal effect. That is, weariness does not mean that users are exposed to ads and react to them less and less *over time*. Weariness is about how two otherwise identical users can have different "stocks" of ad exposure, and the one with lower AdStock might show a superior ad response.

These findings regarding weariness relied on a sufficiently rich model to detect it, specifically one accounting for multiple forms of heterogeneity. However, the key finding about the very existence of a weary class was robust to a wide variety of modeling constructs (including number of classes, differential publisher effects, and temporal smoothing of the AdStock metric) and sample inclusion criteria. The one incontrovertibly critical component was the incorporation of (latent class) heterogeneity in response parameters across *users*. Models lacking this component suggested that all users' (individual posterior) response contours were in fact monotonically increasing in ad response.

Given that advertisers must choose exposure or expenditure levels and target consumers, it is vital to understand the relative effectiveness across users, within-user overrepetition and spacing of exposures, and the channels to reach those users, especially in terms of which regions might show evidence of expenditure waste. The present study takes a first step toward resolving these managerial issues by distinguishing several manifested response shapes that vary in overall size, baseline tendency to visit, and the response contours corresponding to ad repetition, timing, and previous site visits. This means that advertisers armed with appropriate data can profile individual consumers in terms of their response to ad repetitions and thereby get a relative handle on their "value" (e.g., potential gain, risk of overexposure, or inappropriate exposure spacing). We emphasize that this profiling can be based on members' observed browsing behaviors prestored in the ad networks—that is, *before* they are served any ads in a target campaign. Critically, this holds for both users in the estimation sample and new users for whom ads must be served prospectively. Because real-time ad allocation decisions are centrally coordinated by the ad network, the rich information available on individual consumers (e.g., past exposure histories and corresponding publishers) enables the estimation of weariness thresholds and appropriate temporal exposure spacing for target individuals whose baseline visit tendency is high. Moreover, publisher-specific effectiveness estimates (coupled with appropriate cost data) can help advertisers better set detailed expenditure levels. A simulation study, using insights gleaned from our model, demonstrated substantial performance improvements compared with conventional advertising strategy and current practice in the actual data set.

The framework proposed here entails several limitations, some of which owe to data availability and complex downstream real-time optimization. One open issue concerns how to leverage the model's estimates to optimize advertising allocation across time and sites, preferably in a real-time, decision-based system. This question is notoriously challenging, as it requires nontrivial cost information (which itself is often dynamically updated), as well as individual site-usage patterns, which then must be coupled with combinatorial optimization to solve the discrete advertising allocation scheduling problem. A second concern is the presumption that impression records accurately account for all exposures of a particular ad. Even though weariness has clearly been detected, one can never verify whether some ads that were served were not in fact viewed; furthermore, although we have ruled this out decisively for our data set, it is possible that ads served for retargeting would also be seen by users tracked for acquisition; this general issue is exacerbated by the need to combine exposures across various mobile and stationary devices. Third, our results concern only website visits. Future work could extend our examination of the existence of weariness to the upper level of the purchase funnel, up through eventual sales. While the present study investigates various modeling constructs to detect and confirm the existence of weariness, all findings are based on website visit behavior, which primarily reflects the awareness and interest levels of the purchase funnel. Although website visits arguably approximate final conversion values, one cannot rule out the possibility that the weariness may wane over the purchase funnel, a fertile topic for future research. Finally, understanding why certain publishers are more effective than others, and the potentially synergistic effects of combinations of publishers, are compelling research questions. Although the current study cannot address them because of data limitations, we view these extensions as both worthwhile and within the purview of advertisers armed with appropriate additional information and sufficient computational resources.

Associate Editor

Peter Danaher served as associate editor for this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Anand, Punam, and Brian Sternthal (1990), "Ease of Message Processing as a Moderator of Repetition Effects in Advertising," *Journal of Marketing Research*, 27 (3), 345–353.

Balseiro, Santiago R., Jon Feldman, Vahab Mirrokni, and S. Muthukrishnan (2014), "Yield Optimization of Display Advertising with Ad Exchange," *Management Science*, 60 (12), 2886–2907.

Bass, Frank M. (1969), "A Simultaneous Equation Regression Study of Advertising and Sales of Cigarettes," *Journal of Marketing Research*, 6 (3), 291–300.

Bass, Frank M., Norris Bruce, Sumit Majumdar, and B.P.S. Murthi (2007), "Wearout Effects of Different Advertising Themes: A Dynamic Bayesian Model of the Advertising-Sales Relationship," *Marketing Science*, 26 (2), 179–95.

Bass, Frank M., and Darral G. Clarke (1972), "Testing Distributed Lag Models of Advertising Effect," *Journal of Marketing Research*, 9 (3), 298–308.

Batra, Rajeev, and Michael L. Ray (1986), "Situational Effects of Advertising Repetition: The Moderating Influence of Motivation, Ability, and Opportunity to Respond," *Journal of Consumer Research*, 12 (4), 432–45.

Belch, George E. (1982), "The Effects of Television Commercial Repetition on Cognitive Response and Message Acceptance," *Journal of Consumer Research*, 9 (1), 56–65.

Berman, Ron (2018), "Beyond the Last Touch: Attribution in Online Advertising," *Marketing Science*, 37 (5), 771–92.

Braun, Michael, and Wendy W. Moe (2013), "Online Display Advertising: Modeling the Effects of Multiple Creatives and Individual Impression Histories," *Marketing Science*, 32 (5), 753–67.

Brown, Millward (2012), "Do TV Ads 'Wear Out'?" WPP, New York.

Bruce, Norris I., Natasha Z. Foutz, and Ceren Kolsarici (2012), "Dynamic Effectiveness of Advertising and Word of Mouth in Sequential Distribution of New Products," *Journal of Marketing Research*, 49 (4), 469–86.

Calder, Bobby J., and Brian Sternthal (1980), "Television Commercial Wearout: An Information Processing View," *Journal of Marketing Research*, 17 (2), 173–86.

Campbell, Margaret C., and Kevin L. Keller (2003), "Brand Familiarity and Advertising Repetition Effects," *Journal of Consumer Research*, 30 (2), 292–304.

Chakraborty, Subrata (2015), "Generating Discrete Analogues of Continuous Probability Distributions: A Survey of Methods and Constructions," *Journal of Statistical Distributions and Applications*, 2(1), 6, https://doi.org/10.1186/s40488-015-0028-6.

Chatterjee, Patrali, Donna L. Hoffman, and Thomas P. Novak (2003), "Modeling the Clickstream: Implications for Web-Based Advertising Efforts," *Marketing Science*, 22 (4), 520–41.

Corkindale, David, and John Newall (1978), "Advertising Thresholds and Wearout," *European Journal of Marketing*, 12 (5), 329–78.

Craig, C.S., Brian Sternthal, and Clark Leavitt (1976), "Advertising Wearout: An Experimental Analysis," *Journal of Marketing Research*, 13 (4), 365–72.

Dalessandro, Brian, Rod Hook, Claudia Perlich, and Foster Provost (2012), "Evaluating and Optimizing Online Advertising: Forget the Click, but There Are Good Proxies," *Big Data*, 3 (2), 90–102.

Danaher, Peter J., and Tracey S. Dagger (2013), "Comparing the Relative Effectiveness of Advertising Channels: A Case Study of a Multimedia Blitz Campaign," *Journal of Marketing Research*, 50 (4), 517–34.

Dayton, C. Mitchell, and George B. Macready (1988), "Concomitant-Variable Latent-Class Models," *Journal of the American Statistical Association*, 83 (401), 173–8.

Deighton, John, Caroline M. Henderson, and Scott A. Neslin (1994), "The Effects of Advertising on Brand Switching and Repeat Purchasing," *Journal of Marketing Research*, 31 (1), 28–43.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin (1977), "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society, Series B (Methodological)*, 39 (1), 1–38.

Doyle, Peter, and John Saunders (1990), "Multiproduct Advertising Budgeting," *Marketing Science*, 9 (2), 97–113.

Drèze, Xavier, and François-Xavier Hussherr (2003), "Internet Advertising: Is Anybody Watching?" *Journal of Interactive Marketing*, 17 (4), 8–23.

Gelman, Andrew, Jessica Hwang, and Aki Vehtari (2014), "Understanding Predictive Information Criteria for Bayesian Models," *Statistics and Computing*, 24 (6), 997–1016.

Gelman, Andrew, and Donald B. Rubin (1992), "Inference from Iterative Simulation using Multiple Sequences," *Statistical Science*, 7 (4), 457–72.

Germann, Frank, Peter Ebbes, and Rajdeep Grewal (2015), "The Chief Marketing Officer Matters!" *Journal of Marketing*, 79 (3), 1–22.

Google Adsense Usage Statistics (2018), (accessed July 16, 2018), calendar.google.com/calendar/render#main_7.

Haugtvedt, Curtis P., David W. Schumann, Wendy L. Schneier, and Wendy L. Warren (1994), "Advertising Repetition and Variation Strategies: Implications for Understanding Attitude Strength," *Journal of Consumer Research*, 21 (1), 176–89.

Hoban, Paul R., and Randolph E. Bucklin (2014), "Effects of Internet Display Advertising in the Purchase Funnel: Model-Based Insights from a Randomized Field Experiment," *Journal of Marketing Research*, 52 (3), 375–93.

Interactive Advertising Bureau and PricewaterhouseCoopers (2017), "IAB Internet Advertising Revenue Report 2016 Full Year Results," (April), https://www.slideshare.net/frenchweb/iab-internet-advertising-revenue-report-2016-75468917.

Irwin, Julie R., and Gary H. McClelland (2001), "Misleading Heuristics and Moderated Multiple Regression Models," *Journal of Marketing Research*, 38 (1), 100–109.

Kamakura, Wagner A., Michel Wedel, and Jagadish Agrawal (1994), "Concomitant Variable Latent Class Models for Conjoint Analysis," *International Journal of Research in Marketing*, 11 (5), 451–64.

Keele, Luke, and Nathan J. Kelly (2005), "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables," *Political Analysis*, 14 (2), 186–205.

Kireyev, Pavel, Koen Pauwels, and Sunil Gupta (2013), "Do Display Ads Influence Search? Attribution and Dynamics in Online Advertising," *International Journal of Research in Marketing*, 33 (3), 475–49.

Lambrecht, Anja, and Catherine Tucker (2013), "When Does Retargeting Work? Information Specificity in Online Advertising," *Journal of Marketing Research*, 50 (5), 561–76.

Lewis, Randall (2017), "Worn-Out or Just Getting Started? The Impact of Frequency in Online Display Advertising," working paper.

Lewis, Randall A., and Justin M. Rao (2015), "The Unfavorable Economics of Measuring the Returns to Advertising," *Quarterly Journal of Economics*, 130 (4), 1941–73.

Lewis, Randall A., and David H. Reiley (2014), "Online Ads and Offline Sales: Measuring the Effect of Retail Advertising Via a Controlled Experiment on Yahoo!" *Quantitative Marketing and Economics*, 12 (3), 235–66.

Li, Hongshuang (Alice), and P.K. Kannan (2014), "Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment," *Journal of Marketing Research*, 51 (1), 40–56.

Limpert, Eckhard, Werner A. Stahel, and Markus Abbt (2001), "Log-Normal Distributions across the Sciences: Keys and Clues on the Charms of Statistics," *Bioscience*, 51 (5), 341–52.

Little, John D. (1979), "Aggregate Advertising Models: The State of the Art," *Operations Research*, 27 (4), 629–67.

Mahajan, Vijay, and Eitan Muller (1986), "Advertising Pulsing Policies for Generating Awareness for New Products," *Marketing Science*, 5 (2), 89–106.

Manchanda, Puneet, Jean-Pierre Dubé, Khim Yong Goh, and Pradeep K. Chintagunta (2006), "The Effect of Banner Advertising on Internet Purchasing," *Journal of Marketing Research*, 43 (1), 98–108.

Moe, Wendy W., and Peter S. Fader (2004), "Dynamic Conversion Behavior at E-Commerce Sites," *Management Science*, 50 (3), 326–35.

Morrissey, Brian (2013), "15 Alarming Stats About Banner Ads," Digiday (March 21), https://digiday.com/media/15-alarming-stats-about-banner-ads.

Naik, Prasad A., Murali K. Mantrala, and Alan G. Sawyer (1998), "Planning Media Schedules in the Presence of Dynamic Advertising Quality," *Marketing Science*, 17 (3), 214–35.

Nerlove, Marc, and Kenneth J. Arrow (1962), "Optimal Advertising Policy Under Dynamic Conditions," *Economica*, 29 (114), 129–42.

Park, Chang H., and Young-Hoon Park (2016), "Investigating Purchase Conversion by Uncovering Online Visit Patterns," *Marketing Science*, 35 (6), 894–914.

Pechmann, Cornelia, and David W. Stewart (1988), "Advertising Repetition: A Critical Review of Wearin and Wearout," *Current Issues and Research in Advertising*, 11 (1/2), 285–329.

Rao, Ambar G., and Peter B. Miller (1975), "Advertising-Sales Response Functions," *Journal of Advertising Research*, 15, 7–15.

Rao, Vithala R. (1972), "Alternative Econometric Models of Sales-Advertising Relationships," *Journal of Marketing Research*, 9 (2), 177–81.

Rutz, Oliver J., and Randolph E. Bucklin (2012), "Does Banner Advertising Affect Browsing for Brands? Clickstream Choice Model Says Yes, for Some," *Quantitative Marketing and Economics*, 10 (2), 231–57.

Sahni, Navdeep S. (2016), "Advertising Spillovers: Evidence from Online Field Experiments and Implications for Returns on Advertising," *Journal of Marketing Research*, 53 (4), 459–78.

Schwartz, Eric M., Eric T. Bradlow, and Peter S. Fader (2017), "Customer Acquisition Via Display Advertising Using Multi-Armed Bandit Experiments," *Marketing Science*, 36 (4), 500–522.

Shao, Xuhui, and Lexin Li (2011), "Data-Driven Multi-Touch Attribution Models," in *proceedings for KDD '11*. New York: Association for Computing Machinery, 258–64.

Siddarth, Sivaramakrishnan, and Amitava Chattopadhyay (1998), "To Zap or Not to Zap: A Study of the Determinants of Channel Switching During Commercials," *Marketing Science*, 17 (2), 124–38.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64 (4), 583–639.

Steenkamp, Jan-Benedict E.M., Vincent R. Nijs, Dominique M. Hanssens, and Marnik G. Dekimpe (2005), "Competitive Reactions to Advertising and Promotion Attacks," *Marketing Science*, 24 (1), 35–54.

Tellis, Gerard J. (1988), "Advertising Exposure, Loyalty, and Brand Purchase: A Two-Stage Model of Choice," *Journal of Marketing Research*, 25 (2), 134–44.

Terui, Nobuhiko, and Masataka Ban (2008), "Modeling Heterogeneous Effective Advertising Stock Using Single-Source Data," *Quantitative Marketing and Economics*, 6, 415–38.

Upstream (2012), "Digital Advertising Attitude Report/The Consequences of Digital Ad Bombardment," (February), https://www.

upstreamsystems.com/2012-digital-advertising-attitude-report-the-consequences-of-digital-ad-bombardment.

Vakratsas, Demetrios, Fred M. Feinberg, Frank M. Bass, and Gurumurthy Kalyanaram (2004), "The Shape of Advertising Response Functions Revisited: A Model of Dynamic Probabilistic Thresholds," *Marketing Science*, 23 (1), 109–19.

Van Diepen, Merel, Bas Donkers, and Philip H. Franses (2009), "Does Irritation Induced by Charitable Direct Mailings Reduce Donations?" *International Journal of Research in Marketing*, 26, 180–88.

Van Heerde, Harald J., Els Gijsbrechts, and Koen Pauwels (2015), "Fanning the Flames? How Media Coverage of a Price War Affects Retailers, Consumers, and Investors," *Journal of Marketing Research*, 52 (5), 674–93.

Vehtari, Aki, and Andrew Gelman (2014), "WAIC and Cross-Validation in Stan," working paper (May 31), http://www.stat.columbia.edu/~gelman/research/unpublished/waic_stan.pdf.

Watanabe, Sumio (2010), "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory," *Journal of Machine Learning Research*, 11, 3571–94.

Yaveroglu, Idil, and Naveen Donthu (2008), "Advertising Repetition and Placement Issues in Online Environments," *Journal of Advertising*, 37 (2), 31–44.