Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Design of Experiential Services with Acclimation and Memory Decay: Optimal Sequence and Duration

Aparupa Das Gupta, Uday S. Karmarkar, Guillaume Roels

MANAGEMENT SCIENCE
Vol. 62, No. 5, May 2016, pp. 1278–1296
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2015.2172

The Design of Experiential Services with Acclimation and Memory Decay: Optimal Sequence and Duration

Aparupa Das Gupta
Cisco Systems, Inc., San Jose, California 95134, dg.aparupa@gmail.com

Uday S. Karmarkar, Guillaume Roels
UCLA Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095
{uday.karmarkar@anderson.ucla.edu, guillaume.roels@anderson.ucla.edu}

For many consumer-intensive (i.e., business-to-consumer) services, delivering memorable customer experiences is a source of competitive advantage. Yet there are few guidelines available for designing service encounters with a focus on customer satisfaction. In this paper, we show how experiential services should be sequenced and timed to maximize the satisfaction of customers who are subject to memory decay and acclimation. We find that memory decay favors positioning the highest service level near the end, whereas acclimation favors maximizing the gradient of service level. Together, they maximize the gradient of service level near the end. Although memory decay and acclimation lead to the same design individually, they can act as opposing forces when considered jointly. Overall, our analysis suggests that short experiences should have activities scheduled as a crescendo and duration allocated primarily to the activities with the highest service levels, whereas long experiences should have activities scheduled in a U-shaped fashion and duration allocated primarily to activities with the lowest service level so as to ensure a steep gradient at the end.

Keywords: service design; experience; scheduling; social psychology; behavioral operations
History: Received September 14, 2013; accepted January 21, 2015, by Serguei Netessine, operations management. Published online in *Articles in Advance* September 11, 2015.

1. Introduction

As epitomized by Walt Disney's entertainment parks and Benihana's restaurants, firms that deliver outstanding customer experiences achieve greater customer satisfaction and therefore greater customer loyalty (Voss and Zomerdijk 2007, DeVine and Gilson 2010, DeVine et al. 2012), which ultimately drives future sales (Heskett et al. 1997).

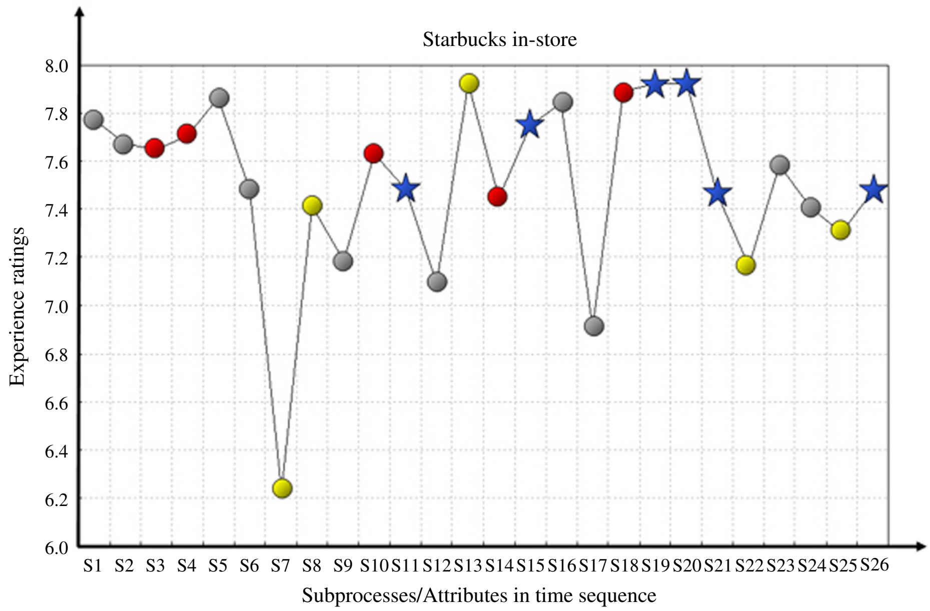
Because service experiences are dynamic processes, the utility that customers derive from an experience evolves dynamically, in response to varying stimuli (Bitran et al. 2008, Baucells and Sarin 2012). For instance, Figure 1 depicts a map, over time, of the customer experience at Starbucks. The *total* experienced utility of a customer from the service will therefore integrate all of the instantaneous utilities experienced throughout the service encounter (Edgeworth 1881).

However, the customers' *remembered* utility, which drives their future purchase decisions, may differ from their total experienced utility (Kahneman et al. 1997). In particular, their remembered utility from an experience tends to be highly correlated to the peak and the final utilities (Fredrickson 2000). In addition, greater weight may be placed on the initial experiences (aka the primacy effect) because of attention decrement (Crano 1977) or assimilation (Asch 1946). Behavioral biases also affect how much (instantaneous) utility is derived from an outside stimulus. For instance, acclimation makes people adapt to states but react to changes (Hsee and Tsai 2008), satiation reduces the marginal utility obtained from accumulated consumption (Baucells and Sarin 2013), and loss aversion weights losses more than gains (Tversky and Kahneman 1991). Savoring and dread may furthermore affect the utility experienced before the service delivery (Loewenstein 1987).

As a result, maximizing customer satisfaction in services requires a deep understanding of customer behavior (Chase and Dasu 2008), both during and between service encounters (Bitran et al. 2008, Watkinson 2013). In fact, Cook et al. (2002) argue that, by drawing on concepts from psychology and sociology, service encounters can be designed with the same depth and rigor as manufacturing processes. However, there exist to date few guidelines for designing service experiences, despite their ubiquity in today's economy (Pine and Gilmore 1998).

In this paper, we show how to sequence and allocate duration to activities in a service encounter so as to maximize the satisfaction of a customer who is

Figure 1 (Color online) Map of Starbucks' In-Store Customer Experience



Source. Lee (2008). Used with permission.
Notes. The stars represent subprocesses that are important to the customer and differentiate the brand. The circles represent subprocesses that either are not important to the customer or do not differentiate the brand.

subject to acclimation and memory decay. Acclimation makes people adapt to states but react to changes (Hsee and Tsai 2008); because of acclimation, the first bites of a dish are usually the most enjoyable (Miller 2014). Memory decay, on the other hand, makes people forget what happened in the past; accordingly, the last moments of an encounter are the most memorable. We focus on acclimation and memory decay for the following reasons. First, they have a long tradition in psychology and have therefore received strong empirical support. Second, these seemingly simple biases are sufficiently rich to explain classical empirical findings on remembered utility, such as that sometimes "more pain is preferred to less" (Kahneman et al. 1993) or that interrupting a pleasant experience with an annoying break can result in greater satisfaction (Nelson and Meyvis 2008). Third, they appear to be mathematically symmetric in our models, and we therefore offer a joint treatment. Nevertheless, we acknowledge that other behavioral biases may also be salient in service experiences, and we leave it for future research to investigate how our service design results are affected by these other biases.

To answer our research question, we build a stylized model of a service encounter with the following characteristics:

- The encounter has a fixed total duration and consists of a given set of activities that have a common acclimation parameter.
- Each activity is associated with a given unidimensional service level, independent of its placement in the sequence and of its duration.
- Customers are captive; i.e., customers cannot arrive after the beginning of the encounter or leave before its end.
- The provider, not the customer, has control over the sequencing of activities and/or the allocation of duration to them.

Although stylized, our model could apply, as a first-order approximation, to an executive education program on a specific topic (e.g., "current trends in supply chain management"). The different activities would correspond there to the different sessions offered (e.g., sustainability, offshoring), and their respective service levels could be measured by their historical teaching ratings. In that context, both memory decay and acclimation are salient (Hoogerheide and Paas 2012). In practice, however, student satisfaction may be based on multiple attributes (Marsh 1984). Moreover, the program director may have additional levers than just the scheduling and duration of each class to influence student satisfaction (such as company visits and the timing of the main reception). Although our stylized model ignores those dimensions, we expect that our managerial insights would remain applicable to those more complex settings.

Other examples of service encounters where sequence and duration of events affect the overall experience are concerts, magic shows, museum exhibitions, fireworks, city walking or bus tours, cruises, spa treatments, fitness classes, or medical procedures such as colonoscopies and dental visits.

We find that a service design that is optimal for memory decay alone is also optimal for acclimation alone, although memory decay and acclimation are fundamentally different mechanisms. In particular, without precedence constraints, it is optimal to sequence activities in *increasing* order of service levels so as to finish at a peak and, with variable durations, to lengthen the duration of those activities with the *highest* service level. Hence, memory decay and acclimation act in the same way when considered individually.

However, when memory decay and acclimation are considered jointly, we find that they may act in opposite directions. Specifically, when both phenomena are present, it may be optimal to sequence activities in a *U-shaped* fashion, beginning and finishing with high service levels, and to lengthen the duration of activities with the *lowest* service level.

Because memory decay favors a high service level near the end and acclimation favors a steep rise in service levels, the combination of these effects favors a *steep gradient* of service level *near the end* of the encounter. Hence, the often-heard "finish strong" recommendation (Chase and Dasu 2001) should be understood as "finish with a steep ascent" in the presence of acclimation and memory decay. A grand finale becomes even more memorable when it comes just after a respite.

Overall, our analysis suggests that short experiences should have activities scheduled as a crescendo and duration allocated primarily to the activities with the highest service levels, whereas long experiences should have activities scheduled in a U-shaped fashion and duration allocated primarily to activities with the lowest service level so as to ensure a steep gradient at the end.

The remainder of this paper is organized as follows. In §2, we review the related literature. In §3, we introduce the model for customer satisfaction. In §4, we characterize the optimal solution for the sequencing and duration allocation problem in three different settings, depending on whether the sequence and durations are fixed or variable. In §5, we numerically compare different sequencing heuristics when customers are heterogeneous in terms of their rates of acclimation and memory decay. We present our concluding remarks in §6. The proofs appear in the appendix. The mathematical programming formulation of the problem, heuristic algorithms, and upper bounds based on Lagrangian relaxations (Fisher 2004) appear in an electronic companion (available as supplemental material at http://dx.doi.org/10.1287/mnsc.2015.2172).

2. Literature Review

This paper is related to three bodies of research: the experimental study of human behavior in psychology and marketing, the analytical models of experienced utility in decision theory, and the prescriptive analytical models of service design in operations management. We next review these three streams of literature and position our paper accordingly.

2.1. Behavioral Aspects

Research in psychology and marketing has identified many factors that influence the retrospective evaluation of experiences, including the trend of an experience (Loewenstein and Prelec 1993), its rate of change (Hsee and Abelson 1991), and its maximum and final intensities (Fredrickson and Kahneman 1993). Underlying these factors are psychological phenomena such as acclimation, loss aversion, and memory decay. We contribute to this literature by adopting a design perspective on two such phenomena—namely, memory decay and acclimation. We then relate these two phenomena to the celebrated "peak–end" rule in retrospective evaluation.

2.1.1. Memory Decay. People naturally tend to forget past events more than recent ones. The most prevalent models of memory decay are an exponential model and a power model; see Anderson (1995) for a review. The exponential model of memory decay was constructed by Ebbinghaus (1913) from his study on how much memory of monosyllabic words was retained after several days. Bahrick (1984) also reports an exponential decay of memory of Spanish words over a time span of 50 years, although memory ultimately does not completely disappear but remains constant after 3-6 years. Wickelgren (1974) derives a combined exponential power model of memory decay and argues that the exponential term provides a better fit for short-term memory data and that the power term is a better fit for longer-term memory data. Nevertheless, the difference between the exponential and the power models of memory decay is often marginal. For instance, Wickelgren (1974) reports that both models account for 99% of the variance of the original data collected by Ebbinghaus.

Moreover, exponential models are widely used. For instance, Naik (1999) uses an exponential decay model to estimate the half-life of advertising campaigns, and Watt et al. (1993) apply it to measure the salience of news. Surveying marketing models of awareness of a new product introduction, Mahajan et al. (1984) report that almost all models explicitly capture forgetting, and the effect is, in general, exponential. In this paper, we adopt Ebbinghaus's (1913) exponential model of memory decay because of its simplicity and relevance.

The rate of memory decay is, in general, context dependent (Anderson 1995): although some experiments show memories over a few seconds (Anderson 1995), data from Ebbinghaus (1913) are over several days, Naik (1999) reports that the half-life of an advertising campaign can be as long as three months, and Bahrick (1984) reports a half-life span of multiple years.

2.1.2. Acclimation. Acclimation makes people adapt to states but react to changes (Hsee and Tsai 2008). This phenomenon is ubiquitous in physiological processes, going as far back as Newton's law of cooling, which states that the rate of change of the temperature of an object is proportional to the difference between its own temperature and the ambient temperature (Incropera et al. 2006). Acclimation is often referred to as adaptation, although life sciences tend to distinguish the two phenomena by associating the former to changes within an organism's lifetime and the latter to changes across several generations of a particular species. Besides physiological processes, Helson (1964) suggests that adaptation also governs cognitive processes, and Brickman and Campbell (1971) propose that adaptation also governs emotions, giving rise to a hedonic treadmill. For instance, in waiting lines, customers have been reported to become gradually demoralized as they wait but to positively respond to each advance of the queue (Carmon and Kahneman 1996). In learning experiences, moderately difficult material generates more utility if it follows difficult material than if it follows easy material (Hoogerheide and Paas 2012). Tversky and Griffin (1991) highlight the duality between endowment and contrast: positive (negative) experiences make us happy (unhappy) but also render similar experiences less exciting (less bad). Hsee and Abelson (1991) emphasize the importance of the rate of change, as opposed to just the contrast, in satisfaction. Because of acclimation, the satisfaction from a pleasant experience may actually be higher when the experience is interrupted by an annoying break than with no interruption (Nelson and Meyvis 2008, Nelson et al. 2009).

Similar to memory decay, the rate of acclimation is highly context dependent. For physiological processes, complete adaptation is often in the order of minutes (Gent and McBurney 1978, Overbosch 1986, Dalton and Wysocki 1996) or even seconds (Dawes and Watanabe 1987). On the other hand, the experiments on commercial interruptions by Nelson et al. (2009) span several minutes. Yet other research fails to find contrast in hedonic consumption (Novemsky and Ratner 2003).

2.1.3. Peak–End Rule. Remembered utility and experienced utility are different constructs and require different modeling approaches (Kahneman et al. 1997). On the one hand, the total experienced utility, which has normative appeal, integrates all (undiscounted) instantaneous utilities (Kahneman et al. 1997) and therefore weights them by their respective durations, in the same spirit as Edgeworth (1881); see Kahneman et al. (2004) for practical measurement methods.

On the other hand, the remembered utility from an experience, which governs future decisions (Kahneman et al. 1993), is highly correlated with its maximum and final intensities, a phenomenon called "the peak–end rule"; see Varey and Kahneman (1992), Fredrickson and Kahneman (1993) and Fredrickson (2000). As a consequence, remembered utility is typically not significantly affected by the duration of the experience and may violate temporal monotonicity, in the sense that adding more pain may result in greater remembered utility (Kahneman et al. 1997). In particular, in waiting lines, a long queue that ends with a very rapid advance may elicit a more favorable retrospective evaluation than a shorter queue (Carmon and Kahneman 1996). In learning experiences, adding moderately difficult material to a difficult lesson may make the overall lesson perceived as easier to learn (Hoogerheide and Paas 2012).

Although the peak–end rule is a well-accepted paradigm, it is only a "good first approximation" (Kahneman 2000a, p. 697). In particular, the peak–end rule suggests that the remembered utility from an experience is independent of the sequence of all but the last activity, i.e., irrespective of whether there are multiple peaks, valleys, breaks, intermissions, etc. However, subsequent research has shown the importance of the valence of trends (Loewenstein and Prelec 1993, Ariely 1998, Rozin et al. 2004, Dixon and Verma 2013) and their rates (Hsee and Abelson 1991, Baumgartner et al. 1997), as well as, in some cases, of duration (Ariely 1998, Schreiber and Kahneman 2000).

Instead of enhancing the peak–end rule to account for trend and/or duration, as is done in some studies (e.g., Dixon and Thompson 2016), we propose here an alternative model of remembered utility. Specifically, we assume that the remembered utility is the discounted sum of instantaneous utilities, where discounting applies backward in time, consistent with the memory decay models reviewed in §2.1.1. In this stylized model of remembered utility, the end utility carries greater weight, because it is the least discounted, and the peak utility has more impact than the other utilities by definition of its being a peak. Moreover, temporal monotonicity may also be violated in our model, as we illustrate in §3.1.1, and trends may affect remembered utility. Hence, our model captures the key features of the peak–end rule, as well as trends and (if the discount rate is not too high) durations. Moreover, our model is rooted in a fundamental behavioral bias, namely memory decay, unlike the peak–end rule, which captures only its symptoms.

2.2. Decision Theory

Kahneman et al. (1997) propose that there exist two kinds of experienced utility: the one reported in real time (instantaneous) and the other based on retrospective evaluation (remembered). Using this idea, our model represents the remembered utility as a function of the instantaneous utility.

Following Constantinides (1990), Wathieu (1997) proposes a model of habit formation, which adapts the reference level to the current consumption level. Baucells and Bellezza (2016) extend that model to account for anticipation and recall, and Baucells and Sarin (2013) embed it into a general framework for happiness. Consistent with Wathieu's habit formation model, our acclimation model adapts the reference level to the current service level.

We contribute to this literature by adopting a design perspective and by focusing on maximizing (ex post) remembered utility as opposed to (ex ante) total experienced utility. There is indeed a lot of evidence that people make decisions based on their memories of experiences (Kahneman et al. 1997). Accordingly, services should be designed to enhance future memories (Watkinson 2013). Because of our different objective, our results differ. For instance, Wathieu (1997) shows that in the presence of acclimation, the optimal consumption pattern that maximizes ex ante discounted utility should be decreasing, increasing, or U-shaped, so as to trade off time discounting (favoring immediate consumption) with acclimation (favoring a positive gradient of consumption). By contrast, we show that, with retrospective evaluation, the optimal sequence could be either increasing or U-shaped, but never decreasing. Baucells and Sarin (2007, 2010) respectively show that U-shaped sequences maximize total experienced utility in the presence of satiation with time discounting and with acclimation. We corroborate their result with memory decay and acclimation by taking the perspective of remembered utility.

2.3. Service Operations

Analytical models of service operations typically focus on managing capacity (Gans and Zhou 2003), scheduling resources (Pinedo 2005, Sampson and Weiss 1995), and pricing (Talluri and Van Ryzin 2004). In the context of service design, Bellos and Kavadias (2014) study how to allocate tasks between the service provider and the customer. In contrast to their model, which considers customer experience as random, we adopt here a behavioral model of customer experience.

Lately, some papers have studied how operational decisions are affected by customer behavior. The literature in this area is diverse in terms of modeling paradigms and behavioral regularities. In particular, Nasiry and Popescu (2011) study pricing, Dixon and Verma (2013) and Dixon and Thompson (2015, 2016) study event scheduling decisions, and Verhoef et al. (2004) study call center satisfaction in light of the peak–end rule. Aflaki and Popescu (2013), Popescu and Wu (2007), and Nasiry and Popescu (2011) study how loss aversion affects pricing and service-level policies. Caro and Martínez-de-Albéniz (2012) show how satiation affects assortment decisions. Ely et al. (2015) characterize the optimal revelation of information when customers have preference for suspense or surprise. Adelman and Mersereau (2013) study how customer memory effects impact a supplier's profits when the supplier dynamically allocates limited capacity among a portfolio of customers. Focusing on queueing experiences, Plambeck and Wang (2013) consider the effect of hyperbolic discounting on unpleasant services, and Carmon et al. (1995) show how customer dissatisfaction can be reduced by spreading the service across the waiting period in queues.

Our work is most closely related to that of Aflaki and Popescu (2013) and Dixon and Thompson (2015, 2016). Aflaki and Popescu (2013) characterize the optimal service level between service encounters so as to maximize long-term customer retention and profitability. By contrast, we focus on activities within a service encounter, assuming that customers are captive, and we control the duration and sequence of activities with fixed service levels. Hence, our approach can be viewed as complementary to theirs. Dixon and Thompson (2015, 2016) study an event scheduling problem by creating bundles of events in the presence of peak–end, trend, and other outcomes of psychological phenomena. By contrast, we directly model the psychological phenomena, and not their outcomes. Moreover we adopt an analytical approach, whereas theirs is computational.

3. Model

We consider a service provider who seeks to maximize ex post customer satisfaction from a service encounter. The service encounter consists of a sequence of n activities. Each activity $i \in \{1, \ldots, n\}$ is designed to offer a fixed service level x_i for a duration t_i, with $\sum_{i=1}^{n} t_i = T$. We assume that the service level is unidimensional and is independent of the activity's placement in the sequence or its duration. For instance, the executive education program in supply chain management may consist of different sessions (e.g., sustainability, offshoring), each with a specific duration (e.g., one or two hours) and a specific service level (e.g., based on historical teaching ratings). Our model is applicable to both positive and negative values of the service level x_i.

A service design is characterized by the following two design variables: (1) the sequence of activities $(i_1, \ldots, i_n)$, where $i_k \in \{1, \ldots, n\}$ is the kth activity in the sequence, and (2) the duration for each activity t_i, $\forall i$. For any sequence $(i_1, \ldots, i_n)$, let $T_k = \sum_{j=1}^{k} t_{i_j}$ be the time passed by the end of the kth activity and let $\bar{T}_k = \sum_{j=k}^{n} t_{i_j}$ be the time remaining from the beginning of the kth activity to the end of the encounter. We next build the customer satisfaction utility model and then formulate the service provider's design problem.

3.1. Customer Satisfaction Utility

We consider a customer who is subject to memory decay and acclimation. We denote by $u_{i_k}(t)$ the customer's utility at time t while experiencing the kth activity in the sequence, with $T_{k-1} < t \le T_k$. The customer satisfaction from the whole experience, evaluated at the end of the encounter, will then integrate all such instantaneous utilities. Acclimation affects how much instantaneous utility $u_{i_k}(t)$ is obtained from each activity, whereas memory decay determines the relative weight of the contribution of $u_{i_k}(t)$ in the overall remembered utility $S((i_1, \ldots, i_n), \mathbf{t})$, as illustrated in Figure 2.

3.1.1. Customer Memory Process. We assume that, in the satisfaction, i.e., remembered utility, from a service encounter, the instantaneous utilities experienced throughout the encounter are weighted by their recency. By contrast, the peak–end rule gives positive weight only to the peak and the end utilities, with no consideration of their duration (Fredrickson 2000). For the memory decay process, we use the exponential forgetting model proposed by Ebbinghaus (1913). According to this model, the memory from an encounter weights activities at the end of the encounter more heavily than the ones at the beginning. Accordingly, memory decay operates like a backward discounting process. When the customer discounts past experiences exponentially with rate $w \in [0, \infty)$, the customer's remembered utility or satisfaction from the service experience is given by

$$S((i_1, \ldots, i_n), \mathbf{t}) = \sum_{k=1}^{n} \int_{T_{k-1}}^{T_k} u_{i_k}(t) e^{-w(T-t)}\, dt. \qquad (1)$$

As an illustration, we apply Equation (1) to the data collected during the colonoscopy experiment by Redelmeier and Kahneman (1996). Figure 3 plots the instantaneous (dis)utility reported by two patients experiencing two different colonoscopy procedures, one short (patient A) and one long (patient B). Assuming a memory decay rate of $w = 0.3$ per minute (which corresponds to a half-life of approximately 2.3 minutes), we find that patient A experiences a total discomfort of $S = -12.1$, whereas patient B experiences a total discomfort of $S = -6.9$.[1] Hence, our memory decay model explains why, in that experiment, the patient subject to the longest procedure reported overall less discomfort than the patient subject to the shortest procedure, similar to what Redelmeier et al. (2003) reported. Although, of course, there exist other interpretations of that result (including acclimation), this illustrates that memory decay alone is rich enough to explain the surprising outcome that more pain can be preferred to less (Kahneman et al. 1993).

3.1.2. Customer Acclimation Process. We next introduce our acclimation model, following the adaptation model of Wathieu (1997) and Baucells and Sarin (2013). As illustrated in Figure 2, the acclimation process affects how much instantaneous utility is derived from a service level. Let $b(t)$ be the customer's reference level at time t. The instantaneous utility experienced at time t is a function of the difference between the current service level and the reference point:

$$u_{i_k}(t) = U(x_{i_k} - b(t)).$$

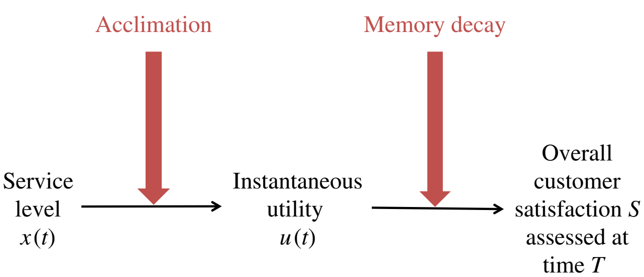
Hence, the instantaneous utility, $u_{i_k}(t)$, captures both the valence (good or bad) and the intensity (mild to extreme) of the instantaneous experience (Kahneman 2000b).

The rate of change of the reference point is proportional to the difference between the service level and the reference point, akin to Newton's law of cooling and other adaptation models in life sciences (e.g., Overbosch 1986):

$$\frac{db(t)}{dt} = \alpha(x(t) - b(t)),$$

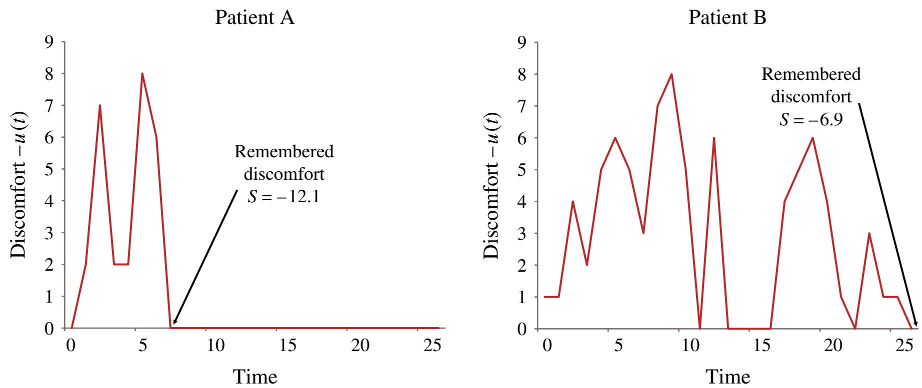
where $\alpha > 0$ is the rate of acclimation. A high value of α implies that the customer is less influenced by

Figure 2 (Color online) Acclimation Affects Instantaneous Utility and Memory Decay Determines Its Relative Weight in the Overall Remembered Utility



[1] We assume here that patients evaluate their satisfaction S just at the end of the procedure. In the experiment, they were asked to report it within an hour after the procedure. Although some memory decay could also happen during that relapse period, potentially at a different rate than the one prevailing during the operation (Baucells and Bellezza 2016), our result would continue to hold in that case, provided that both patients had the same relapse period.

Figure 3 (Color online) Because of Memory Decay, Adding More Pain at the End Can Result in Greater Remembered Utility



Note. To measure overall satisfaction (S), we assume $w = 0.3$ per minute.

past service levels. Solving this differential equation for activity i_k yields

$$b(t) = x_{i_k} - (x_{i_k} - b(T_{k-1}))e^{-\alpha(t-T_{k-1})}, \quad T_{k-1} \le t \le T_k, \qquad (2)$$

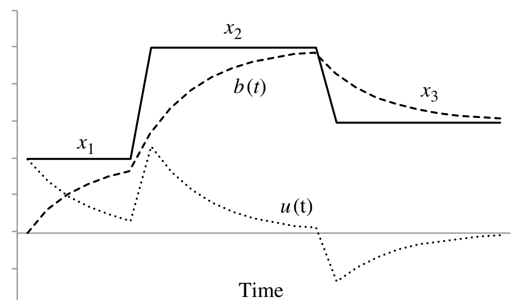
where the initial reference level $b(0)$ captures the history of past experiences.

We assume that the activities in a service encounter have a common acclimation parameter, such as different sessions in a one-day executive program or different art galleries in a guided museum tour. Accordingly, the reference point at the end of an activity carries over to the beginning of the next activity; i.e., $\lim_{t\uparrow T_k} b(t) = \lim_{t\downarrow T_k} b(t)$. If activities potentially have different parameters for acclimation (e.g., waiting for a table and eating dinner), then the reference point could be discontinuously reset once a new activity is started. Baucells and Sarin (2010) make a similar assumption to generate insights.

In addition, we assume that the utility function is linear, $U(x-b) = u_0 + (x-b)$, where u_0 is the intrinsic utility from the experience, normalized to zero in the sequel since it does not affect the design decisions. In practice, utility could be nonlinear, potentially exhibiting different behaviors for gains and losses (Kahneman 2000b). Moreover, it could be nonmonotone. For instance, with temperature, it is likely that one's utility would peak when the temperature is close to 70°F. In those cases, a linear model can be viewed as a first-order approximation of the utility function around x, if the variations in service levels tend to be within the same order of magnitude.

Figure 4 illustrates the acclimation process under those assumptions for a service encounter consisting of three activities such that $x_2 > x_3 > x_1$. The acclimation level $b(t)$ always trails the service level. Therefore, for a constant service level $x(t)$, the instantaneous utility decreases over time and eventually becomes zero (Ariely 1998). By contrast, an upward or downward jump in service level across consecutive activities leads to a rapid increase or decrease in the instantaneous utility level, respectively. Hence,

Figure 4 The Acclimation Process Leads to a Decay in Utility Over Time for a Fixed Service Level and Yields Discontinuous Jumps in Utility After a Change in Service Level



our simple acclimation model captures the observation that people adapt to states but react to changes (Hsee and Tsai 2008).

Under the assumption of same rate of acclimation for all activities, we can expand (2) to obtain a closed-form expression for the reference level:

$$b(t) = x_{i_k} - \left((x_{i_1} - b(0)) + \sum_{j=2}^{k}(x_{i_j} - x_{i_{j-1}})e^{\alpha T_{j-1}}\right)e^{-\alpha t}, \quad T_{k-1} \le t \le T_k.$$

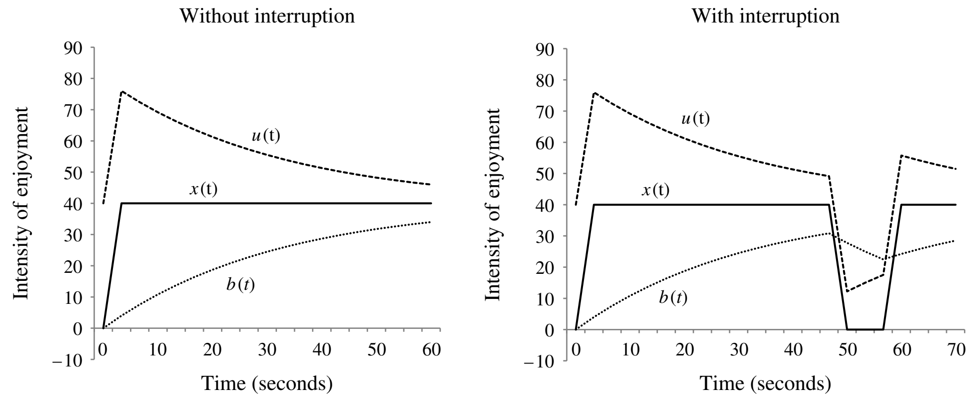
Therefore, because $u_{i_k}(t) = x_{i_k} - b(t)$, we obtain

$$u_{i_k}(t) = (x_{i_1} - b(0))e^{-\alpha t} + \sum_{j=2}^{k}(x_{i_j} - x_{i_{j-1}})e^{-\alpha(t-T_{j-1})}, \quad T_{k-1} \le t \le T_k. \qquad (3)$$

The instantaneous utility $u_{i_k}(t)$ can thus be expressed as a discounted sum of the past changes in service levels, consistent with the intuition that acclimation depends on the gradient of service levels.

As an illustration, we apply the acclimation model to replicate the results from the break insertion experiment by Nelson and Meyvis (2008, p. 660). In their experiment, the inclusion of a break (e.g., an irritating guitar feedback) in an enjoyable experience (e.g., a

Figure 5 Interrupting a Positive Experience Disrupts Acclimation, Making the Experience More Enjoyable



Note. In this experiment, we assume that $\alpha = 1.8$ per minute.

playlist of songs) disrupts the acclimation process. Since in their experiment the break is of the same nature as the experience, we consider the same acclimation rate throughout, and we assume it to be set to $\alpha = 1.8$ per minute (which corresponds to a half-life of 0.4 minutes). As shown in Figure 5, the acclimation model indicates that, if there is no break (left panel), the instantaneous utility gradually decreases over time. Whereas if there is an unpleasant break (right panel), the instantaneous utility first gradually declines, then falls sharply during the break, and then goes back up at the end of the break. Moreover, at the end of the experiment, the utility of the customers who experienced the unpleasant break (right panel) dominates the utility of the customers who did not experience the break (left panel), consistent with in Nelson and Meyvis (2008, Figure 3). Hence, our simple acclimation model can explain why inserting unpleasant breaks may result in higher utility following the break. Moreover, we can show that, with both acclimation and memory decay, the overall satisfaction from the experience may be higher with the unpleasant break than without, as was empirically observed by Nelson and Meyvis (2008).

3.1.3. Customer Satisfaction Model. Combining (1) and (3) leads to the following model of cumulative remembered utility in the presence of acclimation and memory decay:

$$S((i_1, \ldots, i_n), \mathbf{t}) = \sum_{k=1}^{n} \int_{T_{k-1}}^{T_k} u_{i_k}(t) e^{-w(T-t)}\, dt = \sum_{k=1}^{n} (x_{i_k} - x_{i_{k-1}}) \frac{e^{-\alpha \bar{T}_k} - e^{-w\bar{T}_k}}{w - \alpha}. \tag{4}$$

Equation (4) reveals that, under our modeling assumptions, memory decay and acclimation, despite being fundamentally different mechanisms, play a symmetric role mathematically on ex post satisfaction. Specifically, in (4), the customer weights a change in service level from the $(k-1)$th to the kth activity by

$$\Phi(\alpha, w, \bar{T}_k) \doteq \frac{e^{-\alpha \bar{T}_k} - e^{-w\bar{T}_k}}{w - \alpha} = \int_0^{\bar{T}_k} e^{-\alpha t} e^{-w(\bar{T}_k - t)}\, dt. \tag{5}$$

Rewriting (4) as follows,

$$S = \sum_{k=1}^{n} (x_{i_k} - x_{i_{k-1}}) \Phi(\alpha, w, \bar{T}_k) = \sum_{k=1}^{n} x_{i_k} \big(\Phi(\alpha, w, \bar{T}_k) - \Phi(\alpha, w, \bar{T}_{k+1}) \big), \tag{6}$$

indicates that memory decay, which is fundamentally a weighting of the activity service levels x_{i_k}, can also be interpreted as a weighting of the activity increments $(x_{i_k} - x_{i_{k-1}})$ and, conversely, that acclimation, which is fundamentally a weighting of the service-level increments, can also be interpreted as a weighting of the service levels themselves.

3.2. Service Provider's Design Problem: Sequence and Duration

We consider a service provider who seeks to optimize the design of a service encounter to maximize ex post customer satisfaction. We assume that the population of customers is homogeneous, i.e., they have the same values of the parameters α and w, and we test the robustness of our results by that assumption in §5. We consider two design variables, i.e., the sequencing and the duration of activities.

Because of resource limitations, activity durations are typically constrained. In particular, we assume that activity durations must lie within certain bounds, i.e., $\underline{\tau}_i \le t_i \le \bar{\tau}_i$ such that $\underline{\tau}_i \ge 0$, $\bar{\tau}_i > 0$, $\forall\, i$. When $\underline{\tau}_i = 0$, we are effectively allowing for activity i not to be included in the encounter. Moreover, we consider a situation in which the total duration of the service encounter is fixed to T, i.e., $\sum_i t_i = T$, such as a one-day executive program.[2]

[2] With variable total duration, one would also need to explicitly model the pricing decision and how it relates to the customer's

In addition to (or instead of) controlling the duration of activities, a service provider may also have some control over the sequence of activities. The admissible sequence of activities can be restricted by precedence constraints; for instance, a fundamental course in operations management must be taken before an elective. Let $\mathcal{P}$ be the set of all possible feasible sequences of activities.

Then, using (4), the service provider's design problem (SPDP) is given by

$$\max_{(i_1,\ldots,i_n),\,\mathbf{t}} \; S((i_1,\ldots,i_n),\mathbf{t}) = \sum_{k=1}^{n}(x_{i_k}-x_{i_{k-1}})\Phi(\alpha,w,\bar{T}_k) \tag{7}$$

$$\text{subject to} \quad \underline{\tau}_i \le t_i \le \bar{\tau}_i, \ \forall\, i, \tag{8}$$

$$\sum_{i=1}^{n} t_i = T, \tag{9}$$

$$(i_1,\ldots,i_n)\in\mathcal{P}. \tag{10}$$

Going forward, we will denote the optimal sequence by $(i_1^*,\ldots,i_n^*)$ and the optimal duration by $\mathbf{t}^*$. In the electronic companion, we propose a mathematical programming formulation of SPDP.

Different service industries may face different degrees of flexibility regarding the sequencing and the allocation of duration to activities. In certain industries (e.g., music performances), the sequence of activities is variable, but their duration is fixed (VSFD). In SPDP, this can be modeled by setting $\underline{\tau}_i = \bar{\tau}_i$, $\forall\, i$. In other industries (e.g., spa treatments), the sequence of activities is fixed, but their duration is variable (FSVD). In SPDP, this can be modeled by having only one element in $\mathcal{P}$. Finally, other industries (e.g., fitness classes) are characterized by both variable sequences and variable durations (VSVD). Overall, this leads to three different design problems; see Table 1. To obtain a general characterization, we will assume that, whenever the sequence is variable, i.e., in VSFD and in VSVD, there are no precedence constraints; i.e., $\mathcal{P}$ is the set of all permutations.[3]

expectations, since different durations (e.g., a half-day or one-day executive program) are typically associated with different prices and therefore with different customer's expectations. By contrast, the durations of the activities within the encounter (e.g., sessions on specific topics) are typically not specified contractually (within certain bounds) and are typically not reflected in the price of the service. We leave it for future research to explore the interrelationship between total duration, pricing, and customer expectations. In case there is no such interrelationship, such as in colonoscopy procedures (Redelmeier et al. 2003), our model can easily be extended to accommodate variable durations.

[3] Because our design insights remain identical between VSFD (no precedence constraints) and FSVD (fixed sequence), we expect them to carry over to cases of variable sequences with precedence constraints.

Table 1 **Three Stylized Models of a Service Encounter**

	Sequence	
Duration	Fixed	Variable
Fixed		Variable sequence, fixed duration (VSFD): session chair at a conference, music performance
Variable	Fixed sequence, variable duration (FSVD): spa treatment, dental procedure	Variable sequence, variable duration (VSVD): personal fitness class, museum tours, fireworks

4. Analysis

In this section, we characterize the optimal sequence and duration allocation solutions. After some preliminaries, we sequentially consider VSFD, FSVD, and VSVD. Underlying the optimal design in each scenario are the following two properties of memory decay and acclimation:

- Memory decay favors a high service level at the end of the encounter.
- Acclimation favors large positive gradients for successive service levels.

Together, they maximize the gradient of the service level near the end of the service encounter.

4.1. Preliminaries

We first characterize three properties of the weighting function $\Phi(\alpha,w,t)$, defined in (5), which will allow us to characterize next the optimal service designs.

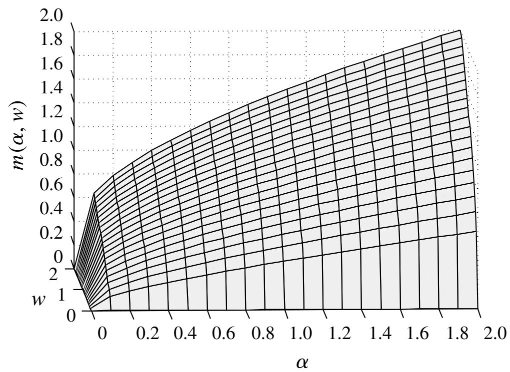
First, the weighting function is symmetric in α and w; i.e., $\Phi(\alpha,w,t)=\Phi(w,\alpha,t)$. Consequently, a design that is optimal with pure acclimation α and no memory decay is also optimal with pure memory decay, with $w=\alpha$, and no acclimation. That is, acclimation and memory decay, despite being fundamentally different phenomena, lead to the same designs when considered individually.

In fact, this mathematical symmetry allows us to reduce the problem dimensionality. Instead of characterizing a problem instance by a pair of acclimation and memory decay rates (α,w), it turns out that we can characterize it by a single number—namely, the logarithmic mean (Carlson 1972) of the acclimation rate and the memory decay rate:

$$m(\alpha,w) \doteq \begin{cases} 0 & \text{if } \alpha=0 \text{ or } w=0,\\ \alpha & \text{if } \alpha=w,\\ \dfrac{w-\alpha}{\ln w-\ln\alpha} & \text{otherwise.}\end{cases}$$

Accordingly, all our design results will be expressed in terms of $m(\alpha,w)$. As depicted in Figure 6, $m(\alpha,w)$ is symmetric and increasing in both α and w. Moreover, it lies between the geometric and the arithmetic means; i.e., $\sqrt{\alpha\cdot w}\le m(\alpha,w)\le(\alpha+w)/2$ for any $\alpha,w\ge 0$.

Figure 6 **The Logarithmic Mean $m(\alpha, w)$ Is Symmetric and Tends to Zero When Either α or w Tends to Zero**



Second, $\Phi(\alpha, w, t)$ is pseudoconcave in t and peaks when $t = 1/m(\alpha, w)$; see Lemma A1 in the appendix. Accordingly, if the service encounter consists of only one activity, i.e., $n = 1$, such that $S = (x_1 - b(0))\Phi(\alpha, w, t_1)$, and $x_1 > b(0)$, satisfaction peaks when the duration is $1/m(\alpha, w)$. Intuitively, the duration of the activity must be long enough to give customers an opportunity to enjoy it but short enough to avoid their acclimating to it.

Third, $\Phi(\alpha, w, t)$ is concave convex in t, and its inflection point occurs at $t = 2/m(\alpha, w)$; see Lemma A3 in the appendix. Based on (6), this shows that the weights on service levels, $\Phi(\alpha, w, \bar{T}_k) - \Phi(\alpha, w, \bar{T}_{k+1})$, are decreasing up to time $\bar{T}_k = 2/m(\alpha, w)$ and then increasing.

Using these three properties, we next characterize the optimal design in terms of the maximum of $\Phi(\alpha, w, \bar{T}_k)$ (i.e., $1/m(\alpha, w)$) and of its inflection point (i.e., $2/m(\alpha, w)$).

4.2. Variable Sequence and Fixed Duration

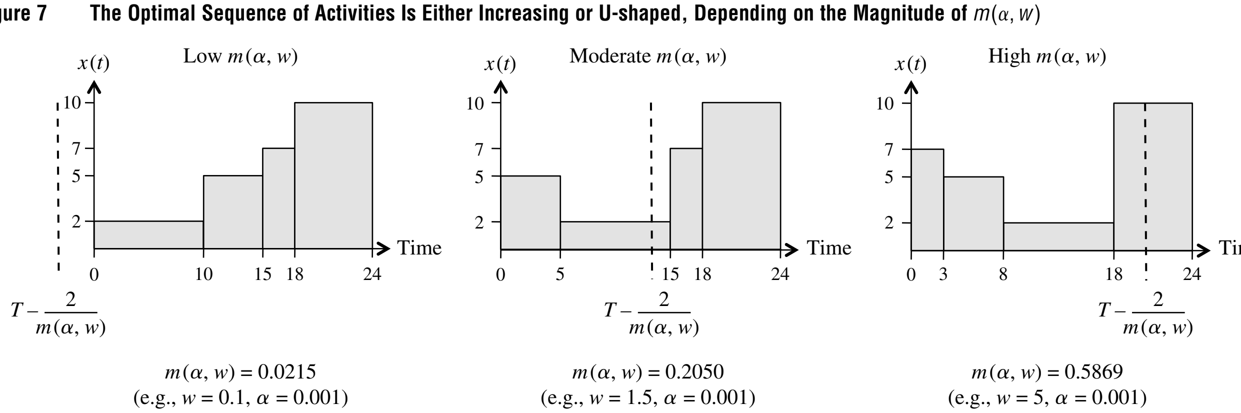
In VSFD, the service provider can freely change the sequence of activities, but their durations are fixed. The next proposition shows that, individually, both memory decay and acclimation favor designs with increasing service levels. However, when they act together, the optimal sequence is, in general, U-shaped. Figure 7 illustrates Proposition 1 for different values of $m(\alpha, w)$.

Proposition 1. *In VSFD,*

1. *When $m(\alpha, w) \le 2/T$, it is optimal to sequence activities in increasing order of service levels; i.e., $x_{i_1^*} < \cdots < x_{i_n^*}$.*
2. *When $m(\alpha, w) > 2/T$, it is optimal to sequence activities in a U-shaped fashion of service levels; i.e., there exists a k such that $x_{i_1^*} > \cdots > x_{i_{k-1}^*} > x_{i_k^*} < x_{i_{k+1}^*} < \cdots < x_{i_n^*}$. In particular, either this kth activity (i_k), or its direct predecessor (i_{k-1}), or its direct successor (i_{k+1}) should take place when there remain $2/m(\alpha, w)$ time units until the end of the encounter. Furthermore, $k < n$; i.e., the last two activities are always sequenced in increasing order of service levels.*

Hence, when $m(\alpha, w) \le 2/T$, i.e., when either α or w is small, a *crescendo* is optimal; see Figure 7, left panel. This is the case, in particular, with either pure acclimation ($w = 0$) or pure memory decay ($\alpha = 0$). We next discuss the intuition for both the case when w is small and the case when α is small. First, when w is small, the customer gives equal weight to past and recent utilities. Because of acclimation, any drop in service level creates a negative instantaneous utility (as illustrated in Figure 4), which will be remembered for a long time since w is small. Hence, with small w, it is optimal to have a monotonically increasing sequence of service levels. Second, when α is small, the customer acclimates to the service level of an activity very slowly. Therefore, the higher the service level, the larger the magnitude of instantaneous utility. Because of memory decay, these high utilities are more likely to be remembered if they happen at the end of the encounter. Hence, with small α, it is also optimal to have a monotonically increasing sequence of service levels. The optimality of crescendo is consistent with people's preferences for increasing outcomes over a time frame (Loewenstein and Prelec 1993), with loss aversion (Tversky and Kahneman 1991), and also with what

Figure 7 **The Optimal Sequence of Activities Is Either Increasing or U-shaped, Depending on the Magnitude of $m(\alpha, w)$**



Note. The last two activities should always be in increasing order.

a basic application of the peak–end rule (Fredrickson 2000) to design would imply.

By contrast, when $m(\alpha, w) > 2/T$, i.e., when both α and w are large, Proposition 1 shows that the optimal sequence of activities is *U-shaped*: the service level should fall gradually in the beginning and rise steeply at the end; see Figure 7, middle and right panels. In particular, the last two service levels should always be increasing. The intuition is that, with both acclimation and memory decay, it is desirable to maximize the gradient of service level toward the end of the encounter. Because high acclimation α generates little utility from slow, gradual increases in service level, and high memory decay w makes the customer heavily discount past experiences, the service provider can gradually decrease the service level in the early part of the sequence, which will be forgotten as a result of high memory decay, so as to ensure a steep positive gradient toward the end.

Hence, although it is sometimes recommended to "finish strong" (Chase and Dasu 2001), Proposition 1 indicates that it is not merely the intensity of the last activity that matters but the steepness of the gradient leading to it; i.e., one should "finish with a steep gradient." For instance, in Tchaikovsky's *1812 Overture*, the finale, which celebrates the Russian victory over Napoleon's army with cannon fire, ringing chimes, and brass fanfare, is especially climatic because it is contrasted with the preceding score, which represents the French invasion of Moscow. When designing song playlists, people derive the highest utility when they not only schedule their favorite song at the end but also precede it with a less preferred song (Kahn et al. 1997).

U-shaped sequences of activities appear in many experiences, such as concerts (Baucells et al. 2013), concertos (Rozin et al. 2004), opera season scheduling (Dixon and Thompson 2015), and movies and theme parks (Hormeß and Lawrence 2013). U-shaped sequences turn out to be also optimal in consumption planning in the presence of either acclimation (Wathieu 1997) or satiation (Baucells and Sarin 2010). We thus reach similar conclusions to those consumption planning models (albeit slightly different since decreasing sequences are never optimal here) from a different perspective, by making evaluation retrospective instead of prospective. We also note that a U-shaped sequence could emerge for other reasons, such as because of the primacy and recency effects, and we thus offer a complementary, yet more subtle, justification of the optimality of such sequences.

Moreover, what happens within the U shape may differ. In particular, the requirement that one should finish with a steep gradient may differ from these other models. For instance, Dixon and Verma (2013) speculate that, for a performing arts season, it may be optimal to start with the peak activity, immediately followed by a steep drop in service levels, and then gradually increase the service levels until the end of the season. Although this is probably motivated by reasons other than acclimation and memory decay (e.g., engagement), we find here that it is optimal to adopt the opposite pattern, i.e., a gradual drop followed by a steep increase.

Also, to maximize the end gradient, it may not always be optimal to place the activity with the highest service level at the end, contrary to the crescendo sequence. In particular, when the activity with the highest service level has a long duration, the customer becomes acclimated to that high service level and the customer's utility diminishes over time. In that case, it may be optimal to move that activity at the beginning of the encounter so as to ensure a steep gradient near the end.[4]

Although Proposition 1 characterizes the optimal sequence, it offers little indication of how to construct the optimal U shape. This problem turns out to be exponentially complex, and commercial solvers may fail to return a solution when the number of activities becomes large.[5] In the electronic companion, we propose a heuristic to construct a U-shaped sequence so as to ensure a steep gradient near the end. Numerical experiments suggest that the heuristic performs well.

4.3. Fixed Sequence and Variable Duration

In FSVD, the service provider controls the duration of activities, but their sequence is fixed. The following proposition characterizes the optimal duration allocation across a subsequence of activities.

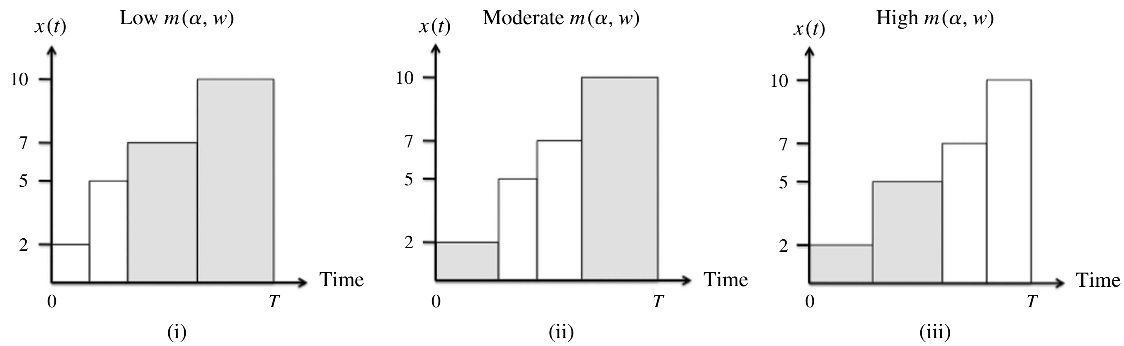
Proposition 2. *In FSVD,*

1. *When $m(\alpha, w) < 1/T$, if it is optimal to allocate more than the minimum duration to activity q, i.e., if $t_q^* > \underline{\tau}_q$, then it is optimal to allocate maximum duration to all subsequent activities if they have increasing service levels, i.e., $t_j^* = \bar{\tau}_j$, $j = q+1, \ldots, r$ if $x_q < \cdots < x_r$, and to all preceding activities if they have decreasing service levels, i.e., $t_j^* = \bar{\tau}_j$, $j = l, \ldots, q-1$ if $x_l > \cdots > x_q$.*
2. *When $1/T \le m(\alpha, w) \le 1/(\sum_{i=r}^{n} \underline{\tau}_i)$, for some r, it is never optimal to allocate more than the minimum duration only in the middle of an increasing sequence, i.e., if $x_l < \cdots < x_r$ and $t_l^* = \underline{\tau}_l$, $t_r^* = \underline{\tau}_r$, then $t_i^* = \underline{\tau}_i$, $\forall\, i, l < i < r$; additionally, it is never optimal to allocate more than*

[4] For instance, when $(x_1, x_2, x_3, x_4) = (2, 5, 7, 10)$ and $(t_1, t_2, t_3, t_4) = (5, 4, 3, 8)$, the optimal sequence when $w = 1$ and $\alpha = 0.7$ is $(i_1, i_2, i_3, i_4) = (4, 2, 1, 3)$; i.e., the activity with the highest service level is placed at the beginning.

[5] Because VSFD can be formulated as a single machine job scheduling problem with the objective of maximizing the total weighted profit $\sum_k x_{i_k} f(T_{i_k})$, where $f(T_{i_k})$ is pseudoconvex (see Lemma A2 in the appendix), the computational complexity of such problems, even without precedence constraints, is an open problem (Höhn and Jacobs 2012).

Figure 8 **Optimal Duration Allocation for an Increasing Sequence**



Note. The shaded activities are allocated duration above their lower bound.

the minimum duration only in the extremities of a decreasing subsequence, i.e., if $x_l > \cdots > x_r$ and $t_l^* > \underline{\tau}_l$, $t_r^* > \underline{\tau}_r$, then $t_i^* > \underline{\tau}_i$, $\forall\, i, l < i < r$.

3. *When $m(\alpha, w) > 1/(\sum_{i=r}^{n} \underline{\tau}_i)$ for some r, if it is optimal to allocate more than the minimum duration to activity q, $q < r$, i.e., if $t_q^* > \underline{\tau}_q$, then it is optimal to allocate maximum duration to all preceding activities if they have increasing service levels, i.e., $t_j^* = \bar{\tau}_j$, $j = l, \ldots, q-1$ if $x_l < \cdots < x_q$, and to all subsequent activities, up to activity r, if they have decreasing service levels, i.e., $t_j^* = \bar{\tau}_j$, $j = q+1, \ldots, r$ if $x_q > \cdots > x_r$.*

Figures 8 and 9 illustrate Proposition 2 with two service encounters: (i) an encounter with all increasing service levels (Figure 8) and (ii) an encounter with all decreasing service levels (Figure 9). In the figures, the shaded activities are allocated duration above their lower bound, i.e., $t_i^* > \underline{\tau}_i$.

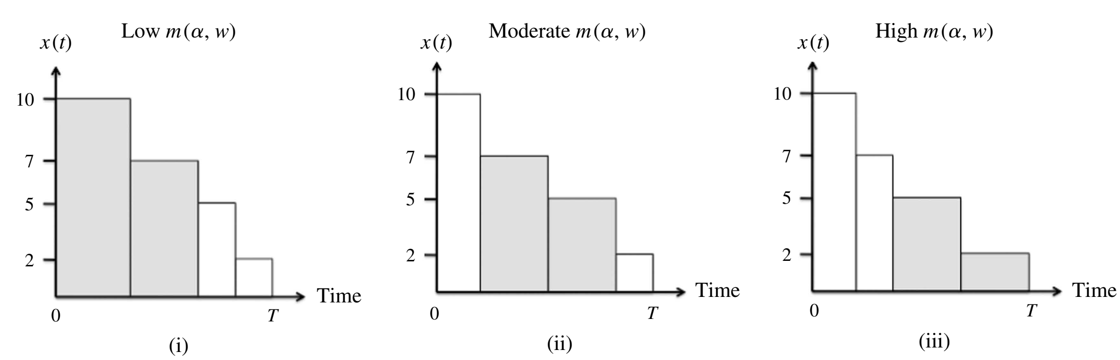
When $m(\alpha, w) < 1/T$, i.e., either α or w is low, Proposition 2 shows that duration should be allocated to the activities in *decreasing* order of service levels (see (i) in Figures 8 and 9). We next discuss the intuition, both for the case when w is small and for the case when α is small. First, when w is small, this design strategy naturally ensures that the maximum time is spent at high service levels. Second, when α is small, this design strategy brings closer to the end the activities associated with a high service level, so that the customer does not heavily discount them.

By contrast, when $m(\alpha, w) > 1/\underline{\tau}_n$, i.e., when both α and w are large, then duration should be allocated in *increasing* order of service levels so as to push the steep rise and gradual fall in service level toward the end of the encounter (see (iii) in Figures 8 and 9). Large memory decay implies that the customer forgets the initial experience of the encounter. Therefore, with an increasing sequence, it is optimal to allocate duration to the activities in the beginning so as to bring closer to the end the steep rise in service levels. Conversely, in a decreasing sequence, it is optimal to allocate duration at the end so as to push the steep drop in service levels as far as possible from the end of the sequence and make the gradient of service level near the end tend to zero.

In the intermediate cases, when $1/T \leq m(\alpha, w) \leq 1/\underline{\tau}_{i_n}$, duration can be allocated in any of the three possible ways shown in Figures 8 and 9 for an increasing and a decreasing sequence, respectively. However, duration should never be allocated only to the activities in the middle of an increasing sequence or in the extremities of a decreasing sequence.

Figure 8(iii) indicates that, in the presence of both acclimation and memory decay, the optimal design should induce a steep gradient at the end of the encounter, even if that involves spending less time on high service-level activities. A grand finale is desirable, but it needs to be short and to contrast with the

Figure 9 **Optimal Duration Allocation for a Decreasing Sequence**



Note. The shaded activities are allocated duration above their lower bound.

preceding activities. This corroborates the optimality of the U-shaped sequence in VSFD, which introduced some respite before the final ascent in service levels.

Similarly, Figure 9(iii) indicates that, in the presence of both acclimation and memory decay, it is preferable to spend more time at low service levels than to provide great service initially and then create disappointment just before the end of the encounter. Pushing that logic further, if $\underline{\tau}_i = 0$, $\forall i$, it would then be optimal to allocate zero duration to the (initial) activities associated with the highest service level, effectively removing them from the encounter. Hence, an anticlimax is never optimal when both acclimation and memory decay are salient.

Together, Figures 8(iii) and 9(iii) indicate that, instead of offering a constant service level, a service provider may actually create more satisfaction by interrupting the experience with a less pleasant break, if the customer is subject to both memory decay and acclimation, consistent with what was empirically observed by Nelson and Meyvis (2008). Hence, deteriorating an experience may be optimal, provided that it induces a steep gradient near the end. In fact, the "service recovery paradox" stipulates that, in some situations, the satisfaction of customers who experienced a service failure and who were recovered by the firm may actually exceed the satisfaction of customers who have not encountered any problems (McCollough and Bharadwaj 1992).

Given the design scope of our model, of course we are not suggesting that the service provider should plan for a service failure and recovery (since customers would be able to internalize the whole process in their ex ante expectations). Nevertheless, the service provider could time a less enjoyable activity in the next-to-last position so as to induce a steep gradient near the end. For instance, a coffee shop may actually take advantage of the disagreeable checkout process by ensuing it with an enjoyable activity (e.g., offering a free cookie). Similarly, an instructor should not relinquish teaching difficult material provided that the lesson ends with easier material (Hoogerheide and Paas 2012). In music, some scores purposely feature a ritardando before finishing with an explosive finale.

We close this section by noting that Proposition 2 offers only a partial characterization of the optimal duration allocation. In general, FSVD is ill behaved, with multiple local optima, unless the sequence of activities is increasing in service levels, in which case the objective function is concave (see Lemma A5 in the appendix). Yet we show in the electronic companion that a simple gradient coordinate algorithm performs reasonably well.

4.4. Variable Sequence and Variable Duration

In VSVD, the service provider controls both the sequence and the duration of activities. Combining Propositions 1 and 2 shows that the optimal solution for VSVD is obtained by constructing the steepest positive gradient near the end.

Corollary 1. *In VSVD,*

1. *When $m(\alpha, w) < 2/T$, a sequence with increasing service levels is optimal. Moreover,*

   (a) *when $m(\alpha, w) < 1/T$, there exists an index k such that minimum duration is allocated to the activities preceding the kth activity, i.e., $t_{i_j} = \underline{\tau}_{i_j}$, $1 \le j \le k-1$, and maximum duration is allocated to the activities subsequent to the kth activity, i.e., $t_{i_j} = \bar{\tau}_{i_j}$, $k < j \le n$;*

   (b) *when $1/T < m(\alpha, w) \le 1/\underline{\tau}_{i_n}$, the set of activities with minimum duration is contiguous; i.e., for any l, r, if $t^*_{i_l} = \underline{\tau}_{i_l}$ and $t^*_{i_r} = \underline{\tau}_{i_r}$, then $t^*_{i_j} = \underline{\tau}_{i_j}$, $\forall j$, $l < j < r$; and*

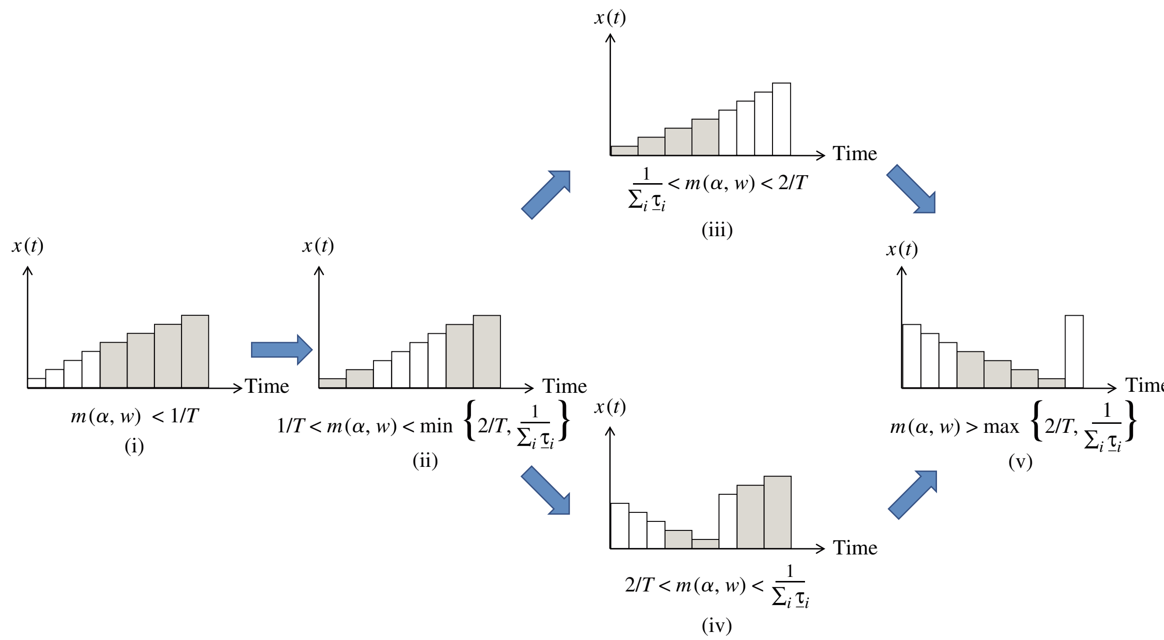
   (c) *when $m(\alpha, w) > 1/\underline{\tau}_{i_n}$, there exists an index k such that maximum duration is allocated to the activities preceding the kth activity, i.e., $t_{i_j} = \underline{\tau}_{i_j}$, $1 \le j \le k-1$, and minimum duration is allocated to the activities subsequent to the kth activity, i.e., $t_{i_j} = \bar{\tau}_{i_j}$, $k < j \le n$.*

2. *If $m(\alpha, w) > 2/T$, a U-shaped sequence bottoming out at the kth activity is optimal such that*
   - *in the increasing part, the set of activities with minimum duration is contiguous; i.e., for any $k \le l < r \le n$ such that $t^*_{i_l} = \underline{\tau}_{i_l}$ and $t^*_{i_r} = \underline{\tau}_{i_r}$, then $t^*_{i_j} = \underline{\tau}_{i_j}$, $\forall j$, $l < j < r$; and*
   - *in the decreasing part, the set of activities with strictly more than the minimum duration is contiguous; i.e., for any $1 \le l < r \le k$ such that $t^*_{i_l} > \underline{\tau}_{i_l}$ and $t^*_{i_r} > \underline{\tau}_{i_r}$, then $t^*_{i_j} > \underline{\tau}_{i_j}$, $\forall j$, $l < j < r$.*

Figure 10 illustrates Corollary 1 as $m(\alpha, w)$ increases, i.e., as either acclimation or memory decay or both become stronger. Initially, when $m(\alpha, w)$ is low, crescendo is the optimal sequence, with duration initially allocated to the activities at the end of the sequence, as in (i), and then, as $m(\alpha, w)$ increases, also at the beginning of the sequence, as in (ii) and (iii), so as to induce a steeper gradient near the end of the encounter. As $m(\alpha, w)$ keeps increasing, the optimal sequence may switch to a U-shaped sequence, as in (iv) and (v), and more duration is then allocated to the activities in the middle of the encounter, so as to induce more gradual transitions at low levels of service and steeper transitions at higher service levels.

Alternatively, one may interpret Corollary 1 in terms of the total service duration T, since most structural results can be expressed in terms of the product $m(\alpha, w) \cdot T$. Keeping $m(\alpha, w)$ fixed, the optimal service design changes from (i) to (v) as T increases. Accordingly, short experiences (e.g., one-day executive courses) should have activities scheduled as a crescendo and duration allocated primarily to the activities with the highest service levels, whereas long

Figure 10 (Color online) Optimal Sequence and Duration Allocation as $m(\alpha, w)$ Increases



Note. The shaded activities are allocated duration above their lower bound.

experiences (e.g., one-week executive courses) should have activities scheduled in a U-shaped fashion and duration allocated primarily to activities with the lowest service level so as to ensure a steep gradient near the end.

In terms of complexity, VSVD is at least as difficult as FSVD or VSFD. In the electronic companion, we propose a heuristic to first construct a U-shaped sequence when $m(\alpha, w) > 2/T$ or a crescendo sequence otherwise, then allocating duration to activities by following the gradient ascent algorithm used for FSVD. Comparing the performance of the heuristic to the optimal solution (for small n) or to an upper bound (for large n) reveals that this heuristic performs reasonably well.

5. Service Design for a Heterogeneous Population of Customers

In practice, a service provider may not be able to adjust the service design to different customers, characterized by different rates of acclimation (α) and memory decay (w), or to precisely assess those rates. In this section, we provide guidelines regarding the best sequence of activities in the absence of complete knowledge of α and w for every individual customer. Based on our structural results in §4.2, we propose three different sequencing heuristics and numerically test their performance for a randomly generated population of customers.

In our experiments, we consider a service encounter consisting of seven activities and generate 150 instances of service levels and durations. For each instance, both the service level and the duration of each activity are randomly drawn from independent Gamma distributions with scale 2 and shape 2, i.e., $x_i \sim \text{Gamma}(2, 2)$ and $t_i \sim \text{Gamma}(2, 2)$ for $i = 1, \ldots, 7$.

For each instance of service encounter, keeping fixed the service levels and durations, we randomly generate a population of 100 customers (α, w), where the distribution α and w are independent Gammas, with means and standard deviations, respectively, equal to $(\mu_\alpha, \sigma_\alpha)$ and (μ_w, σ_w). For each customer, we compare their satisfaction generated with the following sequences:

- the optimal sequence for that particular customer (α, w), denoted as S^*;
- a crescendo sequence, which orders activities in increasing order of service levels, i.e., $x_{i_1} \leq \cdots \leq x_{i_n}$, denoted as S^{cresc};
- a sequence that maximizes the gradient at the end, by ordering the $(n-1)$-smallest service-level activities in decreasing order of service levels and placing the activity with the highest service level at the end, i.e., $x_{i_1} \geq \cdots \geq x_{i_{n-1}} \leq x_{i_n} = \max x_i$, denoted as S^{steep}; and
- the optimal sequence if the provider ignored customer heterogeneity, i.e., based on the means (μ_α, μ_w), denoted as S^{mean}.

Note that the crescendo and steep gradient sequences require no information about the rates of acclimation and memory decay. The last heuristic, by contrast, requires that the provider know the mean rates of acclimation and memory decay.

Table 2 displays the average optimality gaps, averaged across all 150 instances of service encounters

Table 2 **Average Suboptimality Gaps for the Sequence Based on the Mean Rates and the Steep Gradient and Crescendo Sequences**

		$\sigma_\alpha = \sigma_w$ (%)		
	$\mu_\alpha = \mu_w$	0.1	0.3	0.5
$1 - S^{mean}/S^*$	0.2	4.5	9.9	6.1
	0.5	1.5	9.7	16.9
	0.8	0.3	3.9	8.4
$1 - S^{steep}/S^*$	0.2	28.9	20.7	11.6
	0.5	13.1	13.6	16.9
	0.8	32.9	27.5	21.6
$1 - S^{cresc}/S^*$	0.2	16.4	8.3	4.1
	0.5	53.2	50.6	28.2
	0.8	72.1	68.8	64.0

Note. Sequences are based, respectively, on the mean rates (S^{mean}), the steep gradient sequence (S^{steep}), and the crescendo sequence (S^{cresc}) when $\mu_\alpha = \mu_w$ and $\sigma_\alpha = \sigma_w$.

and, for each encounter, across all 100 customers, for the crescendo sequence, the steep gradient sequence, and the sequence based on the mean rates, when $\mu_\alpha = \mu_w \in \{0.2, 0.5, 0.8\}$ and $\sigma_\alpha = \sigma_w \in \{0.1, 0.3, 0.5\}$.

From Table 2, we make the following observations.

- The performance of the sequence based on the mean rates naturally improves when their standard deviation is low (since it is optimal when $\sigma_\alpha = \sigma_w = 0$), but it also improves when the mean rates increase, indicating that its performance mostly depends on the coefficient of variation.
- By contrast, the performance of the crescendo improves when the mean rates are low (since, by Proposition 1, crescendo is optimal when either α or w is small) or when their standard deviation increases. With a large standard deviation, the Gamma distribution indeed tends to have a large probability mass near zero, similar to the exponential distribution.
- The sequence based on the mean rates overall exhibits a very good performance, with an optimality gap mostly in the single digits, especially in comparison to the other two heuristics. This shows that (i) estimating the mean acclimation and memory decay rates and using those means to guide the service design is very beneficial, but that (ii) knowing more than the mean rates, e.g., by attempting to elicit the actual rates for each individual customer and reconfiguring the service design accordingly, tends to have only a marginal impact.
- The steep gradient sequence appears to generally outperform the crescendo, except when the mean rates are low. Moreover, the steep gradient sequence appears to have a relatively stable performance, in contrast to the crescendo, which performs very poorly at high rates of acclimation and memory decay. This suggests that, in the absence of information about the mean rates of acclimation and memory decay, a service provider may be better off finishing the encounter with a steep gradient, even if that involves lowering the service levels initially, rather than gradually increasing the service level, as could be potentially inferred from a "finish strong" recommendation.

As a robustness check, we next consider a situation when $\mu_\alpha \neq \mu_w$. We moreover assume that $\sigma_w = 0.001$ so as to concentrate all random variations on α. (Because of the symmetry between α and w, the same results could have been obtained by setting $\sigma_\alpha = 0$ and varying w.) Table 3 displays the optimality gaps when $\mu_\alpha, \mu_w \in \{0.2, 0.5, 0.8\}$ and $\sigma_\alpha \in \{0.1, 0.3, 0.5\}$.

In addition to confirming our earlier observations, Table 3 reveals that the performance of the sequence based on the mean rates remains very stable with respect to asymmetric changes in means. Hence, that heuristic does not fit a particular mean pattern. By contrast, the performance of crescendo improves when at least one of the means becomes smaller since, by Proposition 1, it becomes nearly optimal when either rate is zero. In that case, crescendo obviously outperforms the steep gradient sequence, which starts with a diminuendo. On the other hand, when both means are large, crescendo does not perform

Table 3 **Average Suboptimality Gap for the Sequence Based on the Mean Rates and, the Steep Gradient and Crescendo Sequences**

		$\mu_w = 0.2$ (%)			$\mu_w = 0.5$ (%)			$\mu_w = 0.8$ (%)		
	μ_α	$\sigma_\alpha = 0.1$	$\sigma_\alpha = 0.3$	$\sigma_\alpha = 0.5$	$\sigma_\alpha = 0.1$	$\sigma_\alpha = 0.3$	$\sigma_\alpha = 0.5$	$\sigma_\alpha = 0.1$	$\sigma_\alpha = 0.3$	$\sigma_\alpha = 0.5$
$1 - S^{mean}/S^*$	0.2	2.9	8.3	9.3	2.5	9.3	10.3	2.6	5.9	9.1
	0.5	0.4	3.8	8.1	0.6	4.8	10.4	0.5	3.4	8.5
	0.8	0.2	1.8	3.3	0.1	1.3	3.5	0.4	2.5	5.4
$1 - S^{steep}/S^*$	0.2	34.9	23.3	23.0	16.7	17.2	15.1	8.5	13.0	14.0
	0.5	17.8	24.7	20.1	8.6	16.2	17.9	21.5	18.6	17.5
	0.8	11.4	11.7	7.9	29.3	18.5	27.3	35.1	30.2	25.4
$1 - S^{cresc}/S^*$	0.2	18.0	12.7	8.4	35.3	25.2	17.4	36.6	27.8	18.2
	0.5	37.2	27.5	26.5	54.5	48.8	44.3	66.8	61.8	55.6
	0.8	44.4	40.9	41.3	71.6	63.7	65.5	72.2	69.3	64.5

Note. Sequences are based, respectively, on the mean rates (S^{mean}), the steep gradient sequence (S^{steep}), and the crescendo sequence (S^{cresc}) when $\mu_\alpha \neq \mu_w$ and $\sigma_w = 0.001$.

well and is significantly outperformed by the steep gradient sequence. Hence, the steep gradient and the crescendo sequences are in some way complementary, although the former generally outperforms the latter, and sometimes by a significant amount.

6. Conclusion

In this paper, we analytically determine how to sequence and allocate duration to activities in a service encounter so as to maximize ex post customer satisfaction in the presence of memory decay and acclimation. Memory decay favors high service levels at the end of the encounter and acclimation favors steep positive gradients of service level.

We show that, individually, the two behavioral biases yield the same service design. However, when considered jointly, they can act as opposing forces. In particular, if either memory decay or acclimation is low, it is optimal to sequence activities in increasing order of service levels and to lengthen the duration of activities with the highest service levels. By contrast, when both memory decay and acclimation are high, it is optimal to sequence activities in a U-shaped fashion and to lengthen the duration of activities with the lowest service levels, so as to accentuate the final gradient.

Our results can alternatively be interpreted in terms of the total duration of the encounter: short experiences should have activities scheduled as a crescendo and duration allocated primarily to the activities with the highest service levels, whereas long experiences should have activities scheduled in a U-shaped fashion and duration allocated primarily to activities with the lowest service level so as to ensure a steep gradient near the end.

Overall, our results suggest that although "finishing strong" is optimal in the presence of either memory decay or acclimation, the optimal design when both effects are present is more subtle because one should aim at maximizing the gradient near the end. Although an anticlimax is never desirable, a never-ending grand finale should also be avoided. Moreover, it may be a good idea to introduce some respite just before the grand finale so as to accentuate its strong character. Going one step beyond the already controversial idea of adding light pain at the end to induce customers to forget a more painful event (Kahneman et al. 1993), our results suggest it may be desirable to voluntarily degrade the service level of the middle activities to accentuate the strong character of the ending activities as empirically observed by Nelson and Meyvis (2008). In fact, our model suggests it may be desirable to even get rid of some high-intensity activities either to avoid a decline in service level (if they are sequenced at the beginning) or to accentuate the final gradient in service levels (if they are sequenced toward the end). This suggests that properly sequencing and allocating duration to a service encounter can more than compensate for inferior service levels.

This paper takes a first step toward building an analytical framework for designing service experiences. Natural research extensions could relax the assumptions outlined in the introduction, such as (i) characterizing the optimal encounter duration and the optimal choice of service levels (subject to a budget constraint); (ii) characterizing the optimal design in the presence of other behavioral biases (e.g., satiation or loss aversion); (iii) considering services where customers are not captive, i.e., can join after the beginning of the encounter or leave before the end, and/or where they self-control the sequence of activities (e.g., amusement parks); and (iv) investigating the effect of precedence constraints and/or sequence-dependent service levels. We hope that this paper will induce further interest in engineering service experiences to maximize customer satisfaction.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2015.2172.

Acknowledgments

The authors thank Manel Baucells, Javad Nasiry, Rakesh Sarin, Serguei Netessine, the associate editor, and three anonymous referees for insightful comments, which have significantly improved this paper.

Appendix. Proofs

In the proofs, we will use, when there is no ambiguity, $\Phi(t)$ instead of $\Phi(\alpha, w, t)$.

Lemma A1. *The function $\Phi(\alpha, w, t)$ is strictly pseudoconcave in t with a stationary point at $1/m(\alpha, w)$.*

Proof. Suppose $\alpha < w$. We have

$$\Phi'(t) = \frac{e^{-\alpha t}(-\alpha) - e^{-wt}(-w)}{w-\alpha} = \frac{e^{-wt}(w - \alpha e^{t(w-\alpha)})}{w-\alpha} > 0$$

$$\Leftrightarrow w > \alpha e^{t(w-\alpha)} \Leftrightarrow \frac{\ln w - \ln \alpha}{w - \alpha} > t.$$

Since $\Phi'(t) > 0$ when $t < 1/m(\alpha, w)$, $\Phi'(t) = 0$ when $t = 1/m(\alpha, w)$, and $\Phi'(t) < 0$ when $t > 1/m(\alpha, w)$, $\Phi(t)$ is strictly pseudoconcave. Note that the result is symmetric for $w < \alpha$. $\square$

Lemma A2. *The function $f(T_{i_k}) = \Phi(\alpha, w, T - T_{i_k} + t_{i_k}) - \Phi(\alpha, w, T - T_{i_k})$ is strictly pseudoconvex in T_{i_k}.*

Proof. Suppose $\alpha < w$. We have

$$f'(T_{i_k}) = \frac{\alpha e^{-\alpha(T - T_{i_k} + t_{i_k})} - w e^{-w(T - T_{i_k} + t_{i_k})}}{w - \alpha} - \frac{\alpha e^{-\alpha(T - T_{i_k})} - w e^{-w(T - T_{i_k})}}{w - \alpha}.$$

Consider $t_0 = T - (1/(w-\alpha)) \cdot \ln(w(1-e^{-wt_{i_k}})/(\alpha(1-e^{-\alpha t_{i_k}})))$. Since $f'(T_{i_k}) < 0$ when $T_{i_k} < t_0$, $f'(T_{i_k}) = 0$ when $T_{i_k} = t_0$, and $f'(T_{i_k}) > 0$ when $T_{i_k} > t_0$, $f(T_{i_k})$ is strictly pseudoconvex. Note that the result is symmetric for $w < \alpha$. $\square$

Lemma A3. *$\Phi(\alpha, w, t)$ is concave in t $\forall t \in [0, 2/m(\alpha, w)]$ and convex in t $\forall t \in [2/m(\alpha, w), T]$.*

Proof. Since $\Phi''(t) = (\alpha^2 e^{-\alpha t} - w^2 e^{-wt})/(w-\alpha)$, $\Phi''(t) \le 0$, $\forall t \in [0, 2/m(\alpha, w)]$ and $\Phi''(t) \ge 0$, $\forall t \in [2/m(\alpha, w), T]$. Hence, $\Phi(t)$ is concave $\forall t \in [0, 2/m(\alpha, w)]$ and convex $\forall t \in [2/m(\alpha, w), T]$. $\square$

Lemma A4. *In VSFD, the optimal sequence $(i_1^*, \dots, i_n^*)$ is such that for any two consecutive activities $x_{i_k^*}$ and $x_{i_{k+1}^*}$, the following applies:*

(i) *If both activities start and finish within $[0, T-2/m(\alpha, w)]$ then $x_{i_k^*} > x_{i_{k+1}^*}$.*

(ii) *If both activities start and finish within $[T-2/m(\alpha, w), T]$ or if $k = n-1$, then $x_{i_k^*} < x_{i_{k+1}^*}$.*

Proof. The proof uses an interchange argument. Let S^* be the satisfaction obtained from the optimal sequence $(i_1^*, \dots, i_n^*)$. For any j, let S_j^* be the satisfaction obtained by interchanging i_{j-1}^* and i_j^* in the optimal sequence. Therefore, by (6),

$$S^* - S_j^* = (x_{i_{j-1}^*} - x_{i_j^*}) \cdot \big((-\Phi(\bar{T}_j) + \Phi(\bar{T}_{j+1})) + (\Phi(\bar{T}_{j-1}) - \Phi(t_{i_{j-1}^*} + \bar{T}_{j+1}))\big).$$

We next consider three scenarios: (i) when the two activities start and end within $[0, T-2/m(\alpha, w)]$, (ii) when the two activities start and end within $[T-2/m(\alpha, w), T]$, and (iii) when $j = n$.

(i) If i_{j-1}^* and i_j^* start and finish within $[0, T-2/m(\alpha, w)]$, then by Lemma A3, the corresponding $\Phi'(t)$ is increasing $\forall t \in [2/m(\alpha, w), T]$. Hence,

$$\Phi(\bar{T}_j) - \Phi(\bar{T}_{j+1}) < \Phi(t_{i_{j-1}^*} + \bar{T}_j) - \Phi(t_{i_{j-1}^*} + \bar{T}_{j+1}).$$

Therefore, by optimality of $(i_1^*, \dots, i_n^*)$, we have $x_{i_j^*} < x_{i_{j-1}^*}$.

(ii) If i_{j-1}^* and i_j^* start and finish within $[T-2/m(\alpha, w), T]$, then by Lemma A3, the corresponding $\Phi'(t)$ is decreasing $\forall t \in [0, 2/m(\alpha, w)]$. Hence,

$$\Phi(\bar{T}_j) - \Phi(\bar{T}_{j+1}) > \Phi(t_{i_{j-1}^*} + \bar{T}_j) - \Phi(t_{i_{j-1}^*} + \bar{T}_{j+1}).$$

Therefore, by optimality of $(i_1^*, \dots, i_n^*)$, we have $x_{i_j^*} > x_{i_{j-1}^*}$.

(iii) By interchanging activities i_{n-1}^* and i_n^*, we obtain a suboptimal sequence with corresponding satisfaction S_n^*. Therefore,

$$\begin{aligned} S^* - S_n^* &= (x_{i_n^*} - x_{i_{n-1}^*}) \cdot (\Phi(t_{i_n^*}) + \Phi(t_{i_{n-1}^*}) - \Phi(t_{i_{n-1}^*} + t_{i_n^*})) \\ &= \frac{x_{i_n^*} - x_{i_{n-1}^*}}{w - \alpha} \{(1 - e^{-wt_{i_{n-1}^*}})(1 - e^{-wt_{i_n^*}}) \\ &\qquad - (1 - e^{-\alpha t_{i_{n-1}^*}})(1 - e^{-\alpha t_{i_n^*}})\}. \end{aligned}$$

Hence, by optimality of $(i_1^*, \dots, i_n^*)$, we have $x_{i_n^*} > x_{i_{n-1}^*}$. $\square$

Proof of Proposition 1. The proof follows from Lemma A4. $\square$

Lemma A5. *Suppose that x_i is the service level of the activity at the ith position in the sequence. When $x_1 < \cdots < x_n$, $S(\mathbf{t})$ is strictly pseudoconcave, whereas, when $x_1 > \cdots > x_n$, $S(\mathbf{t})$ is strictly pseudoconvex.*

Proof. By Equation (6), $S(\mathbf{t}) = \sum_{i=1}^n (x_i - x_{i-1}) \Phi(\bar{T}_i)$. Taking the partial derivatives of $S(\mathbf{t})$ with respect to t_i, $\forall i$ we obtain

$$\frac{\partial S(\mathbf{t})}{\partial t_i} = \sum_{j=1}^{i} (x_j - x_{j-1}) \Phi'(\bar{T}_j), \quad \forall i.$$

Because $\Phi'(t) = 0$ if and only if $t = 1/m(\alpha, w)$, any stationary point has to satisfy the following set of equations:

$$\sum_{j=i}^{n} t_j = \frac{1}{m(\alpha, w)}, \quad \forall i. \tag{11}$$

Therefore, $\mathbf{t}^* = (0, \dots, 0, 1/m(\alpha, w))$ is the unique stationary point satisfying Equation (11).

We next evaluate the Hessian of $S(\mathbf{t})$ at $\mathbf{t}^*$, denoted H, to obtain for any $\mathbf{y} \in \Re^n$

$$\mathbf{y}^T H \mathbf{y} = \alpha \left(\frac{\alpha}{w}\right)^{\alpha/(w-\alpha)} \left(-(x_2 - x_1)\left(\sum_{i=2}^n y_i\right)^2 - (x_3 - x_2)\left(\sum_{i=3}^n y_i\right)^2 \cdots - (x_n - x_{n-1}) y_n^2\right).$$

Clearly, for $x_1 < \cdots < x_n$, $\mathbf{y}^T H \mathbf{y} < 0$, and for $x_1 > \cdots > x_n$, $\mathbf{y}^T H \mathbf{y} > 0$, $\forall \mathbf{y} \in \Re^n$. Hence, $S(\mathbf{t})$ is strictly pseudoconcave when $x_1 < \cdots < x_n$ and strictly pseudoconvex when $x_1 > \cdots > x_n$. $\square$

Lemma A6. *For FSVD the necessary conditions for optimality are given by*

$$\begin{aligned} \frac{\partial S(\mathbf{t})}{\partial t_i} &= \bar{\lambda}_i + \mu, \quad \forall i \in I_U, \\ \frac{\partial S(\mathbf{t})}{\partial t_i} &= \mu, \quad \forall i \in I_M, \\ \frac{\partial S(\mathbf{t})}{\partial t_i} &= \mu - \underline{\lambda}_i, \quad \forall i \in I_L, \end{aligned} \tag{12}$$

together with constraints (8) and (9), in which $\mu \in \Re$, $\underline{\lambda}_i \ge 0$, and $\bar{\lambda}_i \ge 0$; moreover, I_U, I_M, and I_L are three disjoint sets such that $I = I_U \cup I_M \cup I_L$ and $t_i = \bar{\tau}_i$, $\forall i \in I_U$, $t_i = \tau$, $\underline{\tau}_i < \tau < \bar{\tau}_i$, $\forall i \in I_M$, and $t_i = \underline{\tau}_i$, $\forall i \in I_L$.

Proof. The necessary conditions for optimality of $\mathbf{t}^*$ are given by the following Karush-Kuhn-Tucker conditions (Boyd and Vandenberghe 2004):

1. The stationarity condition gives $\partial S(\mathbf{t})/\partial t_i - \mu + \underline{\lambda}_i - \bar{\lambda}_i = 0$, $\forall i$ at $\mathbf{t}^*$.
2. Complementary slackness gives $\mu(T - \sum_{i=1}^n t_i^*) = 0$, $\bar{\lambda}_i(\bar{\tau}_i - t_i^*) = 0$, $\forall i$, and $\underline{\lambda}_i(t_i^* - \underline{\tau}_i) = 0$, $\forall i$.
3. Primal feasibility implies that $\mathbf{t}^*$ satisfies the constraints (8) and (9).
4. Dual feasibility implies that $\mu \in \Re$ and $\underline{\lambda}_i, \bar{\lambda}_i \ge 0$, $\forall i$.

From the complementary slackness condition, $\underline{\lambda}_i = 0$, $\forall i \in I_U$, and $\bar{\lambda}_i = 0$, $\forall i \in I_L$, and both $\underline{\lambda}_i = 0$, $\bar{\lambda}_i = 0$, $\forall i \in I_M$. $\square$

Proof of Proposition 2. We use Lemmas A1 and A6 for this proof. By (6), we have

$$\frac{\partial S(\mathbf{t}^*)}{\partial t_i} = \sum_{j=1}^{i} (x_j - x_{j-1}) \Phi'(\bar{T}_j), \quad \forall i,$$

where $x_0 = b(0)$.

1. Suppose that $x_q < \cdots < x_r$. Since $T < 1/m(\alpha, w)$, $\Phi'(t) > 0$, $\forall t \in [0, T)$ by Lemma A1, and therefore $\partial S(\mathbf{t})/\partial t_q < \cdots < \partial S(\mathbf{t})/\partial t_r$, $\forall \mathbf{t} \geq (\underline{\tau}_1, \ldots, \underline{\tau}_n)$. Hence, from condition (A2), if $t_q > \underline{\tau}_q$, $\partial S(\mathbf{t}^*)/\partial t_q = \mu$ or $\partial S(\mathbf{t}^*)/\partial t_q = \mu + \bar{\lambda}_q$. Therefore, $\partial S(\mathbf{t}^*)/\partial t_j = \mu + \bar{\lambda}_j$, $j = q+1, \ldots, r$. From Lemma A6, $t_j^* = \bar{\tau}_j$, $j = q+1, \ldots, r$. The proof for a decreasing subsequence is similar.

2. We show the result by contradiction. Suppose that $x_l < \cdots < x_r$ and that $t_l = \underline{\tau}_l$ and $t_r = \underline{\tau}_r$. Therefore, condition (A2) implies that $\partial S(\mathbf{t}^*)/\partial t_l = \mu - \underline{\lambda}_l$ and $\partial S(\mathbf{t}^*)/\partial t_r = \mu - \underline{\lambda}_r$. If for any $i, l < i < r$ we have $t_i^* > \underline{\tau}_i$, then $\partial S(\mathbf{t}^*)/\partial t_i = \mu$ or $\partial S(\mathbf{t}^*)/\partial t_i = \mu + \lambda_i$, by condition (A2). If $\sum_{h=i}^n t_h^* \leq 1/m(\alpha, w)$, then $\Phi'(t) > 0$ for $t \leq \sum_{h=i+1}^n t_h^* < 1/m(\alpha, w)$ from Lemma A1, and therefore $\partial S(\mathbf{t}^*)/\partial t_r > \partial S(\mathbf{t}^*)/\partial t_i$, thereby a contradiction. If $\sum_{h=i}^n t_h^* > 1/m(\alpha, w)$, then $\Phi'(t) < 0$ for $t \geq \sum_{h=i}^n t_h^*$ from Lemma A1, and therefore $\partial S(\mathbf{t}^*)/\partial t_l > \partial S(\mathbf{t}^*)/\partial t_i$, thereby a contradiction. Therefore, from Lemma A6, $t_i^* = \underline{\tau}_i$, $\forall i, l < i < r$. The proof for a decreasing subsequence is similar.

3. Suppose that $x_l < \cdots < x_q$. When $\sum_{i=r}^n \underline{\tau}_i > 1/m(\alpha, w)$, we have $\Phi'(t) < 0$, $\forall t \in (\sum_{i=r}^n \underline{\tau}_i, T]$ by Lemma A1, and therefore $\partial S(\mathbf{t})/\partial t_l > \cdots > \partial S(\mathbf{t})/\partial t_q$, $\forall \mathbf{t} \geq (\underline{\tau}_1, \ldots, \underline{\tau}_n)$. From condition (A2), since $t_q > \underline{\tau}_q$, $\partial S(\mathbf{t}^*)/\partial t_q = \mu$ or $\partial S(\mathbf{t}^*)/\partial t_q = \mu + \bar{\lambda}_q$. Therefore, $\partial S(\mathbf{t}^*)/\partial t_j = \mu + \bar{\lambda}_j$, $j = l, \ldots, q-1$. Therefore, from Lemma A6, $t_j^* = \bar{\tau}_j$, $j = l, \ldots, q-1$. The proof for a decreasing subsequence is similar. □

References

Adelman D, Mersereau AJ (2013) Dynamic capacity allocation to customers who remember past service. *Management Sci.* 59(3): 592–612.

Aflaki S, Popescu I (2013) Managing retention in service relationships. *Management Sci.* 60(2):415–433.

Anderson JR (1995) *Learning and Memory: An Integrated Approach* (John Wiley & Sons, New York).

Ariely D (1998) Combining experiences over time: The effects of duration, intensity changes and on-line measurements on retrospective pain evaluations. *J. Behavioral Decision Making* 11(1): 19–45.

Asch SE (1946) Forming impressions of personality. *J. Abnormal Soc. Psych.* 41(3):258–290.

Bahrick HP (1984) Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *J. Experiment. Psych.: General* 113(1):1–29.

Baucells M, Bellezza S (2016) Temporal profiles of instant utility during anticipation, event, and recall. *Management Sci.* Forthcoming.

Baucells M, Sarin R (2007) Satiation in discounted utility. *Oper. Res.* 55(1):170–181.

Baucells M, Sarin RK (2010) Predicting utility under satiation and habit formation. *Management Sci.* 56(2):286–301.

Baucells M, Sarin R (2012) *Engineering Happiness: A New Approach for Building a Joyful Life* (University of California Press, Berkeley).

Baucells M, Sarin RK (2013) Determinants of experienced utility: Laws and implications. *Decision Anal.* 10(2):135–151.

Baucells M, Smith D, Weber M (2013) Preferences over sequences: Empirical evidence from music. Working paper, University of California, Los Angeles, Los Angeles.

Baumgartner H, Sujan M, Padgett D (1997) Patterns of affective reactions to advertisements: The integration of moment-to-moment responses into overall judgments. *J. Marketing Res.* 34(2):219–232.

Bellos I, Kavadias S (2014) A framework for service design. Working paper, Cambridge University, Cambridge, UK.

Bitran GR, Ferrer J-C, Rocha e Oliveira P (2008) Managing customer experiences: Perspectives on the temporal aspects of service encounters. *Manufacturing Service Oper. Management* 10(1): 61–83.

Boyd SP, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).

Brickman P, Campbell DT (1971) Hedonic relativism and planning the good society. Apley MH, ed. *Adaptation-Level Theory: A Symposium* (Academic Press, New York), 287–301.

Carlson BC (1972) The logarithmic mean. *Amer. Math. Monthly* 79(6):615–618.

Carmon Z, Kahneman D (1996) The experienced utility of queuing: Real time affect and retrospective evaluations of simulated queues. Working paper, Duke University, Durham, NC.

Carmon Z, Shanthikumar JG, Carmon TF (1995) A psychological perspective on service segmentation models: The significance of accounting for consumers' perceptions of waiting and service. *Management Sci.* 41(11):1806–1815.

Caro F, Martínez-de-Albéniz V (2012) Product and price competition with satiation effects. *Management Sci.* 58(7):1357–1373.

Chase RB, Dasu S (2001) Want to perfect your company's service? Use behavioral science. *Harvard Bus. Rev.* 79(6):78–84.

Chase RB, Dasu S (2008) Psychology of the experience: The missing link in service science. Hefley B, Murphy W, eds. *Service Science: Research and Innovations in the Service Economy* (Springer, New York), 35–40.

Constantinides GM (1990) Habit formation: A resolution of the equity premium puzzle. *J. Political Econom.* 98(3):519–543.

Cook LS, Bowen DE, Chase RB, Dasu S, Stewart DM, Tansik DA (2002) Human issues in service design. *J. Oper. Management* 20(2):159–174.

Crano WD (1977) Primacy versus recency in retention of information and opinion change. *J. Soc. Psych.* 101(1):87–96.

Dalton P, Wysocki CJ (1996) The nature and duration of adaptation following long-term odor exposure. *Perception Psychophys.* 58(5):781–792.

Dawes C, Watanabe S (1987) The effect of taste adaptation on salivary flow rate and salivary sugar clearance. *J. Dental Res.* 66(3):740–744.

DeVine J, Gilson K (2010) Using behavioral science to improve the customer experience. *McKinsey Quart.* (February), http://www.mckinsey.com/insights/operations/using_behavioral_science_to_improve_the_customer_experience.

DeVine J, Lal S, Zea M (2012) The human factor in service design. *McKinsey Quart.* (January), http://www.mckinsey.com/insights/operations/the_human_factor_in_service_design.

Dixon MJ, Thompson GM (2015) The impact of scheduling and bundling flexibility on sequence effect-based schedules. Working paper, Ivey Business School, London, ON, Canada.

Dixon MJ, Thompson GM (2016) Bundling and scheduling service packages with customer behavior: Model and heuristic. *Production Oper. Management* 25(1):36–55.

Dixon M, Verma R (2013) Sequence effects in service bundles: Implications for service design and scheduling. *J. Oper. Management* 31(3):138–152.

Ebbinghaus H (1913) *Memory: A Contribution to Experimental Psychology*, translated by Ruger HA, Bussenius CE (Teachers College, Columbia University, New York). [Orig. pub. 1885.]

Edgeworth FY (1881) *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* (C. Kegan Paul, London). [Reprinted 1967 (M. Kelly, New York).]

Ely J, Frankel A, Kamenica E (2015) Suspense and surprise. *J. Political Econom.* 123(1):215–260.

Fisher ML (2004) The Lagrangian relaxation method for solving integer programming problems. *Management Sci.* 50(12, Suppl): 1861–1871.

Fredrickson BL (2000) Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition Emotion* 14(4):577–606.

Fredrickson BL, Kahneman D (1993) Duration neglect in retrospective evaluations of affective episodes. *J. Personality Soc. Psych.* 65(1):45–55.

Gans N, Zhou Y-P (2003) A call-routing problem with service-level constraints. *Oper. Res.* 51(2):255–271.

Gent JF, McBurney DH (1978) Time course of gustatory adaptation. *Perception Psychophys.* 23(2):171–175.

Helson H (1964) *Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior* (Harper & Row, New York).

Heskett JL, Sasser WE, Schlesinger LA (1997) *The Service Profit Chain* (Free Press, New York).

Höhn W, Jacobs T (2012) On the performance of Smith's rule in single-machine scheduling with nonlinear cost. Fernández-Baca D, ed. *LATIN 2012: Theoretical Informatics*, Lecture Notes in Computer Science, Vol. 7256 (Springer, Berlin) 482–493.

Hoogerheide V, Paas F (2012) Remembered utility of unpleasant and pleasant learning experiences: Is all well that ends well? *Appl. Cognitive Psych.* 26(6):887–894.

Hormeß M, Lawrence A (2013) Boom! Wow. Wow! WOW! BOOOOM!!!. *Touchpoint* 4(2):24–27.

Hsee CK, Abelson RK (1991) Velocity relation: Satisfaction as a function of the first derivative of outcome over time. *J. Personality Soc. Psych.* 60(3):341–347.

Hsee CK, Tsai I (2008) Hedonomics in consumer behavior. Haugtvedt CP, Herr PM, Kardes FR, eds. *Handbook of Consumer Psychology* (Lawrence Erlbaum Associates, Mahwah, NJ), 639–658.

Incropera FP, DeWitt DP, Bergman TL, Lavine AS (2006) *Introduction to Heat Transfer*, 5th ed. (John Wiley & Sons, Hoboken, NJ).

Kahn B, Ratner R, Kahneman D (1997) Patterns of hedonic consumption over time. *Marketing Lett.* 8(1):85–96.

Kahneman D (2000a) Evaluation by moments: Past and future. Kahneman D, Tversky A, eds. *Choices, Values, and Frames* (Cambridge University Press, Cambridge, UK), 693–708.

Kahneman D (2000b) Experienced utility and objective happiness: A moment-based approach. Kahneman D, Tversky A, eds. *Choices, Values, and Frames* (Cambridge University Press, Cambridge, UK), 673–692.

Kahneman D, Wakker PP, Sarin R (1997) Back to Bentham? Explorations of experienced utility. *Quart. J. Econom.* 112(2): 375–406.

Kahneman D, Fredrickson BL, Schreiber CA, Redelmeier DA (1993) When more pain is preferred to less: Adding a better end. *Psych. Sci.* 4(6):401–405.

Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA (2004) A survey method for characterizing daily life experience: The day reconstruction method. *Science* 306(5702): 1776–1780.

Lee S (2008) To create advocates, you have to differentiate your organization. *CustomerThink.com* (July 19), http://customerthink.com/create_advocates_differentiate/.

Loewenstein G (1987) Anticipation and the valuation of delayed consumption. *Econom. J.* 97(387):666–684.

Loewenstein GF, Prelec D (1993) Preferences for sequences of outcomes. *Psych. Rev.* 100(1):91–108.

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

Marsh HW (1984) Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *J. Ed. Psych.* 76(5):707–754.

McCollough MA, Bharadwaj SG (1992) The recovery paradox: An examination of consumer satisfaction in relation to disconfirmation, service quality, and attribution based theories. Allen CT, Madden TJ, Shimp TA, Howell RD, Zinkhan GM, Heisley DD, Semenik RJ, Dickson P, Zeithaml V, Jenkins RL, eds. *Marketing Theory and Applications* (American Marketing Association, Chicago).

Miller A (2014) Small portions + the economics of eating. *OZY* (July 29), http://www.ozy.com/acumen/eating-less-food-means-you-enjoy-it-more/32827.article.

Naik PA (1999) Estimating the half-life of advertisements. *Marketing Lett.* 10(3):351–362.

Nasiry J, Popescu I (2011) Dynamic pricing with loss-averse consumers and peak-end anchoring. *Oper. Res.* 59(6):1361–1368.

Nelson LD, Meyvis T (2008) Interrupted consumption: Disrupting adaptation to hedonic experiences. *J. Marketing Res.* 45(6): 654–664.

Nelson LD, Meyvis T, Galak J (2009) Enhancing the television-viewing experience through commercial interruptions. *J. Consumer Res.* 36(2):160–172.

Novemsky N, Ratner RK (2003) The time course and impact of consumers' erroneous beliefs about hedonic contrast effects. *J. Consumer Res.* 29(4):507–516.

Overbosch P (1986) A theoretical model for perceived intensity in human taste and smell as a function of time. *Chemical Senses* 11(3):315–329.

Pine BJ, Gilmore JH (1998) Welcome to the experience economy. *Harvard Bus. Rev.* 76(4):97–105.

Pinedo ML (2005) *Planning and Scheduling in Manufacturing and Services* (Springer, New York).

Plambeck EL, Wang Q (2013) Implications of hyperbolic discounting for optimal pricing and scheduling of unpleasant services that generate future benefits. *Management Sci.* 59(8):1927–1946.

Popescu I, Wu Y (2007) Dynamic pricing strategies with reference effects. *Oper. Res.* 55(3):413–429.

Redelmeier DA, Kahneman D (1996) Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain* 66(1):3–8.

Redelmeier DA, Katz J, Kahneman D (2003) Memories of colonoscopy: A randomized trial. *Pain* 104(1–2):187–194.

Rozin A, Rozin P, Goldberg E (2004) The feeling of music past: How listeners remember musical affect. *Music Perception* 22(1):15–39.

Sampson SE, Weiss EN (1995) Increasing service levels in conference and educational scheduling: A heuristic approach. *Management Sci.* 41(11):1816–1825.

Schreiber CA, Kahneman D (2000) Determinants of the remembered utility of aversive sounds. *J. Experiment. Psych.: General* 129(1):27–42.

Talluri KT, Van Ryzin GJ (2004) *The Theory and Practice of Revenue Management* (Springer, New York).

Tversky A, Griffin D (1991) Endowment and contrast in judgments of well-being. Strack F, Argyle M, Schwarz N, eds. *Subjective Well-Being: An Interdisciplinary Perspective*, International Series in Experimental Psychology (Pergammon Press, Elmsford, NY), 101–118.

Tversky A, Kahneman D (1991) Loss aversion in riskless choice: A reference-dependent model. *Quart. J. Econom.* 106(4): 1039–1061.

Varey C, Kahneman D (1992) Experiences extended across time: Evaluations of moments and episodes. *J. Behavioral Decision Making* 5(3):169–185.

Verhoef PC, Antonides G, de Hoog AN (2004) Service encounters as a sequence of events: The importance of peak experiences. *J. Service Res.* 7(1):53–64.

Voss C, Zomerdijk L (2007) Innovation in experiential services—An empirical view. *Innovation in Services*, UK Department of Trade and Industry (DTI) Economic Papers (DTI, London), 97–134.

Wathieu L (1997) Habits and the anomalies in intertemporal choice. *Management Sci.* 43(11):1552–1563.

Watkinson M (2013) *The Ten Principles Behind Great Customer Experiences* (Financial Times Publishing International, Harlow, UK).

Watt JH, Mazza M, Snyder L (1993) Agenda-setting effects of television news coverage and the effects decay curve. *Comm. Res.* 20(3):408–435.

Wickelgren WA (1974) Single-trace fragility theory of memory dynamics. *Memory Cognition* 2(4):775–780.