OPERATIONS RESEARCH
Articles in Advance, pp. 1–13
ISSN 0030-364X (print) | ISSN 1526-5463 (online)
http://dx.doi.org/10.1287/opre.2014.1339

Understanding the Memory Effects in Pulsing Advertising

Ashwin Aravindakshan, Prasad A. Naik
Graduate School of Management, University of California, Davis, Davis, California 95616
{aaravind@ucdavis.edu, panaik@ucdavis.edu}

Extant models assume that awareness decline commences instantly. In contrast, we incorporate the possibility that awareness declines with a delay due to the memory for advertisements. To this end, we use delay differential equations to understand the evolution of awareness in the presence of ad memorability. This extended model generates optimal advertising policies that include the even spending policy, blitz policy, and various cyclic pulsing policies, depending on whether ad memorability exceeds a critical threshold. The extended model not only unifies the various patterns of advertising spending over time, but also augments the prior research by furnishing the optimality of pulsing advertising. Thus ad memorability could drive pulsing. We discuss the implications for practicing managers and identify avenues for future researchers.

Subject classifications: Marketing: advertising and media; optimal control; memory effects; awareness formation; pulsing advertising; Hamiltonian; Lambert W; delay differential equations.
Area of review: Marketing Science.
History: Received August 2013; revisions received April 2014, August 2014, October 2014; accepted October 2014. Published online in *Articles in Advance*.

1. Introduction

Apple's 1984 advertisement retains an iconic position in the pantheon of successful ad campaigns. It emerged as the second most loved ad of all time in a recent *USA Today* poll, and viewers continue to remember it, although it was aired only once. While most ads lack such a lasting effect, and some of them even might be forgotten immediately, many ads are remembered for a span of time. A few studies have shown that awareness seldom decays instantaneously (Hawkins and Hoch 1992, Wansink and Ray 1992, Gregan-Paxton and Loken 1996, Aravindakshan and Naik 2011).

Standard awareness formation models (see Mahajan et al. 1984), however, assume that awareness decline commences instantaneously. In other words, the standard model does not account for the delayed decay of awareness (see Batra et al. 1995, Naik et al. 1998, Bruce 2008, Srinivasan et al. 2010). Hence, the extant literature does not have a model that allows for the two scenarios: awareness decline begins instantaneously or it declines after a delay. For standard models, Hartl (1987) proves that nonmonotonic advertising such as pulsing is not optimal for a monopolist.

More recently, Aravindakshan and Naik (2011) augment the standard awareness formation model to include such memory effects; they allow for the possibility that awareness decline can be delayed due to the memory for ads. This change converts the standard awareness formation model from an ordinary differential equation (ODE) to a delayed differential equation (DDE). DDEs are a special class of differential equations where the argument is allowed to be "delayed," i.e., its effect on the evolution of the state is not instantaneous (e.g., see Győri and Ladas 1991, Bellen and Zennaro 2003, Arino et al. 2006). Using DDEs, Aravindakshan and Naik (2011) establish the existence of memory and show that their model outperforms the standard awareness formation model when consumers remember ads.

However, Aravindakshan and Naik (2011) investigate the scenario when a brand completely stops advertising, i.e., they estimate ad memorability when the brand does not advertise any more. In practice, brands do advertise periodically over time rather than stop advertising. Our research contributes to the literature by answering the following questions:

(1) Is nonmonotonic pulsing advertising optimal in the presence of ad memorability and the absence of competition?

(2) If ads are remembered, then how should managers determine the optimal spending plan to maximize the total awareness over time?

(3) What are the shapes of optimal advertising policies due to memory effects?

To answer these questions, we derive the optimal advertising policy by solving a brand's optimal control problem with diminishing returns for advertising and in a monopoly setting. First, we include diminishing returns because "continuous-time monopolistic models of advertising expenditure that rely on strict response concavity have been shown to prescribe eventual spending at a constant rate" (Feinberg 2001, p. 1476). In contrast, we prove that pulsing advertising is optimal via continuous-time *delayed* dynamic models in the presence of diminishing returns. Second, we consider a monopolist setting because Hartl (1987) proves that pulsing advertising is not optimal under a broad class of scalar differential equations for a profit-maximizing monopolist. In contrast, we prove that pulsing advertising is optimal via scalar delayed differential equations even for monopolists. The optimal ad policy permits several different shapes—from constant spending to blitz to various kinds of pulsing—depending on whether the memory span exceeds a threshold, which we explicitly characterize.

The rest of the paper proceeds as follows: Section 2 reviews the literature on memory, awareness formation models, and advertising scheduling. Section 3 presents the awareness formation model with ad memorability and links it to prior research. Section 4 derives the normative results and discusses the managerial implications. Section 5 concludes by identifying avenues for further research.

2. Literature Review

2.1. Memory for Advertising

Advertising generates awareness that decays immediately or might be remembered for several weeks due to memory. Several behavioral studies document the existence of memory for ads, which is shown to be a function of several factors such as retrieval cues (Keller 1987), verbal and pictorial content (Unnava and Burnkrant 1991), the length and serial position of ads (Pieters and Bijmolt 1997), post-purchase experience (Braun 1999), or spacing and repetition of ads (Janiszewski et al. 2003, Zielske 1959). Specifically, Janiszewski et al. (2003) find that repeated exposures strengthen the memory for an ad even though time between exposures increases. This phenomenon arises because ads seen the first time leave a memory trace, which is strengthened when the ad appears again due to retrieval of the trace created by the first ad.

Memory for ads can be adversely affected due to a time delay since viewing the ad (Hutchinson and Moore 1984, Hawkins and Hoch 1992) or due to interference effects (see Burke and Srull 1988). For example, Wansink and Ray (1992) quantify memory loss due to time delays by showing that up to 70% of the subjects exposed to advertising could recall the target brand after three months, suggesting a time delay of three months. We emphasize that the empirical evidence for time delay has not been established in the extant literature and it needs further empirical investigation as in the recent study by Aravindakshan and Naik (2011), who suggest three weeks of ad memorability for Peugeot brand's advertising even in the absence of continued advertising support.

2.2. Awareness Formation Models

Empirical literature in marketing contains several models of awareness formation (see Tapeiro 1978, Sethi 1979, Mahajan et al. 1984, Mahajan and Muller 1986, Naik and Raman 2003, Bass et al. 2007, Bruce 2008, Srinivasan et al. 2010), of which the Nerlove-Arrow (NA) model is the most commonly used in marketing (see Figure 1 in Aravindakshan and Naik 2011). These models describe the growth and decay of a brand's awareness over time. In these models the evolution of awareness occurs due to its recent state and the current advertising level. For example, Zielske and Henry (1980) or Mahajan and Muller (1986) apply the classical Nerlove and Arrow (1962) model, where awareness decreases in the absence of advertising at the rate proportional to the *recent* awareness level. However, if ads are remembered for a few weeks, then the awareness today would depend on the awareness prevailing a few weeks ago. In the context of sales, previous studies using distributed lag models (e.g., Griliches 1967, Bass and Clarke 1972) have shown that distant sales affect current sales levels. To summarize, awareness formation models imply that awareness decline commences the instant they are viewed, and that awareness levels are unrelated to those in the more distant past.

To account for memory, Aravindakshan and Naik (2011) adapt the commonly used Nerlove-Arrow model by incorporating time delay. Specifically, awareness increases today due to current advertising, but decreases at a rate proportional to awareness levels that existed τ periods ago, where τ denotes the memory span. To distinguish the present study from Aravindakshan and Naik (2011), we note that they do not consider any advertising input because they focus on how memory-driven awareness evolves in the absence of advertising. In contrast, we focus on how it evolves in the presence of advertising. Hence we address how to optimally allocate advertising spending over time, and how the shapes of optimal advertising vary due to the memory span. To this end, we will apply the extended maximum principle (see the lemma), which is novel to the marketing literature. Before we present these analyses, we review the extant literature on advertising pulsing over time.

2.3. Advertising Pulsing

Advertising literature seeks to address how brand managers should allocate gross rating points worth tens of millions of dollars so that a few concentrated bursts of weekly advertising are interspersed with silent periods of no advertising. The resulting on and off media spending patterns over time are called pulsing media schedules. The practical significance between pulsing versus even schedules boils down to making a "big impact periodically" versus maintaining a "continuous presence." To this end, several studies obtain pulsing by introducing some phenomena in discrete-time models. Simon (1982) incorporates advertising wearout; Park and Hahn (1991) and Villas-Boas (1993) introduce competition between two brands; Bronnenberg (1998) formulates a model with brand switchers and repurchasers;

Dubé et al. (2005) relies on the S-shaped response function; finally, Freimer and Horsky (2012) posit the existence of low and high sales in the long run.

Given the discrete-time formulation, the planning of pulses over 52 weeks poses a dimensionality problem. To see this point, in the first week, consider a media planner who faces two decisions: spend or not spend. In the first two weeks, the planner faces four options: spend or not spend for each of the first week's two possibilities. Progressing in this manner, we observe that the annual binary pulsing plan generates 2^{52} plans (less one for not spending at all), which exceed 4,500 *trillion* plans to be evaluated to find the one optimal plan. This set of possibilities is so immense that a manager processing various plans at the rate of one microsecond per plan would take over a century to discover the optimal plan.

To circumvent this curse of dimensionality, continuous-time models are formulated. Sasieni (1971) pioneers the study of pulsing problems and finds that even spending is the optimal policy (also see Sethi 1977). This finding means that a brand should advertise at a constant level. However, few brands advertise this way. Most brands' advertising policies follow a pulsing pattern, i.e., the brand advertises for some period of time and then stops advertising before restarting. To generate pulsing, Mahajan and Muller (1986) postulate an S-shaped sales response to advertising. When the response function is S-shaped, they argued that pulsing can emerge as an optimal strategy. The resulting strategy requires a very rapid switch between "on" and "off" making it difficult to implement in practice. However, Feinberg (2001) shows that for a wide class of S-shaped response functions pulsing cannot be an optimal strategy. Moreover, the empirical support for the existence of S-shaped response functions is mixed (Rao and Miller 1975, Simon and Arndt 1980, Vakratsas et al. 2004).

Hartl (1987) proved a theorem that shows the optimal ad spending $u^*(t)$ cannot be periodic for models with one state variable, namely, univariate response function $\dot{A} = g(u, A, \theta)$, where $u(t)$ denotes the advertising rate and the parameter vector θ specifies the function $g(\cdot)$. This result implies that a second state variable is required to obtain nonconstant spending plans. For example, besides sales growth, Luhmer et al. (1988) incorporate an additional state variable called adaptation level as per Simon's (1982) ADPULS model. Hahn and Hyun (1991) assume an additional binary state for the presence or absence of advertising costs; Feinberg (1992) introduces the state variable called "filter" that induces inertia on the advertising rate; Mesak (1992) adds the state variable called peak sales from which the actual sales wearout during sustained advertising; Naik et al. (1998) introduce the second state variable via time-varying ad effectiveness. In a personal communication, John D. C. Little characterized this pulsing literature as a "veritable subfield of marketing."

Even the above continuous-time models result in optimal *binary* pulsing with only two levels in a spending plan; that is, $u(t) \in [0, \bar{u}]$. Such binary pulsing does not corroborate with the observed pulsing patterns, which exhibit different spending levels across different weeks. Consequently, extant continuous-time models do not inform managers how much to spend in each week. As Vakratsas and Naik (2007, p. 334) note, "for multipulse schedules, how long should each pulse last? Or should they be equally long? What should be the spacing between pulses? These questions—simple to state, but hard to answer—have remained open for a long time (see, e.g., Corkindale and Newall 1978, Simon 1982, and Table 8.1 Hanssens et al. 1998, p. 254)."

In sum, no study in either discrete- or continuous-time traditions derives multilevel cyclic optimal pulsing plans. Hence the lead question that Little (1986, p. 107) raised—are there any response models for which pulsing (other than chattering) would be optimal?—still remains open. Next, we analyze a continuous-time dynamic model with memory effects and derive the optimal pulsing plan in the presence of diminishing returns (see Feinberg 2001) and the absence of competition (see Hartl 1987).

3. Model Development

3.1. Awareness Formation Model

Based on Aravindakshan and Naik (2011), awareness evolves according to the delay differential equation

$$\dot{A} = \beta\sqrt{u(t)} - \delta A(t-\tau), \tag{1}$$

where $\dot{A} = dA/dt$ and the initial function $A_0(t) = A_0$ over the interval $[-\tau, 0]$. In Equation (1), awareness $A(t)$ and advertising $u(t)$ are nonnegative, the square root function captures the diminishing return to advertising (i.e., marginal impact of advertising decreases as its level increases), and the parameters (β, δ, τ) belong to the nonnegative octant. In their analysis, Aravindakshan and Naik (2011) assume $u(t) = 0$ for all t.

Equation (1) states that the awareness growth depends on not only the brand's advertising $u(t)$ at time t, but also the awareness level $A(t-\tau)$ that prevailed τ periods ago. Consequently, awareness decline does not commence instantly; awareness stays constant for τ periods before the decay sets in. The parameters β and δ measure ad effectiveness and the forgetting rate, respectively. The forgetting rate δ quantifies the proportion of previously aware individuals who no longer remember the ad. The parameter τ represents memorability for advertisements (Aravindakshan and Naik 2011). The parameters (δ, τ) differ conceptually: the forgetting rate δ represents a *proportion* with no units of measurement, whereas the memory span τ reflects *longevity* that is measured in the units of time (e.g., seconds, hours, days). Next, to connect with the extant marketing literature, we map the awareness formation in Equation (1) to current and past advertising in discrete time, although the rest of the analyses proceed in continuous time.

3.2. Mapping the Memory Process

Equation (1) in discrete time can be represented as $A_t - A_{t-1} = \beta\sqrt{u_t} - \delta A_{t-\tau}$. Using the lagged operator L^n to denote the n-period lag, namely, $L^n A_t = A_{t-n}$, we get $A_t - LA_t = \beta\sqrt{u_t} - \delta L^\tau A_t$, which can be expressed as $A_t - LA_t + \delta L^\tau A_t = \beta\sqrt{u_t}$. Then we extract A_t as the common term to obtain $(1 - L(1-\delta L^{\tau-1}))A_t = \beta\sqrt{u_t}$. Thus $A_t = \beta(1 - L(1-\delta L^{\tau-1}))^{-1}\sqrt{u_t}$. To obtain the coefficients that link awareness to current and past advertising, we substitute $x = L(1-\delta L^{\tau-1})$ in the result $(1-x)^{-1} = x^0 + x^1 + x^2 + \cdots$. Hence $(1 - L(1-\delta L^{\tau-1}))^{-1} = \sum_{k=0}^{\infty} L^k(1-\delta L^{\tau-1})^k$. To expand the right-hand side parenthesis, we invoke the binomial theorem and obtain $(1-\delta L^{\tau-1})^k = \sum_{j=0}^{k}\binom{k}{j}(-\delta L^{\tau-1})^j$, where $\binom{k}{j} = k!/(j!(k-j)!)$, and $j = 0, \ldots, k$ and $k \geqslant j$. Consequently, $\sum_{k=0}^{\infty} L^k(1-\delta L^{\tau-1})^k = \sum_{k=0}^{\infty}\sum_{j=0}^{k}\binom{k}{j}(-\delta)^j L^{(\tau-1)j+k} = \sum_{n=0}^{\infty}\sum_{j=0}^{k}\binom{k}{j}(-\delta)^j L^n$, where the summation and combinatorial indices (j, k) are linked to the memory span via $n = (\tau-1)j + k$. Collecting the above results, we find that

$$\begin{aligned} A_t &= \beta(1 - L(1-\delta L^{\tau-1}))^{-1}\sqrt{u_t} \\ &= \beta\sum_{k=0}^{\infty} L^k(1-\delta L^{\tau-1})^k\sqrt{u_t} \\ &= \beta\sum_{k=0}^{\infty}\sum_{j=0}^{k}\binom{k}{j}(-\delta)^j L^{(\tau-1)j+k}\sqrt{u_t} \\ &= \beta\sum_{n=0}^{\infty}\sum_{j=0}^{k}\binom{k}{j}(-\delta)^j L^n\sqrt{u_t}, \\ &\qquad \text{where } n = (\tau-1)j + k \\ &= \beta\sum_{n=0}^{\infty}\sum_{j=0}^{k}\binom{k}{j}(-\delta)^j\sqrt{u_{t-n}}. \end{aligned} \tag{2}$$

In general, starting from u_0 with any feasible β, δ, τ, u_t and A_t, the awareness links to current and past advertising via $A_t = \beta\sum_{n=0}^{t} c_n\sqrt{u_{t-n}}$, where the coefficients are given by

$$c_n = \sum_{j=0\,|\,(j,k)\in(\tau-1)j+k=n}^{k}\binom{k}{j}(-\delta)^j. \tag{3}$$

For example, Table 1 presents the coefficients c_n for $\tau = 3$. Its final column lists the coefficients that link awareness to current and past advertising values.

Finally, we provide an illustration of how the extant marketing literature relates to Equation (3). Specifically, in the standard model $A_t = \beta\sqrt{u_t} + (1-\delta)A_{t-1}$ used in the extant marketing literature, awareness depends on current and past advertising as follows: $A_t = \beta\sqrt{u_t} + \beta(1-\delta)\sqrt{u_{t-1}} + \beta(1-\delta)^2\sqrt{u_{t-2}} + \beta(1-\delta)^3\sqrt{u_{t-3}} + \cdots$. Using Equation (3) for $\tau = 1$, we derive the coefficients $c_n = \sum_{j=0}^{n}\binom{n}{j}(-\delta)^j = (1-\delta)^n$; see Table 2 for details. Comparing the coefficients in the above equation with those in Table 2, we find them to be exactly equal. Thus, Equation (3) not only furnishes a closed-form solution linking awareness to current and past advertising for any memory span τ, but also nests the standard awareness formation model. The next section offers the normative analysis by deriving the optimal advertising strategy when awareness evolves in the presence of ad memorability.

4. Normative Analysis

4.1. Brand's Objective Function

The media planner seeks to maximize a brand's stock of awareness in the most cost-effective manner starting from the awareness levels prevailing prior to the initial time $t = 0$. Let J denote the discounted value of the total awareness generated by an advertising spending plan $u(t)$ over an infinite horizon. Then the resulting performance index is expressed as follows:

$$J(u) = \int_0^\infty e^{-\rho t}(A(t) - cu(t))\,dt, \tag{4}$$

where ρ is the discount rate, and the constant c converts advertising units to awareness points. Equation (4) says that a media planner considers several possible ad spending plans over time, each of which potentially generates a sequence of awareness via Equation (1) that sums up to yield the performance index $J(u(t))$. Additionally, because advertising is nonnegative, we need to ensure that

$$u(t) \geqslant 0. \tag{5}$$

Thus, Equations (1), (4) and (5) formulate the pulsing problem.

We aim to discover whether the optimal trajectory $u^*(t)$—one that maximizes the performance index—exhibits periodicity over time. Specifically, for some future time t', does $u^*(t) = u^*(t+t')$ for patterns other than the even policy or binary pulsing?

The complexity arises from the fact that advertising takes different values from $[0, \infty)$ at different instants t, and so infinitely many trajectories are admissible in the space $\mathcal{U} = \{u(t)\colon \mathbb{R}^+ \to \mathbb{R}^+\}$, where $\mathbb{R}^+$ denotes the nonnegative real line. In contrast, previous pulsing models restricted the set of admissible policies to binary pulsing $u(t) \in [0, \bar{u}]$. Consequently, even a quintessentially periodic advertising such as $u(t) = \bar{u}|Cos(2\pi t)|$ would not be admissible because of the ex ante restriction on the decision space. This restricted set excludes general trajectories, which Park and Hahn (1991, p. 399) acknowledge by stating that such binary pulsing is "suboptimal among various pulsing strategies." Hence, following Feinberg's (2001, p. 1486) call, we have specified the admissible set at the "fullest temporal generality." That is, our decision space admits *any* piecewise continuous function, including multilevel patterns with long stretches of zero spending.

Table 1. Coefficients for $\tau = 3$.

Lag	Feasible (j, k) that satisfy $(\tau - 1)j + k = n$	$c_n = \sum_{j=0}^{k} \binom{k}{j}(-\delta)^j$	Coefficients $\beta \times c_n$
$n = 0$	$\{2j + k = 0\} \Rightarrow (0, 0)$	$\binom{0}{0}(-\delta)^0$	β
$n = 1$	$\{2j + k = 1\} \Rightarrow (0, 1)$	$\binom{1}{0}(-\delta)^0$	β
$n = 2$	$\{2j + k = 2\} \Rightarrow (0, 2)$	$\binom{2}{0}(-\delta)^0$	β
$n = 3$	$\{2j + k = 3\} \Rightarrow (1, 1), (0, 3)$	$\binom{1}{1}(-\delta)^1 + \binom{3}{0}(-\delta)^0$	$\beta(1 - \delta)$
$n = 4$	$\{2j + k = 4\} \Rightarrow (1, 2), (0, 4)$	$\binom{2}{1}(-\delta)^1 + \binom{4}{0}(-\delta)^0$	$\beta(1 - 2\delta)$
$n = 5$	$\{2j + k = 5\} \Rightarrow (1, 3), (0, 5)$	$\binom{3}{1}(-\delta)^1 + \binom{5}{0}(-\delta)^0$	$\beta(1 - 3\delta)$
$n = 6$	$\{2j + k = 6\} \Rightarrow (2, 2), (1, 4), (0, 6)$	$\binom{2}{2}(-\delta)^2 + \binom{4}{1}(-\delta)^1 + \binom{6}{0}(-\delta)^0$	$\beta(1 - 4\delta + \delta^2)$
$n = 7$	$\{2j + k = 7\} \Rightarrow (2, 3), (1, 5), (0, 7)$	$\binom{3}{2}(-\delta)^2 + \binom{5}{1}(-\delta)^1 + \binom{7}{0}(-\delta)^0$	$\beta(1 - 5\delta + 3\delta^2)$
$n = 8$	$\{2j + k = 8\} \Rightarrow (2, 4), (1, 6), (0, 8)$	$\binom{4}{2}(-\delta)^2 + \binom{6}{1}(-\delta)^1 + \binom{8}{0}(-\delta)^0$	$\beta(1 - 6\delta + 6\delta^2)$

Table 2. Coefficients for $\tau = 1$.

Lag	Feasible (j, k) that satisfy $(\tau - 1)j + k = n$	$c_n = \sum_{j=0}^{k} \binom{k}{j}(-\delta)^j$	Coefficients $\beta \times c_n$
$n = 0$	$\{0j + k = 0\} \Rightarrow (0, 0)$	$\binom{0}{0}(-\delta)^0$	β
$n = 1$	$\{0j + k = 1\} \Rightarrow (0, 1), (1, 1)$	$\binom{1}{0}(-\delta)^0 + \binom{1}{1}(-\delta)^1$	$\beta(1 - \delta)$
$n = 2$	$\{0j + k = 2\} \Rightarrow (0, 2), (1, 2), (2, 2)$	$\binom{2}{0}(-\delta)^0 + \binom{2}{1}(-\delta)^1 + \binom{2}{2}(-\delta)^2$	$\beta(1 - \delta)^2$
$n = 3$	$\{0j + k = 3\} \Rightarrow (0, 3), (1, 3), (2, 3), (3, 3)$	$\binom{3}{0}(-\delta)^0 + \binom{3}{1}(-\delta)^1 + \binom{3}{2}(-\delta)^2 + \binom{3}{3}(-\delta)^3$	$\beta(1 - \delta)^3$

To discover the uniquely optimal advertising policy $u^*(t)$, we maximize the performance index in (4) subject to the awareness formation model in (1) and the nonnegativity constraint in (5). Applying the optimal control theory (e.g., Kamien and Schwartz 1991, Sethi and Thompson 2000), we construct the Hamiltonian function as follows:

$$\mathscr{H}(t) = [A(t) - cu(t)] + \lambda(\beta\sqrt{u(t)} - \delta A(t - \tau)) + \kappa u(t). \tag{6}$$

Equation (6) is the long-term performance index similar to the Bellman's value function. Its right-hand side consists of three terms. The first term comes from Equation (4), and it captures the direct contribution due to the current awareness less advertising expenditure.

The second term captures the contribution from a small increase in awareness due to the awareness dynamics in Equation (1). In other words, λ measures the long-term *value of incremental awareness*, and the term $(\beta\sqrt{u(t)} - \delta A(t - \tau))$ predicts the increment in awareness expected due to the current spending given the ad effectiveness, forgetting rate, and memory span. Also known as the costate variable, $\lambda(t)$ quantifies the awareness valuation at each instant t.

The third term ensures the nonnegativity of advertising in Equation (5). Specifically, $u(t) > 0$ if $\kappa(t) = 0$; otherwise $u(t) = 0$ if $\kappa(t) > 0$.

In the absence of memory, when $\tau = 0$, we apply the standard Pontryagin's maximum principle (see Sethi and Thompson 2000, Chap. 2) to obtain the first-order conditions: $\partial\mathscr{H}/\partial u = 0$, $d\lambda/dt = \rho\lambda - \partial\mathscr{H}/\partial A$, and $\partial\mathscr{H}/\partial\kappa = 0$. Solving them simultaneously we obtain the best plan $u^*(t)$, which maximizes both the Hamiltonian function in (6) and the objective function in (4). For applications, see Sasieni

(1971) or Feinberg (1992, 2001). In contrast, when $\tau \neq 0$, the standard maximum principle does not hold. Hence, we first extend this principle and then derive the optimal advertising policy.

4.2. Optimal Advertising Policy

The awareness formation model in (1) is a delay differential equation. The delay occurs because ads are remembered for τ periods. Because this delay exists, we cannot obtain the optimal policy $u^*(t)$ using the standard maximum principle. Specifically, we must not only account for the immediate awareness formation due to current advertising, but also account for the fact that awareness levels linger on because of the memory for advertisements—*the presence of the past*. Hence, based on Kharatishvili (1967), we apply the extended maximum principle:

Lemma (Extended Maximum Principle). *For the control problem with the delay differential Equation* (1), *the first-order conditions are given by*
(*i*) $\partial \mathcal{H}(t)/\partial u = 0$, *and*
(*ii*) $d\lambda/dt = \rho\lambda - \partial\mathcal{H}(t)/\partial A(t) - \partial\mathcal{H}(t+\tau)/\partial A(t)$.

Proof. See the appendix.

Condition (i) is the same as in the standard principle; however, condition (ii) for the costate evolution differs. The right-hand side of the costate evolution consists of three terms. The first term, $\rho\lambda$, captures the current yield of the awareness valuation. The second term, $\partial\mathcal{H}/\partial A$, captures the immediate performance gain due to incremental awareness. The first two terms are the same as in the standard maximum principle. But, because awareness lingers in the presence of memory, the third term $\partial\mathcal{H}(t+\tau)/\partial A$ incorporates the performance gain due to the "after effects" of memory. Thus, today's awareness gain affects both the immediate performance and the distant performance τ periods later.

Intuitively, when consumers who become aware today remember the ads for τ periods, today's awareness increases affect not only the present profit, but also the future profits for subsequent τ periods. Hence, apart from a profit bump today, the "memory effect" provides a continuing bump in profit that lasts for τ periods.

Applying the Lemma stated above, we derive the optimal advertising in the proposition:

Proposition 1. *The optimal periodic advertising is given by*

$$u^*(t) = \left(\max\left[0, \frac{\beta\lambda(t)}{2c}\right]\right)^2, \tag{7}$$

where $\lambda(t)$ is the oscillatory solution to the advanced differential equation:

$$\frac{d\lambda}{dt} = \rho\lambda + \delta\lambda(t+\tau) - 1. \tag{8}$$

Proof. See the appendix.

According to Equation (7), at any time t, the optimal policy equals either zero or it depends on the magnitude of awareness valuation. Equation (8) specifies how awareness valuation, $\lambda(t)$, changes over time. Equation (8) reveals that the current valuation $\lambda(t)$ depends on the future valuation $\lambda(t+\tau)$. This effect of the future-in-present arises because today's ad spending generates awareness that lingers for τ periods subsequently. Hence, the effect of brand's ad spending today lasts up to τ periods into the future. As this memory effect sustains itself for τ periods, today's awareness valuation embodies its future effect for the subsequent τ periods.

To establish that the optimal advertising can be cyclic, we analytically solve the advanced differential equation in (8). We relegate its derivation to the appendix (see Equation (33)) and present the solution here:

$$\lambda(t) = \frac{1}{\rho+\delta} + 2\exp\left(-\frac{\tilde{a}t}{\tau}\right)\cos\left(\frac{bt}{\tau}\right), \tag{9}$$

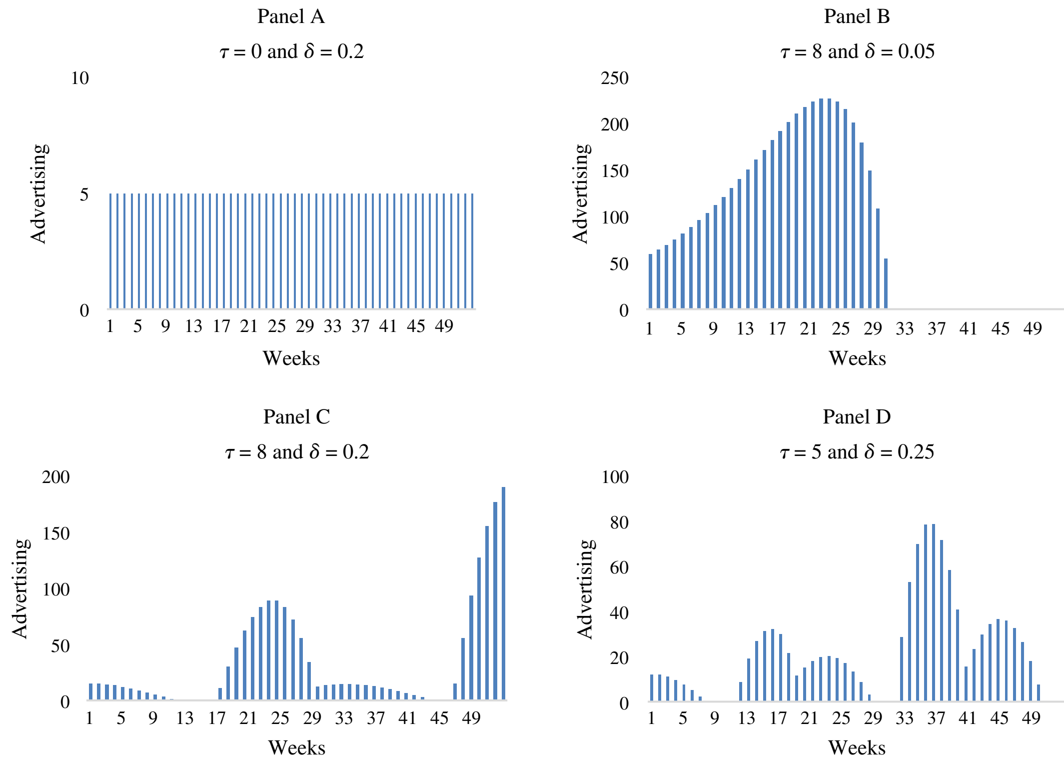
where $\tilde{a} = a - \rho\tau$, and the coefficients a and b are obtained from the Lambert's $W(x)$ evaluated at $x = -\delta\tau e^{\rho\tau}$. See the appendix for details. (For another application of Lambert's W function in static logit models, see Aravindakshan and Ratchford 2011.) The cosine term in Equation (9) induces the oscillations in Equation (7), thus generating pulsing advertising. Note that pulsing does not dampen when $a = \rho\tau$. Intuitively, suppose consumers remember ads for four weeks, then the brand manager advertises in the first week, ceases advertising for the next three weeks because awareness does not decay during that period, and commences advertising again in the fifth week. Thus ad memorability could drive pulsing. Given that the extant literature has established only binary pulsing as optimal, this result marks the first explicit characterization of multilevel optimal cyclic advertising in the presence of diminishing returns and the absence of competition.

Finally, when $\tau = 0$ the awareness valuation remains constant at $\lambda(t) = 1/(\rho+\delta)$, yielding an optimal advertising strategy that is identical to that resulting from the standard model (Sethi 1977, Naik and Raman 2003).

4.3. Shapes of Optimal Advertising Policies

The optimal ad policy in (7) depends on the memory span τ, the awareness decay rate δ, and the ad effectiveness β. Two important findings are that (i) the memory span and decay rate affects the shape of pulsing advertising, whereas (ii) ad effectiveness influences the amount of spending but not the shape. Although these findings can be proven analytically, we illustrate them via numerical examples for clarity. Specifically, over a 52 week span we compute the optimal advertising policy via Proposition (1) and Equation (9) assuming $\rho = 5\%$ per annum, $\beta = 2$, $c = 1$, and

Figure 1. (Color online) Shapes of optimal advertising policy.



varying (δ, τ) from 0 to 1 in steps of 0.05 and 0 to 20 in steps of 1, respectively. To construct Figures 1 and 2 over the full parameter space of (δ, τ), we apply $u(t) = \delta A(t-\tau)/\beta$ as necessary to ensure nonnegative awareness.

When $\tau = 0$, awareness decline begins instantaneously. To maintain awareness levels, a brand must advertise continually (i.e., not periodically) at a constant rate. Hence the optimal advertising remains constant over time. We illustrate the resulting policy in panel A of Figure 1, which displays the constant advertising function. When $\tau \neq 0$, awareness decline is delayed. We illustrate the resulting policies in panels B, C, and D, which display pulsing patterns for various (δ, τ). The pulses emerge because, as advertising builds awareness, a brand can afford to reduce or stop spending for a certain time given that the memory for ads lingers on. Because of the memory effects, the brand is better off by not advertising and relying on consumers' ability to remember the ads.

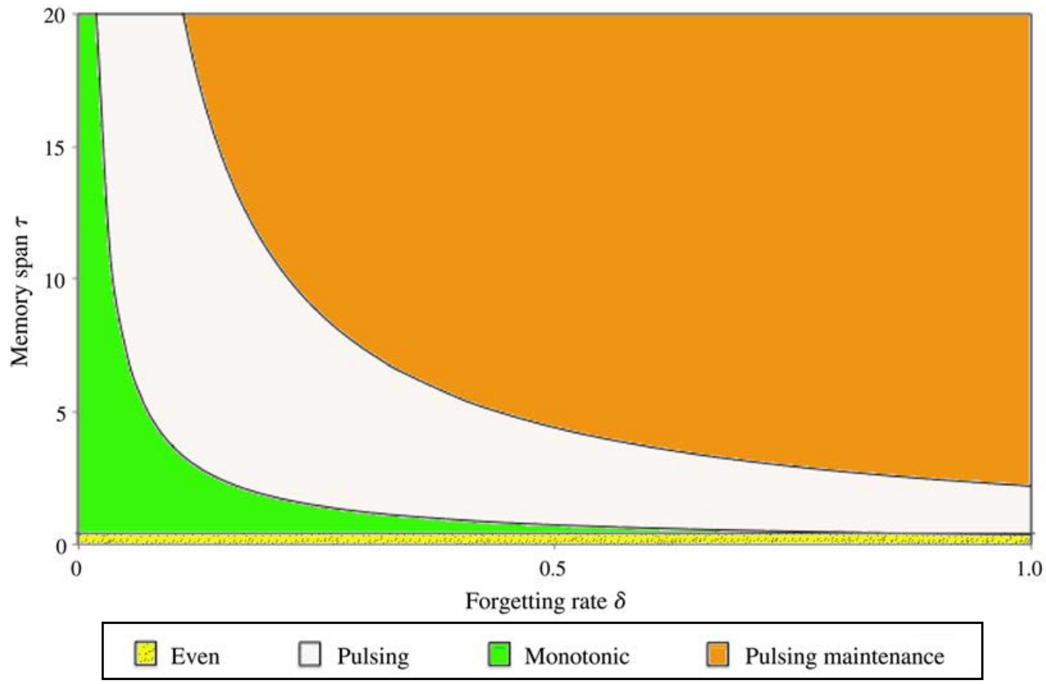
Next, we characterize the regions of optimal pulsing advertising based on the memory span and the forgetting rate. That is, how do the above shapes change as (δ, τ) changes? Because Equation (7) exhibits a variety of shapes, we present Figure 2 that highlights the different patterns for different combinations of (δ, τ). The graph informs managers of the different pulsing advertising possible for different levels of memory span and forgetting rate. We observe that even spending or monotonic spending, blitz advertising, pulsing advertising, and pulsing maintenance—various pulsing types classified and described by Mahajan and Muller (1986)—emerge as *optimal* policies for different memory spans and decay rates. In contrast to the extant pulsing literature that yields only binary pulsing as the optimal policy, this result marks the first illustration of truly cyclic pulsing patterns as optimal.

Moreover, for a wide range of parameter values for the memory span and forgetting rates—in the north-east region above the lower hyperbolic curve in Figure 2—the pulsing patterns occupy a larger area than that under the monotonic or even spending patterns. This fact not only justifies the prevalence of pulsing advertising in practice, but also reveals that the memory span could be the critical factor for the existence of pulsing (because managers afford to reduce or stop ad spending when consumers remember ads and past awareness lingers on). Next, we specify the conditions to decide whether or not to pulse.

4.4. To Pulse or Not to Pulse?

The mere presence of delay does not automatically guarantee that pulsing advertising is optimal. The delay could be critical for pulsing advertising, but its magnitude should be large enough to justify reducing or stopping advertising for a certain period. In the next proposition, we characterize the exact threshold for the memory span, yielding an insight into when brands should engage in pulsing advertising.

Figure 2. (Color online) Optimal advertising scheduling.



Proposition 2. *Pulsing advertising is optimal when the memory span exceeds the threshold*

$$\tau^c > \frac{W(\rho/(\delta e))}{\rho}, \tag{10}$$

where $W(\cdot)$ *is the Lambert's* W *function.*

Proof. See the appendix.

If the memory span falls below the threshold τ^c, managers should not use pulsing—even or monotonic advertising is optimal. For longer spans, they can afford to oscillate the spending levels as in pulsing advertising shown in Figure 2.

Is pulsing advertising optimal if managers were extremely patient? To this end, we take the limits of Equation (10) as the discount rate vanishes. We learn that, even as ρ tends to zero, when managers possess infinite patience, pulsing advertising remains optimal if the memory span exceeds the critical threshold $\tau^c > 1/(\delta e)$. In other words, managers' patience does not jettison the optimality of pulsing, which hinges on consumers' memory for advertisements.

5. Conclusions

A few marketing studies have shown that awareness seldom declines instantaneously. Recently, Aravindakshan and Naik (2011) incorporate the concept of memory span in awareness formation models and show that the memory span for Peugeot brand advertising is about three weeks even after the brands stops advertising. But they do not characterize the nature of optimal advertising over time (e.g., cyclic advertising under certain conditions). Hence the extant literature lacks the understanding of memory effects in pulsing advertising.

In a nutshell, this paper shows that ad memorability plays an important role in determining the optimal advertising policy. It creates the waxing and waning of awareness (state); this rhythm propels the awareness valuation (costate) and influences the spending decisions (control). Intuitively, as awareness lingers due to the memory for ads, this presence of the past awareness creates the after effects on distant awareness valuations, which are reflected in the current spending decisions. This joint dynamic of awareness evolution (past-in-the-present) and awareness valuation (future-in-the-present) endogenously leads to the optimal pulsing advertising. Thus ad memorability plays an important role in driving pulsing.

In closing, we identify three avenues for further research. The first extension is an empirical investigation into the existence of memory span in the presence of advertising. Awareness data are collected by many fast moving consumer goods companies; for example, Srinivasan et al. (2010) present awareness for 21 cereal brands, 19 bottled waters, 19 fruit juices, and 21 shampoo brands on a weekly basis over a seven-year span. To establish the presence of memory span using such data, researchers need to develop an estimation method that accounts for the forward evolution of awareness, backward propagation of the awareness valuation, and the nonnegativity of the estimated parameters as well as advertising and awareness levels as in Equation (1). If awareness data are not available, awareness can be replaced by sales. Indeed, Little (1986, p. 107) states that "Mahajan and Muller study awareness whereas Sasieni considers sales, but the mathematics does not care." The resulting model then allows researchers to explore the role of *delayed* state dependence, which is novel to marketing science.

Another extension deals with sales seasonality, which can be incorporated in two ways. First, dummy variables capture seasonal effects as in Tellis and Franses (2006), who study intraday seasonality using 24 hourly dummies. Second, time transformation can capture seasonal effects as in Radas and Shugan (1998), who shrink time so that it moves at a faster rate when sales are high during high seasons, and stretch time so it progresses slowly when sales are low during low seasons.

Finally, we recall that Hartl's (1987) monotonicity theorem rules out pulsing optimality for a monopolist. To generate pulsing strategy, the state space needs to be augmented, for example, by including a competing brand. In contrast, we generate optimal pulsing policies for a monopolist. In other words, competition is not necessary to yield optimal pulsing; rather a monopolist has incentives to pulse even in the absence of competition. To affirm this insight, we analyzed the monopoly setting in our model. Future researchers should extend our analyses to dynamic oligopoly markets by combining the techniques in this paper (e.g., extended maximum principle) with those from differential games literature (see Jørgensen and Zaccour (2004), Naik et al. 2008). Based on such rigorous methods, would pulsing still be optimal for both the brands? If so, the resulting insights can broaden our understanding because most studies in the extant literature on pulsing in competitive markets find conditions for the optimality of "binary pulsing" (i.e., some sequence of 1010001), which is not realistic because brand managers advertise unequally in various weeks (i.e., multilevel pulsing as in panels B, C, or D of Figure 1). We hope the proposed model and methods help attain congruence between the optimal and actual pulsing strategies in competitive markets.

Acknowledgments

The authors benefited from the invaluable comments of John D. C. Little, Gary Lilien, Marnik Dekimpe, Bart Bronnenberg, and Dan Romik (UC Davis Mathematics); and the helpful suggestions from participants in the seminars at the Penn State University, Indiana University, National University of Singapore, the 2012 Winter Camp of the Catholic University in Leuven, and the 2012 Marketing Science Conference in Boston. The authors also thank the anonymous reviewers, associate editor, and area editor for a constructive review process.

Appendix

Proof of Lemma

We derive the necessary conditions to solve control problems in the presence of delays. Specifically, we aim to maximize the objective function with respect to the functional $u(t)$,

$$\int_0^\infty F(t, x(t), u(t))\, dt, \tag{11}$$

and subject to the dynamic constraint,

$$x' = \frac{dx}{dt} = g(x(t-\tau), u(t)), \tag{12}$$

and the preshape function over $[-\tau, 0]$,

$$x(t) = x_0. \tag{13}$$

As in nonlinear programming, we augment (11) by adjoining (12) via the Lagrange multiplier $\lambda(t)$ to get

$$J = \int_0^\infty \{F(t, x, u) + \lambda(g(\tilde{x}, u) - x')\}\, dt, \tag{14}$$

where $\tilde{x} = x(t-\tau)$. We then apply integration-by-parts to the last term to obtain,

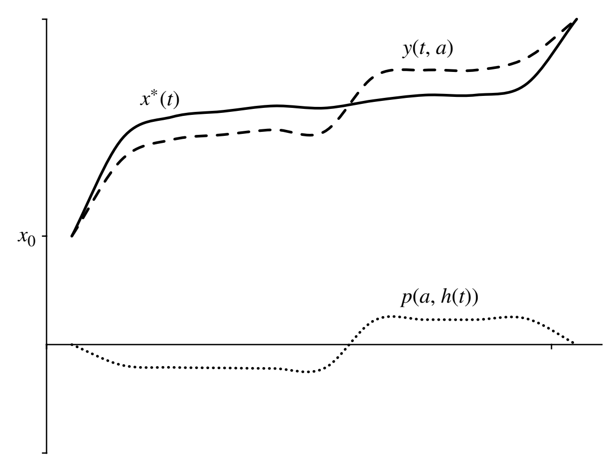
$$\int_0^\infty \lambda x'\, dt = \lambda(\infty)x(\infty) - \lambda(0)x(0) - \int_0^\infty x\lambda'\, dt. \tag{15}$$

Because $\lambda(\infty)x(\infty) = 0$ due to the transversality condition, we substitute (15) in (14) to get,

$$J = \lambda(0)x_0 + \int_0^\infty \{F(t, x, u) + \lambda g(\tilde{x}, u) + x\lambda'\}\, dt. \tag{16}$$

Let $u^*(t)$ denote the optimal control that maximizes (11), and $x^*(t)$ be the corresponding optimal state trajectory. Let us consider an alternative control $u(t) = u^*(t) + ah(t)$, where $h(t)$ is any arbitrary function and a is a parameter. The new (suboptimal) control $u(t)$ results in an alternative state trajectory $y(t, a) = x^*(t) + p(a, h(t))$, where $p(\cdot)$ denotes the "perturbation" due to (a, h). When $a = 0$, both the new control and state trajectories coincide with the optimal control and the state trajectory. Figure A.1 shows the optimal state trajectory $x^*(t)$ and the alternative state $y(t, a)$ trajectory, starting from the same initial condition $y(0, a) = x_0$ for all a.

Figure A.1. State trajectories.



Next, we evaluate the performance metric (16) under the alternative control path $u(t)$ and the resulting state trajectory $y(t,a)$ to obtain

$$J(a)=\lambda(0)x_0+\int_0^\infty \{F(t,y(t,a),u^*+ah(t)) + \lambda g(y(t-\tau,a),u^*+ah(t))+y(t,a)\lambda'\}\,dt. \quad (17)$$

Taking the first total variation of (17) at $a=0$ and denoting $dJ(a)/da\,|_{a=0}=J'(0)$, we obtain

$$J'(0)=\int_0^\infty \{(F_x(t,x^*,u^*)+\lambda')\Delta x+(F_u(t,x^*,u^*) + \lambda g_u(\tilde{x}^*,u^*))\Delta u+\lambda g_{\tilde{x}}(\tilde{x}^*,u^*)\Delta\tilde{x}\}\,dt, \quad (18)$$

where the subscripts denote partial derivatives. Also, the first term in (17) vanishes because x_0 is a constant.

Then, we incorporate the preshape function (13) in (18). To this end, we isolate the term $\int_0^\infty \lambda g_{\tilde{x}}(\tilde{x}^*,u^*)\Delta\tilde{x}\,dt$, change its variables to $\xi=t-\tau$, and express it over the two intervals $[-\tau,0)$ and $[0,\infty)$ as follows:

$$\int_{-\tau}^0 \lambda(\xi+\tau)\frac{\partial g(x^*(\xi),u^*(\xi+\tau))}{\partial x(\xi)}\Delta x(\xi)\,d\xi + \int_0^\infty \lambda(\xi+\tau)\frac{\partial g(x^*(\xi),u^*(\xi+\tau))}{\partial x(\xi)}\Delta x(\xi)\,d\xi. \quad (19)$$

Observe that $\Delta x(\xi)=0$ because $x(\xi)$ is a constant over $[-\tau,0)$ due to (13). Thus, only the second term of (19) remains. Substituting (19) back in (18) and noting that ξ is an integration dummy, we obtain

$$J'(0)=\int_0^\infty \left\{\left(F_x(t,x^*,u^*)+\lambda'+\lambda(t+\tau)\frac{\partial g(x^*,u^*(t+\tau))}{\partial x}\right) \cdot\Delta x+(F_u(t,x^*,u^*)+\lambda g_u(\tilde{x}^*,u^*))\Delta u\right\}dt. \quad (20)$$

Because $u^*(t)$ is the maximizing control, $J(a)$ attains its maximum at $a=0$; that is, $J'(0)=0$. Hence we select $\lambda(t)$ to make the first integrand of (20) vanish, and so we get one of the necessary conditions:

$$\lambda^{*'}=-F_x(t,x^*,u^*)-\lambda^*(t+\tau)\frac{\partial g(x^*,u^*(t+\tau))}{\partial x}. \quad (21)$$

Also the second integrand on the right-hand side of (20) must equal zero for any arbitrary value of Δu. In particular, it must hold for the specific value of $\Delta u=F_u(t,x^*,u^*)+\lambda g_u(\tilde{x}^*,u^*)$ which yields $\int_0^\infty\{[(F_u(t,x^*,u^*)+\lambda g_u(\tilde{x}^*,u^*))]^2\}\,dt$. This latter integral is always positive and so it vanishes if

$$F_u(t,x^*,u^*)+\lambda g_u(\tilde{x}^*,u^*)=0, \quad (22)$$

which provides the other necessary condition.

Finally, we define the Hamiltonian function,

$$\mathcal{H}(t)=H(t,x,\tilde{x},u)=F(t,x,u)+\lambda g(\tilde{x},u), \quad (23)$$

so that the time-translated Hamiltonian is $\mathcal{H}(t+\tau)=F(t+\tau,x(t+\tau),u(t+\tau))+\lambda(t+\tau)g(x(t),u(t+\tau))$, and then (21) can be reexpressed equivalently as

$$\lambda^{*'}=-\frac{\partial\mathcal{H}(t)}{\partial x}-\frac{\partial\mathcal{H}(t+\tau)}{\partial x},$$

proving condition (ii) in the Lemma after noting $F(t,x,u)=e^{-\rho t}f(x,u)$ and using the current value Hamiltonian and costate (see Sethi and Thompson 2000). Similarly, (22) can be reexpressed equivalently as

$$\frac{\partial\mathcal{H}(t)}{\partial u}=0,$$

proving condition (i) in the lemma.

Proof of Proposition 1

Below we apply the above Lemma to derive the optimal periodic advertising. Using Equations (1) and (4), we formulate the Hamiltonian as

$$\mathcal{H}(t)=A(t)-cu(t)+\lambda(t)(\beta\sqrt{u(t)}-\delta A(t-\tau)). \quad (24)$$

To ensure the nonnegativity constraint in Equation (5), we augment the Hamiltonian by introducing the multiplier $\kappa(t)$ in the Lagrangian function:

$$L(t)=\mathcal{H}(t)+\kappa(t)u(t). \quad (25)$$

We obtain the optimal advertising by solving the first-order condition, $\partial L(t)/\partial u(t)=-c+\lambda(t)\beta/(2\sqrt{u(t)})+\kappa(t)=0$. Rearranging its terms, we find that

$$\sqrt{u^*(t)}=\frac{\lambda(t)\beta}{2(c-\kappa(t))}. \quad (26)$$

Next, by applying the condition (ii) in the lemma we get

$$\lambda'=\rho\lambda+\delta\lambda(t+\tau)-1. \quad (27)$$

The costate values in (27) can be positive ($\lambda(t)>0$) or negative ($\lambda(t)<0$). Hence we find the optimal advertising under both the cases:

i. Positive Costate: $\lambda(t)>0$. When $\lambda(t)>0$, $(\lambda(t)\beta)/(2(c-\kappa(t)))>0$, and so $u^*(t)>0$, which implies $\kappa(t)=0$ due to the complementary slackness condition; therefore $u^*(t)=(\lambda(t)\beta/(2c))^2$.

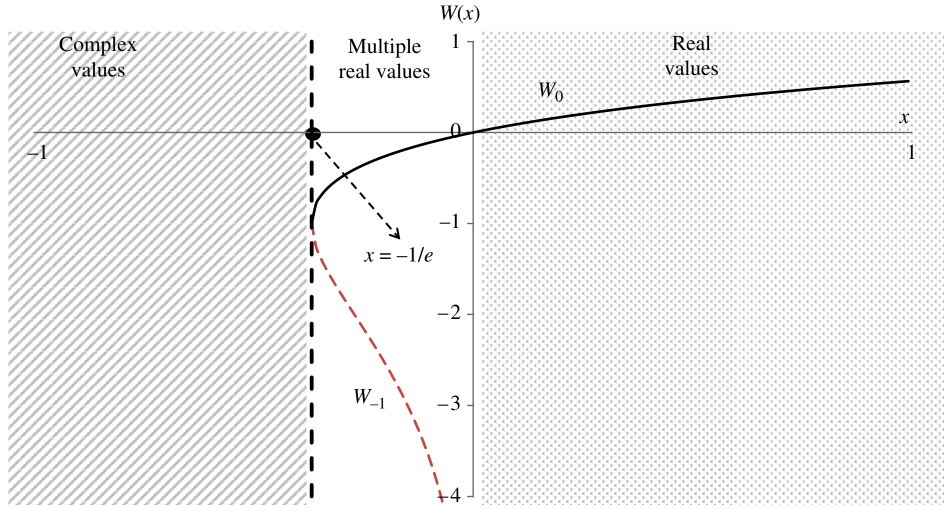
ii. Negative Costate: $\lambda(t)<0$. Rewriting (26), we find $2\sqrt{u^*(t)}(c-\kappa(t))=\lambda(t)\beta$. When $\lambda(t)<0$, $\lambda(t)\beta<0$ because $\beta>0$, and so $2\sqrt{u^*(t)}(c-\kappa(t))<0$, or $c-\kappa(t)<0$, which implies $\kappa(t)>c>0$, which implies $\kappa(t)>0$. Therefore $u^*(t)=0$ due to the complementary slackness condition.

Together, cases (i) and (ii) along with the case when $\lambda(t)=0$ furnish the optimal advertising strategy:

$$u^*(t)=\left(\max\left[0,\frac{\beta\lambda(t)}{2c}\right]\right)^2. \quad (28)$$

This ends the proof for Proposition 1.

Figure A.2. Lambert W function.



Proof of Proposition 2

Below we characterize the explicit oscillatory solution to awareness valuation, then identify the condition for oscillatory solutions, and finally discover the threshold of memory span for oscillations.

To explicitly solve the advanced differential Equation (8), we denote $\lambda_H(t)$ and $\lambda_P(t)$ as the homogeneous and particular solutions, respectively, so that the overall solution is given by,

$$\lambda(t) = \lambda_H(t) + \lambda_P(t). \tag{29}$$

To obtain the homogeneous solution, let $\lambda_H(t) = e^{st}$, so $\dot{\lambda}_H = se^{st}$ and $\lambda_H(t+\tau) = e^{s(t+\tau)}$. Substituting $(\lambda_H, \dot{\lambda}_H)$ in $\dot{\lambda}(t) = \rho\lambda + \delta\lambda(t+\tau)$ gives $se^{st} = \rho e^{st} + \delta e^{s(t+\tau)}$, which upon rearranging yields $se^{-s\tau} - \rho e^{-s\tau} = \delta$. Multiplying both sides by $-\tau e^{\rho\tau}$ we obtain

$$(-s\tau + \rho\tau)e^{-s\tau+\rho\tau} = -\delta\tau e^{\rho\tau}. \tag{30}$$

To solve (30), we apply the Lambert's W function, which solves the equations of the form $ye^y = x$ by $y = W(x)$. It is related to the logarithm function, which solves the equations of the form $e^y = x$ by $y = \text{Log}(x)$. Corless et al. (1996) discuss the properties and applications of the W function. Applying the W function to (30), we get $-s\tau + \rho\tau = W(-\delta\tau e^{\rho\tau})$ and so $s = -(-\rho\tau + W(-\delta\tau e^{\rho\tau}))/\tau$.

Next, we describe some properties of the Lambert's $W(x)$ function needed for the proof. Specifically, it possesses two real-valued branches: $W_0(x)$ and $W_{-1}(x)$. Figure A.2 presents $W_0(x)$ as the bold curve, $W_{-1}(x)$ as the dashed curve, and the three regions: (i) when $x > 0$, $W_0(x) > 0$; (ii) when $-1/e < x < 0$, $W_0(x) < 0$ and $W_{-1}(x) < 0$ (hence multivalued); and (iii) when $x < -1/e$, both $W_0(x)$ and $W_{-1}(x)$ are complex valued.

Then, using $s = -(-\rho\tau + W(-\delta\tau e^{\rho\tau}))/\tau$, we obtain the homogenous solution $\lambda_H(t) = \exp(-((-\rho\tau + W_0(-\delta\tau e^{\rho\tau}))/\tau)t) + \exp(-((-\rho\tau + W_{-1}(-\delta\tau e^{\rho\tau}))/\tau)t)$.

To obtain the particular solution, we let $\lambda_P(t) = A + Bt$, so then $\lambda_P(t+\tau) = A + Bt + B\tau$ and $\dot{\lambda}_P = B$. Substituting $(\lambda_P, \dot{\lambda}_P)$ in $\dot{\lambda}(t) = \delta\lambda(t+\tau) - 1$, we get $B = \rho A + \delta Bt + \delta B\tau + \delta A - 1$. Equating coefficients on both sides of the equality, we find that $B = 0$ and $A = 1/(\rho+\delta)$ and so the particular solution is $\lambda_P(t) = 1/(\rho+\delta)$.

By substituting $\lambda_H(t)$ and $\lambda_P(t)$ in (29), we solve the dynamics of awareness valuation in closed form:

$$\lambda(t) = \frac{1}{\rho+\delta} + \exp\left(-\frac{-\rho\tau + W_0(-\delta\tau e^{\rho\tau})}{\tau}t\right) + \exp\left(-\frac{-\rho\tau + W_{-1}(-\delta\tau e^{\rho\tau})}{\tau}t\right). \tag{31}$$

To identify when the solution (31) oscillates, we recall that both $W_0(x)$ and $W_{-1}(x)$ are complex valued when $x < -1/e$. So Equation (31) oscillates when $-\delta\tau e^{\rho\tau} = x < -1/e$, offering the explicit condition:

$$\delta\tau e^{\rho\tau} > \frac{1}{e}. \tag{32}$$

Finally, to explicitly characterize the oscillatory solutions when (32) holds, we note the conjugacy of $W_0(-\delta\tau e^{\rho\tau}) = a + bi$ and $W_{-1}(-\delta\tau e^{\rho\tau}) = a - bi$, where $i = \sqrt{-1}$. Then (31) becomes $\lambda(t) = 1/(\rho+\delta) + \exp(-(\tilde{a} + bi)t/\tau) + \exp(-(\tilde{a} - bi)t/\tau)$, where $\tilde{a} = -\rho\tau + a$. Applying Euler's formula, we simplify $\lambda(t) = 1/(\rho+\delta) + \exp(-\tilde{a}t/\tau)[\cos(bt/\tau) - i\cdot\sin(bt/\tau)] + \exp(-\tilde{a}t/\tau)[\cos(bt/\tau) + i\cdot\sin(bt/\tau)]$, which results in

$$\lambda(t) = \frac{1}{\rho+\delta} + 2\exp\left(-\frac{\tilde{a}t}{\tau}\right)\cos\left(\frac{bt}{\tau}\right). \tag{33}$$

Thus, $\lambda(t)$ oscillates because of the cosine term and so does the optimal advertising $u^*(t) = (\max[0, \beta\lambda^*(t)/2c])^2$. The oscillations do not dampen when $a = \rho\tau$.

Finally, to derive the critical threshold, we rewrite (32) as $\rho\tau e^{\rho\tau} > \rho/(\delta e)$. Then using the Lambert W function we obtain the threshold $\tau^c > W(\rho/(\delta e))/\rho$. This ends the proof for Proposition 2.

References

Aravindakshan A, Naik P (2011) How does awareness evolve when advertising stops? The role of memory. *Marketing Lett.* 22(3):315–326.

Aravindakshan A, Ratchford B (2011) Solving share equations in logit models using the Lambert W function. *Rev. Marketing Sci.* 9(1):1–19.

Arino O, Hbid ML, Dads EA (2006) *Delay Differential Equations and Applications*, NATO Science Series (Springer-Verlag, Dordrecht, The Netherlands).

Bass FM, Bruce N, Majumdar S, Murthi BPS (2007) Wearout effects of different advertising themes: A dynamic Bayesian model of the ad-sales relationship. *Marketing Sci.* 26(2):179–195.

Bass FM, Clarke DG (1972) Testing distributed lag models of advertising effect. *J. Marketing Res.* 9(3):298–308.

Batra R, Lehmann DR, Burke J, Pae J (1995) When does advertising have an impact? A study of tracking data. *J. Advertising Res.* 35(5):19–32.

Bellen A, Zennaro M (2003) *Numerical Methods for Delay Differential Equations* (Oxford Science Publications, Clarendon Press, Oxford, UK).

Braun KA (1999) Postexperience advertising effects on consumer memory. *J. Consumer Res.* 25(March):319–334.

Bronnenberg BJ (1998) Advertising frequency decisions in a discrete Markov process under a budget constraint. *J. Marketing Res.* 35(August):399–406.

Bruce NI (2008) Pooling and dynamic forgetting effects in multi-theme advertising: Tracking the advertising sales relationship with particle filters. *Marketing Sci.* 27(4):659–673.

Burke RR, Srull TK (1988) Competitive interference and consumer memory for advertising. *J. Consumer Res.* 15(1):55–68.

Corkindale D, Newall J (1978) *Advertising Threshold and Wearout* (MCB Publications, Bradford, UK).

Corless RM, Gonnet GH, Hare DEG, Jeffrey DJ, Knuth DE (1996) On the Lambert *W* function. *Advances Comput. Math.* 5(1):329–359.

Dubé JP, Hitsch G, Manchanda P (2005) An empirical model of advertising dynamics. *Quant. Marketing Econom.* 3(2):107–144.

Feinberg FM (1992) Pulsing policies for aggregate advertising models. *Marketing Sci.* 11(3):221–234.

Feinberg FM (2001) On continuous-time optimal advertising under *S*-shaped response. *Management Sci.* 47(12):1476–1487.

Freimer M, Horsky D (2012) Periodic advertising pulsing in a competitive market. *Marketing Sci.* 31(4):637–648.

Gregan-Paxton J, Loken B (1996) Understanding memory for ads: A process view. Wells WD, ed. *Measuring Advertising Effectiveness* (Lawrence Erlbaum Associates, Mahwah, NJ), 183–202.

Griliches Z (1967) Distributed lags: A survey. *Econometrica* 35(1):16–49.

Győri I, Ladas GE (1991) Oscillation theory of delay diferential equations: With applications. *Oxford Mathematical Monographs* (Clarendon Press, Oxford, UK).

Hahn M, Hyun JS (1991) Advertising cost interactions and the optimality of pulsing. *Management Sci.* 37(2):157–169.

Hanssens D, Parsons L, Schultz R (1998) *Market Response Models: Econometric and Time Series Analysis.* 5th Printing (Kluwer Academic Publishers, Boston).

Hartl RF (1987) A simple proof of the monotonicity of the state trajectories in autonomous control problems. *J. Econom. Theory* 41:211–215.

Hawkins SA, Hoch SJ (1992) Low-involvement learning: Memory without evaluation. *J. Consumer Res.* 19(2):212–225.

Hutchinson WJ, Moore DL (1984) Issues surrounding the examination of delay effects in advertising. Kinnear T, ed. *Advances in Consumer Research* (Association for Consumer Research, Provo, UT), 650–655.

Janiszewski C, Noel H, Sawyer AG (2003) A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *J. Consumer Res.* 30(June):138–149.

Jørgensen S, Zaccour G (2004) *Differential Games in Marketing*, International Series in Quantitative Marketing (Kluwer Academic Publishers, The Netherlands).

Kamien M, Schwartz N (1991) *Dynamic Optimization* (Elsevier Science Publishing , New York).

Keller KL (1987) Memory factors in advertising: The effect of advertising retrieval cues on brand evaluations. *J. Consumer Res.* 14(December):316–333.

Kharatishvili GL (1967) A maximum principle in extremal problems with delays. Balakrishnan AV, Neustadt LW, eds. *Mathematical Theory of Control* (Academic Press, New York), 26–34.

Little JDC (1986) Comments: Advertising pulsing policies for generating awareness for new products. *Marketing Sci.* 5(2):107–108.

Luhmer A, Steindl A, Feichtinger G, Hartl R, Sorger G (1988) ADPULS in continuous time. *Eur. J. Oper. Res.* 34:171–177.

Mahajan V, Muller E (1986) Advertising pulsing policies for generating awareness for new products. *Marketing Sci.* 5(2):89–111.

Mahajan V, Muller E, Sharma S (1984) An empirical comparison of awareness forecasting models of new product introduction. *Marketing Sci.* 3(3):179–197.

Mesak HI (1992) An aggregate advertising pulsing model with wearout effects. *Marketing Sci.* 11(3):310–326.

Naik PA, Raman K (2003) Understanding the impact of media synergy in multimedia communications. *J. Marketing Res.* 40(4):375–388.

Naik PA, Mantrala MK, Sawyer A (1998) Planning media schedules in the presence of dynamic advertising quality. *Marketing Sci.* 17(3): 214–235.

Naik PA, Prasad A, Sethi SP (2008) Building brand awareness in dynamic oligopoly markets. *Management Sci.* 54(1):129–138.

Nerlove M, Arrow KJ (1962) Optimal advertising policy under dynamic conditions. *Economica* 29(114):129–142.

Park S, Hahn M (1991) Pulsing in a discrete model of advertising competition. *J. Marketing Res.* 28(4):397–405.

Pieters RGM, Bijmolt THA (1997) Consumer memory for television advertising: A field study of duration, serial position, and competition effects. *J. Consumer Res.* 23(March):362–372.

Radas S, Shugan SM (1998) Seasonal marketing and timing of new product introductions. *J. Marketing Res.* 35(3):296–315.

Rao AG, Miller PB (1975) Advertising/sales response functions. *J. Advertising Res.* 15(2):1–15.

Sasieni M (1971) Optimal advertising expenditure. *Management Sci.* 18(December):64–72.

Sethi SP (1977) Optimal advertising for the Nerlove-Arrow model under a budget constraint. *Oper. Res. Quart.* 28(3):683–693.

Sethi SP (1979) A note on Nerlove-Arrow model under uncertainty. *Oper. Res.* 27(4):839–842.

Sethi SP, Thompson GL (2000) *Optimal Control Theory: Applications to Management Science and Economics*, Second ed. (Springer, Heidelberg, Germany).

Simon H (1982) ADPULS: An advertising model with wearout and pulsation. *J. Marketing Res.* 19(August):352–363.

Simon JL, Arndt J (1980) The shape of the advertising response function. *J. Advertising Res.* 20(August):11–28.

Srinivasan S, Vanhuele M, Pauwels K (2010) Mind-set metrics in market response models: An integrative approach. *J. Marketing Res.* 47(4):672–684.

Tapeiro CS (1978) Optimum advertising and goodwill under uncertainty. *Oper. Res.* 26(3):450–463.

Tellis GJ, Franses PH (2006) Optimal data interval for estimating advertising response. *Marketing Sci.* 25(3):217–229.

Unnava HR, Burnkrant RE (1991) An imagery-processing view of the role of pictures in print advertisements. *J. Marketing Res.* 28(May): 226–231.

Vakratsas D, Feinberg FM, Bass FM, Kalyanraman G (2004) The shape of advertising response functions revisited: A model of dynamic probabilistic thresholds. *Marketing Sci.* 23(1):109–119.

Vakratsas D, Naik PA (2007) Essentials of planning media schedules. Tellis G, Ambler T, eds. *Handbook of Advertising* (Sage Publications, London), 333–348.

Villas-Boas M (1993) Predicting advertising pulsing policies in oligopoly: A model and empirical test. *Marketing Sci.* 12(1):88–102.

Wansink B, Ray ML (1992) Goal-related consumption and extension advertising: The impact on memory and consumption. Sherry JF Jr, Sternthal B, eds. *Advances in Consumer Research*, Vol. 19 (Association for Consumer Research, Provo, UT), 806–812.

Zielske HA (1959) The remembering and forgetting of advertising. *J. Marketing* 23(3):239–243.

Zielske HA, Henry WA (1980) Remembering and forgetting television ads. *J. Advertising Res.* 20(2):7–13.

Ashwin Aravindakshan is an assistant professor of marketing at the University of California, Davis. His research interests focus on understanding the role of marketing communications at both the aggregate and individual level. He develops new approaches to determine the effectiveness of marketing instruments and to optimize the budget and the allocation of these marketing instruments over time. His research has been published in *Management Science*, *Marketing Science*, and the *Journal of Marketing Research*, among others.

Prasad A. Naik is a professor of marketing at the University of California, Davis. His research appears in top journals including *Nature Reviews*, *Automatica*, *Biometrika*, the *Journal of Marketing Research*, *Marketing Science*, *Management Science*, the *Journal of the American Statistical Association*, the *Journal of the Royal Statistical Society*, the *Journal of Econometrics*, and the *Accounting Review*. He is a recipient of the Chancellor's Fellow, AMS Doctoral Dissertation Award, Frank Bass Award, MSI Young Scholar, AMA Consortium Faculty, O'Dell Award Finalist, JIM Best Paper Award, and Professor of the Year for outstanding teaching on multiple occasions.